American Society of Human Genetics 65th Annual Meeting
October 6–10, 2015 Baltimore, MD

POSTER ABSTRACTS

Session Topic/Title	Abstract/Poster Board Numbers	
	Start #	End #
Bioinformatics and Genomic Technology	1588	1917
Cancer Genetics	2567	2794
Cardiovascular Genetics	508	647
Clinical Genetics and Dysmorphology	2344	2489
Clinical Genetic Testing	1918	2099
Complex Traits and Polygenic Disorders	704	1021
Cytogenetics	2490	2538
Development	2539	2566
Epigenetics	391	467
Ethical, Legal, Social, and Policy Issues in Genetics	2155	2195
Evolutionary and Population Genetics	1447	1587
Genetic Counseling	2196	2215
Genetics/Genomics Education	2216	2241
Genome Structure, Variation, and Function	3026	3193
Health Services Research	2242	2262
Metabolic Disorders	2263	2343
Molecular Basis of Mendelian Disorders	2795	3025
Pharmacogenetics	648	703
Prenatal, Perinatal, and Reproductive Genetics	2100	2154
Psychiatric Genetics, Neurogenetics, and Neurodegeneration	1022	1253
Statistical Genetics and Genetic Epidemiology	1254	1446
Therapy for Genetic Disorders	468	507

Posters should remain on the board for all three days (Wednesday through Friday)

The program and abstract/poster board number next to each listing is followed by an **W** (Wednesday), **T** (Thursday), or **F** (Friday) to indicate the day on which authors must be present at their poster boards.

391T

The functional MTHFR C677T polymorphism modulates global DNA methylation and associates with non-random differentially methylated regions (DMRs) in autism spectrum disorder. *A.J.M. McNaughton[1], P. Liang[2], K. Calli[3], S. Gannon[3], Y. Qiao[3], M. Hudson[1], F. Solehdin[3], N. Nebesio[2], M. Ayub[1], E. Rajcan-Separovic[3], S. Lewis[3], X. Liu[1].* 1) Dept. Psychiatry, Queen's University, Kingston, Ontario, Canada; 2) Dept. Biological Sciences, Brock University, St. Catharines, Ontario, Canada; 3) Dept. Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada.

5,10-methylenetetrahydrofolate reductase (MTHFR; EC 1.5.1.20) is a key enzyme in DNA methylation. Two functional polymorphisms in *MTHFR*, C677T (rs1801133, p.Asp233Asn) and A1298C (rs1801131, p.Glu429Ala), were extensively studied in many disorders and C677T was implicated to be linked with autism in multiple studies with the low activity allele 677T associated with increased risk for autism especially in sporadic cases. We hypothesized that the 677T allele may be associated with reduce global DNA methylation levels in a non-random fashion which could influence condition-specific gene pathways. In this study, we investigated C677T in relation to global methylation levels in subjects with autism spectrum disorder (ASD) using methyl binding protein sequencing (MBD-Seq) approach. We found that the 677T allele carriers had lower global methylation levels than the 677CC allele group, although the significance is marginal. 154 significant differentially methylated regions (DMRs) were identified between the two allele groups. Interestingly, despite a genome-wide hypomethylation, a significant number of DMRs (93/154 or 60.4%) show a hypermethylation status in the 677T allele group. The identified DMRs demonstrate a highly heterogeneous distribution pattern, with chromosomes 3 and 9 having significantly more DMRs than expected for their chromosome size, while chromosomes 2, 5 and 13 showed an opposite pattern. Furthermore, these DMRs are highly biased towards CpG islands, especially those in non-repetitive intergenic regions. Most interestingly, about 10% (14/154) of these DMRs are found in proximity to genes known to be associated with ASD and related conditions, despite an apparent under-representation of these genes. Altogether, these results suggest that *MTHFR* 677T may modulate global DNA methylation and target autism related genes in a non-random fashion to increase the risk for autism.

392F

Alignment and Methylation Analysis of Tandem Repeat Elements in Cancer using Targeted Next-Gen Bisulfite Sequencing. *J. Alexander[1], J. Zhou[1], A. Meyer[1], A. Miller[1], M. Poulin[1], L. Yan[1], W. Timp[2].* 1) EpigenDx Inc., Hopkinton, MA., USA; 2) Johns Hopkins University, Baltimore, MD, USA.

Global DNA hypomethylation is a common epigenetic alteration in tumor cells, and is closely associated with hypomethylation of interspersed DNA repetitive elements. This hypomethylation is thought to promote the initiation and progression of tumorigenesis through the aberrant activation of repetitive elements. Tandem repeat elements such as NBL-2, Sat-2 and D4Z4 have shown more variability in DNA methylation from normal to cancer tissue. Some cancer specimens show NBL-2 and D4Z4 hypomethylation and others hypermethylation. It appears that the changes in DNA methylation levels of individual tandem repetitive elements may be unique for each type of cancer. This indicates that distinct epigenetic factors may have unique roles. The chromosomal locations of these repeats may be an important factor in whether they are hypo or hypo methylated. Bisulfite PCR primers for repetitive elements were designed to a relatively conserved region of the element to avoid using degenerated primers. The PCR region between the primers is more variable depending on their location within the genome. Multiplex PCR including assays for Sat-2 and NBL-2 was performed on tissues from 40 cancerous and adjacent normal tissue in four cancer types (breast, colon, ovarian and cervical). Following the Ion Torrent protocol, libraries were generated and sequenced on a PGM system. The resulting FASTQ files were then aligned to two references: the whole bisulfite genome, and each chromosome individually. Results from the two alignments were compared to identify any differences, which could indicate artificial alignments. Simplex PCRs of the assays for Sat-2 and NBL-2 were then performed and subjected to the same sequencing preparation and analysis. The comparison of multiplex and simplex PCRs, as well as alignment to different references, may provide insight into the most accurate approach for repetitive element analysis on Next-Gen Sequencing. By mapping the differences in the locations of Sat-2 and NBL-2 the 40 pairs of samples, valuable information about the role epigenetic regulation may play in repetitive element transposition and cancer.

393T

Integrating single-base resolution quantitative epigenomic and transcriptomic sequence data to analyze differential gene expression in normal and tumor liver tissue samples. *K.R. Booher, H. Chung, D. Tan, Y.C. Chew, D. Cubberley, E. Putnam, K. Giang, X. Sun.* Epigenetic Analysis Services, Zymo Research Corporation, Irvine, CA.

Recent technological advances driving the next-generation sequencing revolution finally allow the generation of single-base resolution datasets that fully examine the diverse molecular mechanisms influencing gene expression. Though one-dimensional studies utilizing either RNA-seq, ChIP-seq, or DNA methyl-seq are able to partially explain differential expression, it is increasingly evident that only a pursuit of multi-omics approaches can provide the in depth and detailed information necessary to fully understand gene expression changes associated with human disease. However, combining data from different labs or commercial service providers is subject to user and platform-specific batch effects that may adversely affect data quality and thus interpretation. Using an integrated epigenomics/transcriptomics approach that combines sequence data generated entirely from the same matched liver tumor and adjacent normal tissue sample set in one laboratory setting, we demonstrate a thorough relationship between the interplay of epigenetic influences and gene expression. We assessed genome-wide DNA methylation (5-mC) patterns using an expanded Reduced Representation Bisulfite Sequencing (RRBS)-style protocol that increases genome coverage compared to the classic RRBS method. We overlapped 5-mC data with genome-wide, single-base resolution DNA hydroxymethylation (5-hmC) data generated using a unique protocol known as Reduced Representation Hydroxymethylation Profiling (RRHP) that is specifically designed to pair with RRBS data and differentiate 5-hmC from 5-mC modifications. In addition, chromatin immunoprecipitation followed by sequencing (ChIP-seq), and transcription profiling (RNA-seq) facilitated coupling of the 5-mC/5-hmC data with differential histone modification and gene expression profiles. Complete QC steps to assess starting sample purity and quality (TapeStation, nano-drop, and Qubit), efficient chromatin immunoprecipitation enrichment (qPCR), library construction and quantification (TapeStation and qPCR), and Illumina sequencing read quality score filtering (Trim Galore!) ensured suitable data for downstream analysis. Surprisingly, we found that many sites with similar RRBS-determined 5-mC levels have differential 5-hmC modifications based on the RRHP data. Particularly informative gene regions identified using genome-wide approaches were validated through targeted bisulfite sequencing.

394F

Characterization of differential DNA methylation in transcription factor binding sites across human cancers. *B.N. Lasseigne[1], R.C. Ramaker[1, 2], M.K. Kirby[1], D.S. Gunther[1], S.J. Cooper[1], R.M. Myers[1].* 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) University of Alabama in Birmingham, Birmingham, AL.

CpG DNA methylation (DNAm) is an established driver of cancer etiology, progression, and metastasis. However, genome-wide surveys of DNAm differences between tumor and paired normal tissues often result in thousands of differentially methylated CpGs, complicating interpretation of such results. In order to better understand which CpGs might be important cancer drivers and identify important transcription factors involved in different cancers, we developed and implemented a bioinformatics pipeline for CpG enrichment in the binding sites of 161 transcription factors (TFs) studied by ENCODE. With this pipeline, we examined many in-house and publically available cancer data sets for differential DNAm enrichment or depletion in TF binding sites. We found that most TF binding sites appear to be significantly protected from tumor-induced DNAm changes ("protected TFs"). Additionally, a subset of TFs with binding sites enriched for tumor-induced DNAm changes ("unprotected TFs") appear to be specific to each cancer, implicating particular TFs as drivers of tissue-specific tumor biology. We characterized CpGs in protected and unprotected TFs by many genomic features including local CpG density, CADD score, proximity to transcription start site, proximity to binding motif, and copy number variation. We also examined the relationship between binding site enrichment and TF expression, as well as expression of genes nearest unprotected binding sites with concentrated tumor-induced DNAm changes. As part of an ongoing project to incorporate genome-wide functional regulatory element data with large-scale functional genomic disease datasets we have developed an R package for high-throughput statistical analysis of these datasets to aid in identification of CpGs and TFs that may be important in cancer. This methodology provides a useful tool for interpreting epigenetic alterations in the context of transcription factor regulation in cancer.

395T

HiTMAP: A High-Throughput Methylation Analysis Program for Targeted Bisulfite Sequencing. *B.S. Pullman[1], Y. Yang[1], I. Feng[1], A.B. Costa[2,3], S.A. Scott[1].* 1) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Neurosurgery, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Scientific Computing, Icahn School of Medicine at Mount Sinai, New York, NY.

DNA methylation plays essential roles in gene expression, imprinting and X chromosome inactivation, and epigenetic deregulation has been directly implicated in Mendelian disorders and tumorigenesis. Many techniques have been developed to detect CpG methylation, the most common involving bisulfite conversion coupled with DNA sequencing. The availability of high-throughput sequencing has prompted the development of both whole-genome bisulfite sequencing as well as highly quantitative and multiplexed targeted bisulfite sequencing; however, these methods require significant computational resources and expertise to manage and analyze the data. Therefore, we developed HiTMAP: a High Throughput Methylation Analysis Program to address the need for a stand-alone program capable of analyzing high-throughput targeted bisulfite sequencing data. HiTMAP is a comprehensive web tool that takes raw, targeted bisulfite sequence data (FASTA) and demultiplexes against sample barcodes, aligns sequencing reads to *in silico* bisulfite-converted genomic reference sequences, quantitates CpG methylation levels, and exports resulting methylation data for both individual CpG sites and amplicon regions. The user-facing side of HiTMAP provides an online interface for uploading raw sequence and reference files, setting alignment, methylation quantitation, and quality metric parameters, and for retrieving and saving analysis output data and result figures. The front end was written in JavaScript with AngularJS and D3. Backend computation, including BWA alignment, is accomplished on a server cluster that runs Scala code with the Akka HTTP library for asynchronous data processing. Compute intensive steps, like sequence alignment, are dynamically queued and queried through a generic DRMAA interface to Mount Sinai's batch-processing supercomputer, Minerva. CpG methylation quantitation by HiTMAP was validated using a targeted single-molecule real-time bisulfite sequencing (SMRT-BS; Pacific Biosciences) dataset (four amplicons, 30 cell lines; 21,967 reads), which revealed a very strong correlation (r=0.98) between HiTMAP and the commonly used Bismark bisulfite sequencing analysis pipeline. Coupled with an online raw data submission system (*hitmap.stuartscottlab.org*), HiTMAP eliminates the need for manual data manipulation, local computational resources and expertise, and provides a fast and efficient mechanism to measure CpG methylation from targeted high-throughput bisulfite sequencing data.

396F

A systematic study of normalization methods for Infinium 450K methylation data using whole-genome bisulfite sequencing data. *T. Wang[1], W. Guan[2], J. Lin[5], N. Boutaoui[1], G. Canino[3], J. Luo[4], J.C. Celedón[1], W. Chen[1,5].* 1) Division of Pediatric Pulmonology, University of Pittsburgh, Children's Hospital of Pittsburgh, Pittsburgh, PA; 2) Department of Biostatistics, University of Minnesota, Minneapolis, MN; 3) Behavioral Sciences Research Institute, University of Puerto Rico, San Juan, PR; 4) Department of Pathology, University of Pittsburgh, Pittsburgh, PA; 5) Departments of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA.

DNA methylation, an epigenetic mechanism, plays an important role in disease etiology. The Illumina Infinium HumanMethylation450 (450K) BeadChip is a widely used platform in large-scale epidemiologic studies. This platform can efficiently and simultaneously measure methylation levels at ~480,000 CpG sites in the human genome in multiple study samples. Due to the intrinsic chip design of two types of chemistry probes, data normalization or preprocessing is a critical step to consider before data analysis. To date, numerous methods and pipelines have been developed for this purpose, and some studies have been conducted to evaluate different methods. However, validation studies have often been limited to a small number of CpG sites to reduce the variability in technical replicates.In this study, we measured methylation on a set of samples using both whole-genome bisulfite sequencing (WGBS) and 450K chips. We used WGBS data as a gold standard of true methylation states in cells to compare the performances of eight normalization methods for 450K data on a genome-wide scale. Analyses on our dataset indicate that the most effective methods are peak-based correction (PBC) and quantile normalization plus beta-mixture quantile normalization (QN.BMIQ). To our knowledge, this is the first study to systematically compare existing normalization methods for Illumina 450K data using novel WGBS data. Our results provide a benchmark reference for the analysis of DNA methylation chip data, particularly in white blood cells.

397T

Microsatellite Instability (MSI) is associated with Differential DNA Methylation in Colorectal carcinoma and may have interaction with location of the tumor. *F. Jasmine[1], M. Kamal[2], M. Rahman[2], S. Roy[1], M. Raza[2], R. Paul[2], H. Ahsan[1], Z. Haq[2], M.G. Kibriya[1].* 1) Public Health Sciences, The University of Chicago, Chicago, IL; 2) Bangabandhu Sheikh Mujib Medical University, Dhaka, Bangladesh.

Introduction: Colorectal Cancer (CRC) is one of the most common malignancies worldwide. The role of MSI and DNA methylation in tumor tissue is well known in CRC. MSI may cause error in DNA replication resulting in loss of function of tumor suppressor genes leading to carcinogenesis. In past, using paired tumor-normal tissue on Infinium 27K methylation array we demonstrated a large number of differentially methylated loci to be associated with CRC. The array mainly focused on the promoter CpG islands. In the present study using larger number of samples and 450K methylation array, that addresses promoter sites, shores, shelves, gene body and deep sea on a genome-wide scale, we explored whether differential methylation is associated with MSI status. **Methods:** We carried out a genome-wide methylation assay for a total of 172 paired samples from 86 CRC patients (m=50, f=36) at different stages (stage1:16, stageII: 25 and stageIII: 45). Of them 24 had MSI, and 15 had somatic mutation in KRAS (rs112445441).**Results:** First, we compared the methylation data between tumor and corresponding normal tissue using mixed model 2-way ANOVA and identified a total of 1700 tumor-specific DML covering 634 genes that were significant at FDR 0.05 and the magnitude of difference (delta beta) was at least 20%, which was very close to what we previously reported using the other 27K array. In the next step, we explored if (a) methylation status of these tumor-specific DML were different among the MSI and MSS tumor tissues and (b) if there was interaction between MSI status and location of the tumor (right-sided and left-sided CRC). We found that methylation status was significantly different (at FDR 0.05, with delta beta 20%) in 22 loci (representing 12 genes) between MSI and MSS tumors. All these loci were hypermethylated in MSI tumors compared to MSS. Of these 22 loci, 21 were in CpG Islands near the gene and only one was in shore (closest gene *ETS1*). We also found significant interaction ($p<0.05$) between MSI status and location of the tumor for 4 loci (*GATA2, ZNF568, LMO2* and one inter-genic loci) where difference of methylation is more pronounced only if the tumor is MSI and on the right side of the colon. **Conclusions:** In this genome-wide methylation study of CRC, we identified a number of genes that are hypermethylated in MSI compared to MSS tumors suggesting an association between DNA methylation and MSI status in CRC.

398F

DNA methylation landscape of sporadic and melanoma-prone patients. *A.C.V. Krepischi[1], E.S.S. de Araújo[2],R.F. Ramalho[2], M.I. Achatz[3], L.F. Moredo[4], J.P. Duprat[4], C. Rosenberg[1], D.M. Carraro[2].* 1) Institute of Biosciences, University of São Paulo, São Paulo, SP, Brazil; 2) International Research Center, A. C. Camargo Cancer Center, Sao Paulo, Brazil; 3) Department of Oncogenetics, A. C. Camargo Cancer Center, Sao Paulo, Brazil; 4) Department of Skin Cancer, A. C. Camargo Cancer Center, Sao Paulo, Brazil.

Melanoma is the most deadly form of skin cancer. Around 10% of melanoma cases are hereditary, and nearly half of them have unknown etiology. Abnormal DNA methylation as either predisposition factor or biomarker has been poorly explored in skin melanomas. Our goal was investigating whether melanoma-prone and sporadic melanoma patients carry distinctive DNA methylation patterns compared to healthy controls. The leukocytes' methylomes from 19 unrelated melanoma-prone and 12 sporadic melanoma patients, all of them negative for germline *CDKN2A* mutations, were obtained using the HM450K platform (Illumina). Methylation data was compared with a control group of 11 healthy controls. Bias adjustment and normalization were achieved by BMIQ, and differential methylation analysis was processed using the RnBeads package. Genes associated with the top 100 differentially methylated sites (significant p-value and higher methylation differences) were annotated by DAVID and cBioPortal tools. Melanoma-prone patients exhibited DNA methylation differences compared to controls mainly at CpG islands, whereas sporadic melanoma patients presented differential methylation at genic sequences located outside promoters and islands. Seventy-four genes were found to be differentially methylated mainly at promoters regions (58%) in melanoma-prone patients. On the other hand, sporadic melanoma group displayed 72 genes mostly hypermethylated (94%) at gene bodies. An *in silico* analysis of each set of genes using melanoma data from TCGA studies revealed that above 50% of all of the genes (53% and 66% for sporadic and melanoma-prone patients, respectively) presented at least 10% of genetic alterations in the tumors. However, only the gene set detected in the sporadic melanoma group showed statistical enrichment for cancer pathways. Our preliminary analysis revealed a distinctive epigenomic profile in leukocytes from sporadic and melanoma-prone patients compared to controls, which were divergent regarding DNA methylation profile and associated genes. Additional studies are necessary to confirm this findings, and to elucidate if such results are related to the melanoma etiology.Funding: FAPESP 2013/07480-8.

399T

Quantification and application of potential epigenetic markers in maternal plasma of pregnancies with hypertensive disorders. *H.J. Kim[1], S.Y. Kim[1], J.H. Lim[1], D.E. Lee[1], D.J. Kim[1], S.Y. Park[1], M.Y. Kim[2], J.Y. Han[2], J.S. Choi[2], H.K. Ahn[2], D.W. Kwak[2], H.M. Ryu[1,2].* 1) Laboratory of Medical Genetics, Cheil General Hospital and Women's Healthcare Center, Seoul, Korea, South Korea; 2) Department of Obstetrics and Gynecology, Cheil General Hospital and Women's Healthcare Center, Dankook University College of Medicine, Seoul, Korea.

ABSTRCTSObjective: To identify novel epigenetic markers that are specifically hypermethylated in placenta and hypomethylated in maternal blood cells, we used methylation profiling array data from the our previous study, and evaluated the quantitative aberrations of candidate marker in maternal plasma of pregnancies with hypertensive disorders. **Methods:** Selected hypermethylated regions by methylation profiling analysis were validated by bisulfite direct sequencing. Using real-time PCR, we compared the concentrations of *DSCR3, RASSF1A and SRY* as cell-free fetal DNA markers and *HYP2 and GAPDH* as cell-free total DNA markers in 72 normal pregnancies, 16 early-onset preeclampsia (EO-PE), 47 late-onset preeclampsia (LO-PE) and 29 gestational hypertension (GH) at 6 - 42 gestational weeks .**Results:** Among 7 identified regions, the most differentially hypermethylated *DSCR3* was selected for potential fetal-specific epigenetic marker. The concentrations of all target sequences were significantly correlated with gestational age at sampling in the control group ($P < 0.001$ for all target sequences). A strong positive correlation was observed between *DSCR3* and *SRY* ($r = 0.369$, $P = 0.006$), *DSCR3* and *RASSF1A* ($r = 0.696$, $P < 0.001$), *DSCR3* and *HYP2* ($r = 0.842$, $P < 0.001$), *RASSF1A* and *HYP2* ($r = 0.598$, $P < 0.001$). In the EO-PE, *DSCR3* and *RASSF1A* concentrations were significantly higher at 24-32 weeks onwards ($P < 0.05$ for both). Moreover, *HYP2* concentration was significantly elevated from 15 weeks of gestation compared with control group. In the LO-PE, *DSCR3, RASSF1A* and *HYP2* concentrations were significantly higher only at 33-42 weeks compared with control group. Concentrations of all markers in GH group were not significantly different from control group. There were no statistical differences among the patients groups.**Conclusions;** Along with the well-known *RASSF1A, DSCR3* and *HYP2* quantification is a promising epigenetic markers for prediction of PE, especially for the early prediction of EO-PE.

400F

Abnormal DNA methylation in T-lymphocytes from patients with CF. *E. Kvaratskhelia[1], N. Dabrundashvili[1], E. Maisuradze[1], M. Gagua[1], T. Topuria[1], L. Margvelashvili[2], E. Abzianidze[3].* 1) Tbilisi State Medical University, Institute of Medical Biotechnology, Tbilisi, Georgia; 2) Hospital 'New Life', Department of Medical Genetics and Diagnostics, Tbilisi, Georgia; 3) Tbilisi State Medical University, Department of Molecular and Medical Genetics, Tbilisi, Georgia.

Cystic fibrosis (CF) is caused by the mutations of the CFTR gene. T-lymphocytes from patients with CF are characterized by reduced expression of anti-inflammatory cytokines and elevate level of pro-inflammatory cytokines. Although the mechanisms responsible for cytokine dysregulation are not well defined, epigenetic factors such as DNA methylation or histone modification might contribute.In the previous work we have found that T-lymphocytes from CF patients express elevated level of DNMT3a, while no significant changes was observed at levels of DNMT1 compared to healthy subjects. In presence of the high dose of 5-AzaC the level of DNMT3a was decreased which was accompanied by an increased amount of IL-10.In this study we analyzed the DNA global methylation levels in T-Lymphocytes derived from CF subjects. Peripheral blood was obtained by venepuncture from CF and healthy subjects. The CF patients were in stable clinical condition and were not receiving systemic corticosteroids. The written informed consent was obtained from the parents of all patients. PBMC were isolated by Ficoll-Paque-1077 (Sigma-Aldrich, Germany) density gradient centrifugation and washed with PBS twice. CD4+ T cells were isolated from PBMC using the CD4+ T Cell Isolation Kit (Miltenyi Biotec GmbH). Cells were activated with 20ng/ml PMA and 250 ng/ml ionomycin . At 48 h after incubation genomic DNA was isolated using QIAamp Mini kit (QIAGEN). Global DNA methylation was measured using the Methylated DNA Quantification Kit as described by the manufacturer (abcam). This kit measures the5-methylcytosine (5-mC)content. The degree of DNA methylation determined was based on the obtained OD readings.Our data showed that DNA methylation levels were significantly reduced in stimulated CF T-cells compared with controls. We also analyzed the relationship between global DNA methylation status and expression ofDNMTsin CD4+T cells obtained from CF individuals. This analysis demonstrated that DNMT3alevel positively correlated with the overall methylation levels in 5-AzaCtreated CD4+T cells, while there was no significant correlation betweenDNMT1 levels and global methylation status in CF CD4+T cells compared with healthy subjects. In summary, data presented in this report demonstrates that abnormal DNA methylation in CD4+ T cells from CF individuals likely plays the significant role in the aberrant immune response during Cystic Fibrosis.

401T

Early prediction of hypertensive disorders of pregnancy using maternal characteristics, epigenetic markers and serum markers. *S.Y. Kim[1], H.J. Kim[1], D.E. Lee[1], J.H. Lim[1], S.Y. Park[1], M.Y. Kim[2], J.Y. Han[2], D.W. Kwak[2], S.Y. Park[2], H.M. Ryu[1,2].* 1) Laboratory of Medical Genetics, Medical Research Institute, Cheil General Hospital and Women's Healthcare Center, Seoul, Korea; 2) Department of Obstetrics and Gynecology, Cheil General Hospital and Women's Healthcare Center, Dankook University College of Medicine, Seoul, Korea.

Objective: To investigate the maternal plasma levels of cell-free fetal DNA and cell-free total DNA throughout gestation using tissue-specific epigenetic markers in pregnancies with hypertensive disorders and healthy pregnancies and to evaluate the predictive value of these epigenetic markers, maternal characteristics and serum markers for developinghypertensive disorders of pregnancy.Methods: A nested case-control study was conducted with maternal plasma throughout gestation collected from 16 early-onset preeclampsia (EO-PE), 47 late-onset preeclampsia (LO-PE), 29 gestational hypertension (GH) and 72 uncomplicated pregnancies. All women underwent routine first- and second-trimester prenatal serum screening at the Cheil General Hospital. We performed a real-time PCR to quantify levels of *RASSF1A*, *DSCR3* and *SRY* as cell-free fetal DNA markers and *HYP2* and *GADH* as cell-free total DNA markers.Results: Significantly positive correlation between level of *RASSF1A, DSCR3* or *HYP2* and gestational age at sampling was observed in all study groups ($P < 0.005$ for all). There were also significant positive correlations between *RASSF1A* and *HYP2, DSCR3* and *HYP2, SRY and RASSF1A* as well as *SRY* and *DSCR3* in PE, GH, all case and control groups ($P < 0.05$ for all). The level of *HYP2* was significantly elevated in both PE and all case groups compared with control group from 6-14 weeks of gestation. At 15-23 weeks and 24-32 weeks, *HYP2* level was a significantly different between EO-PE and controls, PE and controls as well as all cases and controls ($P < 0.05$ for all). At 33-41 weeks, the levels of *RASSF1A, DSCR3* and *HYP2* were significantly elevated in EO-PE, LO-PE, PE and all cases with controls ($P < 0.05$ for all). The final parameters for the prediction of pregnancies with hypertensive disorders included *RASSF1A*, *DSCR3*, *HYP2* at 15-23 weeks' gestation, PAPP-A, AFP, uE_3, β-hCG, InhA, maternal age, and body mass index. ROC curve analysis showed that *HYP2* appeared the most discriminative marker for EO-PE [AUC 0.98 (95% CI 0.86-1.00)], PE [AUC 0.71 (95% CI 0.56-0.83)] and all cases [AUC 0.67 (95% CI 0.57-0.76)], followed by the first trimester PAPP-A for EO-PE [AUC 0.81 (95% CI 0.69-0.90)], PE [AUC 0.69 (95% CI 0.58-0.78)] and all cases [AUC 0.66 (95% CI 0.53-0.78)]. Conclusion: The addition of *HYP2* gene as a cell-free total DNA marker to PAPP-A may be useful for the early prediction of PE, especially for EO-PE.

402F

Missing data imputation using genome-wide DNA methylation data. *W. Guan[1], C. Chong[1], B. Wu[1], Y. Li[2], J.S. Pankow[1], E.W. Demerath[1], J. Bressler[3], M. Fornage[3], M.L. Grove[3], E. Boerwinkle[3].* 1) University of Minnesota, Minneapolis, MN; 2) University of North Carolina, Chapel Hill, NC; 3) University of Texas Health Science Center at Houston, Houston, TX.

DNA methylation is a widely studied epigenetic mechanism. Alterations in methylation patterns may be involved in the development of common diseases. Unlike inherited changes in genetic sequence, variation in site-specific methylation varies by tissue, developmental stage, disease status, and may be impacted by aging and exposure to environmental factors such as diet or smoking. While these wide-range correlations pose analytical challenges in epigenome-wide association studies (EWAS), including reverse causality and confounding by non-genetic factors, it brings opportunities to impute missing values for phenotypes such as blood cell counts using rich methylation data such as that provided by the Illumina Infinium HumanMethylation450 (HM450) BeadChip. We evaluated two statistical methods for missing data imputation: 1) a projection-based method (Houseman et al., 2012); and 2) a variable selection based multiple imputation method. Through simulations, we showed that both methods can properly control false positive rate, but the projection-based method may lead to biased estimates in subsequent association testing. On the other hand, the multiple-imputation method can lose power when the missing rate is high. We illustrated the methods with data from the Atherosclerosis Risk in Communities (ARIC) study to carry out an EWAS between methylation levels and smoking status, in which missing cell type compositions and white blood cell counts were imputed.

403T

A genome-wide study of DNA methylation and prediction of diabetic nephropathy. *G.D. Fufaa[1], R.L. Hanson[1], S. Kobes[1], R.G. Nelson[1], K. Susztak[2].* 1) NIDD/NIH, Phoenix, AZ; 2) University of Pennsylvania, Philadelphia, PA.

Cross-sectional studies suggest that DNA methylation at some CpG sites differs between those with and without diabetic nephropathy, but there are few longitudinal studies that support these observations. We examined the role of DNA methylation in prediction of diabetic nephropathy in a longitudinal study of American Indians. We designed two nested case-control samples for different stages of kidney disease: one involving cases who developed severe albuminuria (SA) (urinary albumin:creatinine ratio ≥300 mg/g) among individuals who initially did not have SA (167 cases, 195 controls), and one involving cases who developed end-stage renal disease (ESRD) among those who initially had SA or estimated glomerular filtration rate <60 ml/min (82 cases, 101 controls). Genome-wide DNA methylation was measured in peripheral blood leukocytes on the Illumina 450 Infinium array. After exclusion of probes containing SNPs and restriction to probes which were highly reproducible (false discovery rate [FDR]<0.05 for the correlation among duplicates), 77,110 CpG sites were tested for association with nephropathy. Conditional logistic regression was used to estimate the hazard ratio (HR) associated with a 1 standard deviation difference in methylation at each site, controlled for age, sex, diabetes duration and the first 4 principal components. P-values for SA and ESRD were combined for a global test of association with development of nephropathy. No site achieved a FDR <0.05, but several suggestive signals (FDR<0.50) were observed. The strongest association with SA occurred at a site near *ZC3H12D* on chromosome 6 (HR=0.51, $P=3.2\times10^{-6}$), while the strongest association with ESRD occurred near *PTPRD* on chromosome 20 (HR=0.41, $P=2.8\times10^{-6}$). The strongest combined association was near *ALHD1A3* on chromosome 15 ($P=1.9\times10^{-6}$; HR=0.60 for SA, HR=0.66 for ESRD). Analysis of the top 1% of differentially methylated sites showed that they were more often hypomethylated in those who progressed to diabetic nephropathy compared with the remaining 99% of sites ($P=7.4\times10^{-6}$ for SA, $P=1.3\times10^{-3}$ for ESRD, $P=1.1\times10^{-4}$ for combined). Pathway analyses using DAVID showed that the top 1% of probes for the combined analysis were significantly enriched for genes involved in focal adhesion and in T-cell receptor signaling. These studies identify several CpG sites and pathways at which DNA methylation (particularly hypomethylation) may predict development of diabetic nephropathy.

404F

Comparison of DNA methylation profiles in sib-pairs discordant for intrauterine exposure to maternal gestational diabetes mellitus. *S. Kwak[1], E. Kim[2], S. Choi[3], K. Park[1], J. Sung[2], H. Jang[3].* 1) Department of Internal Medicine, Seoul National University Hospital, Seoul, South Korea; 2) Department of Epidemiology, School of Public Health, Seoul National University, Seoul, South Korea; 3) Department of Internal Medicine, Seoul National University Bundang Hospital, Gyeonggi-do, South Korea.

Offspring of women with a history of gestational diabetes mellitus (GDM) is at significantly increased risk for development of obesity and diabetes. It is suggested that intrauterine hyperglycemia induce epigenetic change in fetus which might have detrimental effect on future metabolic phenotypes. In this study, we compared pair-wise DNA methylation status between siblings whose intrauterine exposure to maternal GDM are discordant to each other. A total of 19 sib-pairs born from 18 Korean women who experienced both normal pregnancy and GDM pregnancy were included. DNA was extracted from peripheral leukocytes when the offspring were at age between 4 and 15 years. An epigenome-wide association study was conducted using Illumina Infinium HumanMethylation 450 BeadChip assays. Differential methylation was compared within each sib-pairs discordant for GDM pregnancy adjusting for their age, and sex. In the unsupervised clustering based on Manhattan distance, 33 samples were closely related to intrauterine GDM status. Children of GDM pregnancy were generally younger, leaner and smaller than their counterparts, but the differences in anthropometrics disappeared when quartile growth at ages were compared. A total of six CpG sites were differential methylated within sib-pairs with false discovery rate of less than 0.1 ($P < 1.50 \times 10^{-6}$). The six sites were located in *MFHAS1*, *LOC92973*, *PACRG*, *HNF4A*, *PITPNM3*, and *RREB1*. Among these CpG sites, cg08407434 which is located at *HNF4A* was consistently hypermethylated in offspring of GDM with mean pairwise difference methylation of 1.3% ($P = 9.10 \times 10^{-7}$). Using Ingenuity pathway analysis of differentially methylated regions, we found that immune response was overrepresented by hypermethylated genes. To the best of our knowledge, this is the first study to investigate the epigenome-wide difference in methylation within sib-pairs discordant for intrauterine hyperglycemia. We found several suggestive CpG sites with differential methylation, including the one located in *HNF4A* which warrants further investigation.

405T

DNA methylation may mediate the association of cigarette smoking and prostate cancer progression. *I.M. Shui[1,2], R. Rubicz[1], M.S. Geybels[1], J.L. Wright[1], S. Kolb[1], S. Zhao[3], M. Bibikova[4], J. Fan[4], B. Klotzle[4], E.A. Ostrander[5], Z. Feng[6], J.L. Stanford[1].* 1) Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Harvard TH Chan School of Public Health, Boston, MA; 3) National Institute of Environmental Health Sciences, Research Triangle Park, NC; 4) Illumina, San Diego, CA; 5) National Human Genome Research Institute, Bethesda, MD; 6) MD Anderson Cancer Center, Houston, TX.

Background: Smoking is associated with prostate cancer progression and DNA methylation could mediate this relationship. Aims of this study were to assess if a) smoking at the time of diagnosis is associated with disease recurrence following radical prostatectomy, b) smoking is associated with differential methylation profiles in prostate tumor tissue, and c) the smoking-related differential methylation is also associated with disease recurrence.Methods: Men with prostate cancer were identified from 2 population-based case-control studies. We included 799 men who had radical prostatectomy as primary therapy. Of these men, 523 had Illumina 450K methylation data from tumor DNA. Cox proportional hazards models evaluated the association between smoking and prostate cancer recurrence. Multivariable linear regression with empirical Bayes shrinkage (*limma*) assessed differential methylation by individual CpG positions (DMPs) and *Probe Lasso* was used to identify differentially methylated regions (DMRs). DMPs identified in Aim b were carried forward to test their association with prostate cancer recurrence.Results: For the primary analysis, men were categorized as never (43%), former (47%), or current (10%) smokers. Current smoking vs. never smoking was associated with a significantly increased risk of fatal prostate cancer (HR: 4.55, 95% CI: 2.10, 9.83, p=0.0001). There was substantial differential methylation when comparing current vs. never smokers. 33194 CpG sites had an FDR q<0.05 and ≥5% difference in beta value; 2098 had ≥10% difference. Moreover, 71% of the 33194 CpGs associated with smoking were also associated with fatal prostate cancer (q<0.05 and ≥5% difference in beta value); 28% had ≥10% difference. Top DMPs and DMRs associated with recurrence included regions known to be imprinted and those associated with bone disorders, cancer susceptibility, cholesterol metabolism, and immune function. There was no evidence of differential methylation between former and never smokers, suggesting that methylation changes may be reversible upon smoking cessation. Conclusions: Our results provide evidence that DNA methylation could be an important mediator for smoking and prostate cancer recurrence. Importantly, former smokers had a reduced risk of prostate cancer recurrence and a methylation profile similar to never smokers. Future steps will be to identify specific biologic mechanisms relevant to the smoking-methylation-prostate cancer relationship.

406F

DNA Methylation at residue cg05575921 in the aryl hydrocarbon receptor repressor is the most sensitive and specific indicator of smoking status in epigenome. *A. Andersen[1], N. Hollenbeck[1], E. Andersen[2], T. Osborn[2], M. Gerrard[3], F. Gibbons[3], K. Wang[4], R. Philibert[1,2].* 1) Psychiatry, University of Iowa, Iowa City, IA; 2) Behavioral Diagnostics, Iowa City, IA; 3) Department of Psychology, Center for Health Intervention and Prevention, University of Connecticut, Storrs, CT; 4) Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, IA.

Smoking is the largest preventable cause of morbidity and mortality in the world. Despite the development of numerous preventive interventions and public policy changes in the last 50 years, one in five US adults is a smoker. The effectiveness of ehavioral and pharmacologic interventions to aid smoking cessation have been limited by lack of early identification of smoking cases, when interventions are most effective. Unfortunately, the validity of self-report measures of smoking have been shown to be limited, particularly in younger smokers. Biomarkers for smoking such as cotinine and carbon monoxide have significant limitations in this group, particularly when smoking is intermittent. Furthermore, both e-cigarettes and nicotine replacement therapies may confound cotinine measurement as a sensitive and specific indicator of smoking. Epigenetic measurement has the potential to overcome the limitations of these approaches and guide early intervention, before smoking behavior has become ingrained and health consequences have emerged. We have previously demonstrated that smoking is associated with genome-wide changes in DNA methylation. Here, we determine the sensitivity and specificity of ten of the most consistently replicated CpG loci with respect to smoking status using data from a publically available dataset. We show that methylation at many of the more commonly cited loci, such as in F2RL3, display ethnically contextual effects that render them unacceptable for general usage. We further show that methylation status at a CpG locus in the aryl hydrocarbon receptor repressor, cg05575921, is both sensitive and specific for smoking status in adults of all ethnicities, with a receiver operated curve characteristic (ROC) area under the curve of 0.99. Modeling analyses of cg05575921 consumption show that demethylation at this locus is a function of both average recent consumption and total lifetime consumption. The locus is responsive to smoking over short periods of time with notable changes evident in less than 4 weeks. Given recent demonstrations that methylation at this locus reflects both intensity of smoking and the degree of smoking cessation, we conclude that a methylation-based assay at this locus could have a prominent role in understanding the impact of new products, such as e-cigarettes on initiation of cigarette smoking among adolescents, while improving the prevention and treatment of smoking and smoking related disorders.

407T

Estimation of Cell Type Specific DNA Methylation Effects using Whole Blood Methylation Data. *R. Barfield[1], A. Bacarelli[2], X. Lin[2].* 1) Harvard University, Cambridge, MA; 2) Harvard T.H. Chan School of Public Health, Boston, MA.

Association analysis of DNA methylation data is challenged by cell type heterogeneity, as DNA methylation collected in studies is typically a mixture of different cell types. For example, studies using whole blood measure DNA methylation from a mixture of different lymphocytes. This cell type heterogeneity can potentially bias results, as DNA methylation is known to be an important mechanism of tissue differentiation, and thus may differ by cells. Current methods to adjust for this involve including observed or estimated cell type counts as additional covariates in the analysis. These methods however do not estimate the exposure effects on cell type specific DNA methylations. Direct measurements of cell type specific methylations involve intensive lab work and are costly. We develop in this paper a statistical method to estimate exposure effects on cell-specific methylations using whole blood methylation data when cell type counts are available but cell-specific methylations are unavailable. Specifically, we assume cell type specific regression models of the exposure effects on cell type specific methylations. We treat cell specific methylations as missing data, and develop an EM algorithm to estimate the exposure effects on cell type specific methylations adjusting for covariates by using whole blood methylation data and cell type counts. To the best of our knowledge, this is the first method to estimate these effects without assaying each cell type. We applied our method to the Normative Aging Study to study the smoking effect on cell type specific methylations.

408F

Genome-Wide DNA Methylation Signatures of Salivary Gland Inflammation in Sjögren's Syndrome. *M.B. Cole[1], D. Quach[2], H. Quach[2], L.F. Barcellos[2], L.A. Criswell[3].* 1) Department of Physics, University of California, Berkeley, Berkeley, CA; 2) Division of Epidemiology, Genetic Epidemiology and Genomics Laboratory, School of Public Health, University of California, Berkeley, Berkeley, CA; 3) Rosalind Russell / Ephraim P. Engleman Rheumatology Research Center, Department of Medicine, University of California, San Francisco, CA.

Sjögren's Syndrome (SS, OMIM #270150) is a chronic, multi-system autoimmune disease characterized by progressive destruction of the exocrine glands, with subsequent mucosal and conjunctival dryness. A growing body of evidence indicates that many epigenetic changes are associated with disease status and that epigenetic marks can provide unique insights into complex disease mechanisms. We report on results of a case-control study of DNA methylation (DNAm) differences within labial salivary gland tissue, using biopsies sampled from 13 severe SS cases and 13 controls in the Sjögren's International Collaborative Clinical Alliance (SICCA; http://sicca.ucsf.edu/; HHSN268201300057C) registry. We have methylotyped gland tissue and sorted PBMCs from these subjects using the Illumina HumanMethylation450 (450k) BeadChip platform. In addition to standard background correction and normalization techniques, we applied an adaptive normalization scheme to adjust for residual associations between DNAm and control measures. Principal component analysis (PCA), applied to the 404,353 CpG sites passing strict QC criteria, clearly distinguishes glands of primary SS cases from those of controls. We find over 10,000 significant differentially methylated positions (DMPs) that show significant overlap with the promoters of genetic risk loci as a whole, with particular enrichment seen in the promoters of *CXCR5* and *BLK*. We also observe an extended region of differential methylation surrounding *PSMB8* and *TAP1* in the MHC class II region. Despite a previous report of association between global DNAm and ICAM1expression in the context of SS, we fail to detect differential methylation in the promoter of the *ICAM1* locus. On the other hand, differential methylation patterns surrounding several non-coding RNAs suggest many indirect consequences of the observed salivary gland *hypo*-methylation in SS. Transcription factor motif enrichment analysis highlights the specific nature of these methylation differences, demonstrating co-localization of DMPs with interferon-stimulated response element (ISRE) and PU-Box motifs. DNAm signatures derived from sorted PBMCs show that disease-associated changes in DNAm are linearly correlated with cell-type specific patterns, resolving differential lymphocyte proportions. Our results emphasize the utility of CpG methylation not only as a biomarker of disease status, but also as an independent probe of underlying disease processes.

409T

DNA methylation profiles in ADHD: comparison between boys and girls. *T.V.M.M. Costa[1], C. Milani[1], F.A. Marchi[2], G.M. Novo-Filho[1], E.A. Zanardo[1], M.M. Montenegro[1], F.A.R. Madia[1], R.L. Dutra[1], F.B. Piazzon[1], A.M. Nascimento[1], M. Rocha[1,3], M.B. Veronesi[1], C.A. Kim[4], V. Schuch[5], C.B. Mello[5], M. Muszkat[5], L.D. Kulikowski[1,3,4].* 1) Laboratório de Citogenômica, Departamento de Patologia, Faculdade de Medicina da USP, São Paulo, Brasil; 2) Laboratório NeoGene, AC Camargo Cancer Center, São Paulo, Brazil; 3) Centro de Reprodução Humana e Genética,Departamento de Saúde Coletiva, Faculdade de Medicina do ABC, Santo André, Brasil; 4) Unidade de Genética, Departamento de Pediatria, Instituto da Criança, HC/FMUSP, São Paulo, Brasil; 5) Núcleo de Atendimento Neuropsicológico Infantil Interdisciplinar (NANI), Departamento de Psicobiologia da Escola Paulista de Medicina, UNIFESP, São Paulo, Brasil.

Attention deficit hyperactivity disorder (ADHD) is one of the most common neurobehavioral disorder diagnosed in childhood. ADHD is etiologically heterogeneous. On the other hand, it has a high heritability (around 80%) and it is diagnosed twice more often in boys than in girls. Although it has a high heritability, the genetic architecture of ADHD is still largely unknown and molecular markers for diagnosis havenotbeen identified yet. DNA methylation is an important epigenetic mechanism associated with silencing of genes in CpG islands. Thus, the investigation of methylation profile in patients with ADHD is very important and could reveal different and interesting aspects of the disorder. DNA methylation profiles were performed using DNA extracted from blood lymphocytes of 13 carrying ADHD (9 boys and 4 girls, ages 08-15) by Illumina Infinium HumanMethylation450 BeadChip. Children were diagnosed according to Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR). We pre-processed the array data using GenomeStudio software and full analysis was performed using specific packages within the environment R. The comparison between boys and girls considered delta-beta ($\Delta\beta$) <-0.20 or > +0.20. The probes were annotated according to the data provided by Illumina, using the genome hg19 reference. After the comparison between boys and girls we obtained 20 differentially hipomethylated probes and 20 differentially hipermethylated probes. Our previous results showed significantly methylated probes involved in several networks, including Glutamate pathways, pointing out the relevance of glutamate, which is closely associated to the Central Nervous System and could be modified by epigenetic factors. In this sense, tissue-specific DNA methylation profiles considering gender may provide novel insights into pathogenic mechanisms of ADHD and also help in future epigenetic therapies.Grants: FAPESP: 14/02565-8 and FINEP-CT INFRA 0160/12 SP8.

410F

A genome-wide analysis of differential methylation regions in the nucleus accumbens of rhesus macaques after long-term alcohol use. *B. Ferguson, R. Cervera-Juanes, L. Wilhelm, S. Gonzales, K. Grant.* Div. of Neurosciences, ONPRC, Oregon Health Science Univ., Portland, OR.

The long-lasting behavioral plasticity associated with alcohol addiction may be related to changes in epigenetic mechanisms that regulate neural gene expression, contributing to the widespread gene expression changes observed after alcohol use. The nucleus accumbens (NAc), a critical component of the reward system, is associated with risk for alcohol use. Here, we investigated the long-term effects of alcohol on the NAc methylome of male rhesus macaques after 12-months of alcohol or water (controls) self-administration. We used targeted bisulfite DNA sequencing to identify significant differentially methylated regions (DMRs) between controls (C), low-moderate drinkers (LMD) and heavy drinkers (HD) using Comb-p software. Our results revealed 51 significant, gene-specific DMRs among the three groups. Interestingly, some of the DMRs were uniquely detected in either the HD or LMD groups (39), while others appeared to be correlated with alcohol dose (16). The location of the DMRs in the promoter and within the gene body, together with the preliminary gene expression data suggest a potential role of the DMRs in regulating total transcript level and the expression of alternative transcripts. The DMRs map to genes encoding proteins with functions linked to ion transport, cytoskelton binding, cell surface recptors and transcriptional regulation. The majority of the genes associated with DMRs are involved in neuronal functions, suggesting that DNA methylation may play a central role in the neural adaptation to chronic alcohol use.

411T

Intra-individual dynamics of transcriptome and genome-wide stability of DNA methylation during three months. *R. Furukawa[1], T. Hachiya[1], H. Ohmomo[1], Y. Shiwa[1], K. Ono[1], S. Suzuki[2], M. Satoh[1], J. Hitomi[1], K. Sobue[1], A. Shimizu[1].* 1) Iwate Medical University, Shiwa-gun,Iwate, Japan; 2) Keio University,Shinjuku-ku, Tokyo,Japan.

Cytosine methylation in CpG dinucleotides is one of the key epigenetic mechanisms that regulate gene expression. DNA methylation is considered to be stable, and to define "fields" for producing the transcriptomes that characterize the functional differences of cells. Although gene expression patterns are well known to change in response to various stimuli, it remains unclear to what extent DNA methylation contributes to short-term dynamics in gene expression. In the present study, we examined the short-time dynamics and stability of gene expression and DNA methylation in peripheral blood mononuclear cells (PBMCs) and monocytes. We collected blood 24 times from one healthy volunteer in three months and obtained PBMCs and $CD14^{high}CD16^{low}$ monocytes using cell sorter (SH800, Sony). After extraction DNA and RNA, we measured gene expression and DNA methylation levels by RNA sequencing and by Illumina Infinium HumanMethylation450 BeadChip microarrays, respectively. First, to extract genes showing dynamic change in expression, we calculated the coefficient of variation for each gene as an indicator of the expression change, and defined the dynamic expression genes by detecting outliers from a baseline. From these dynamic genes, we removed the genes that the expressions were associated with the change in cell-type composition by using a linear regression model and the ANOVA test to detect Cell-type composition/Expression Associations (CEA). As results, we obtained 1,637 and 5,279 dynamic non-CEA genes in PBMCs and monocytes, respectively. Next, we assessed to what extent the DNA methylation levels explain the expression of the dynamic non-CEA genes. As an index of the contribution of DNA methylation to the gene expression, R^2 values were obtained by a linear regression model. Unexpectedly, there was no significant difference between the distributions of R^2 values to the dynamic and the non-dynamic non-CEA genes, meaning that the contribution of DNA methylation to the dynamics of non-CEA genes was not significant. Additionally, the standard deviation of the methylation levels of CpG loci near the dynamic genes was less than that of CpG loci located near the non-dynamic genes. From these results, it was considered that the DNA methylation levels of the CpG that located the vicinity of the dynamic non-CEA genes were quite stable, and the extent of the contribution of the DNA methylation to the dynamics of gene expression was small at least in several months.

412F

Characterizing a genomic map of 5-hydroxymethylcytosine in human brain at single base resolution through next-generation sequencing. *J.A. Gross[1], A. Pacis[2,3], G.G. Chen[1], L.B. Barreiro[2,3], C. Ernst[1], G. Turecki[1].* 1) Douglas Mental Health Univ Inst, Montreal, Quebec, Canada; 2) CHU Sainte-Justine Research Centre, Department of Genetics, Montreal, Quebec, Canada; 3) Departments of Biochemistry and Pediatrics, University of Montreal, Montreal, Quebec, Canada.

The recent discovery that methylated cytosines are converted to 5-hydroxymethylated cytosines (5hmC) by the family of ten-eleven translocation enzymes has sparked significant interest on the genomic location, the abundance in different tissues, the putative functions, and the stability of this epigenetic mark. 5hmC plays a key role in the brain, where it is particularly abundant and dynamic during development. Using AbaSI-Seq, we comprehensively characterize 5hmC in the prefrontal cortices of 24 subjects. We show that, although there is inter-individual variability in 5hmC content among unrelated individuals, approximately 8% of all CpGs on autosomal chromosomes contain 5hmC, while sex chromosomes contain far less. Our data also provide evidence suggesting that 5hmC has transcriptional regulatory properties, as the density of 5hmC was highest in enhancer regions and within exons. Furthermore, we link increased 5hmC density to histone modification binding sites, to the gene bodies of actively transcribed genes, and to exon-intron boundaries. Finally, we provide several genomic regions of interest that contain gender-specific 5hmC. Collectively, these results present an important reference for the growing number of studies that are interested in the investigation of the role of 5hmC in brain and mental disorders. Understanding the differences in the genomic locations of hmC will shed light on the growing debate as to whether hmC represents a novel epigenetic mark or whether it is simply an intermediate product of active DNA demethylation.

413T

Strong components of epigenetic memory in cultured human fibroblasts. *N.A. Ivanov[1], R. Tao[1], J.G. Chenoweth[1], A. Brandtjen[1], M.I. Mighdoll[1], J.D. Genova[1], R.D. McKay[1], Y. Jia[1], D.R. Weinberger[1,2,3,4,5], J.E. Kleinman[1], T.M. Hyde[1,5,6], A.E. Jaffe[1,7,8,9].* 1) Lieber Institute for Brain Development, Baltimore, MD; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD; 4) Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD; 5) Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD; 6) Department of Biological Sciences, Johns Hopkins School of Medicine, Baltimore, MD; 7) Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore MD; 8) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore MD; 9) Center for Computational Biology, Johns Hopkins University, Baltimore MD.

Differentiation of pluripotent cells from fibroblast progenitors are potentially transformative tools in personalized medicine. We previously described culture of scalp- and dura-derived fibroblasts from postmortem tissue, and identified relatively greater success culturing the dura-derived cell lines. We hypothesized that these differences in growth in culture were related to epigenetic differences between the cultured fibroblasts by sampling location, and therefore generated genome-wide DNA methylation and transcriptome data on 11 intrinsically matched pairs of dural and scalp fibroblasts from donors across the lifespan (infant to 85 years). While these cultured fibroblasts were several generations removed from the primary tissue and morphologically indistinguishable, we found widespread epigenetic differences by sampling location at the single CpG (N=101,989), region (N=697), "block" (N=243), and global spatial scales suggesting a strong epigenetic memory of original fibroblast location. Furthermore, many of these epigenetic differences manifested as differences in gene expression levels, particularly at the region-level (69.4%). We further identified 7,265 CpGs and 10 regions showing significant epigenetic memory related to the age of the donor, as well as an overall increased epigenetic variability, preferentially in scalp-derived fibroblasts (83% of loci were more variable in scalp). This was hypothesized to result from cumulative exposure to environmental stimuli in the primary tissue. By integrating publicly available DNA methylation datasets on individual cell populations in blood and brain, we identified significantly increased inter-individual variability in our scalp- and other skin-derived fibroblasts. This increased inter-individual variability is on a similar scale as epigenetic differences between different lineages of blood cells. Lastly, these epigenetic differences did not appear to be driven by somatic mutation - while we identified 63 probable variants across the 11 subjects, there was no association between mutation burden and age of the donor (p=0.86). These results depict a strong component of epigenetic memory in cell culture from primary tissue, even after several generations of daughter cells, related to cell state and donor age.

414F

Epigenome Characterization of Human Genomes using the PacBio® Platform. *J. Korlach.* Pacific Biosciences, Menlo Park, CA.

In addition to the genome and transcriptome, epigenetic information is essential to understand biological processes and their regulation, and their misregulation underlying disease. Traditionally, epigenetic DNA modifications are detected using upfront sample preparation steps such as bisulfite conversion, followed by sequencing. Bisulfite sequencing has provided a wealth of knowledge about human epigenetics, however it does not access the entire genome due to limitations in read length and GC-bias of the sequencing technologies used.In contrast, Single Molecule, Real-Time (SMRT®) DNA Sequencing is unique in that it can detect DNA base modifications as part of the sequencing process. It can thereby leverage the long read lengths and lack of GC bias for more comprehensive views of the human epigenome. I will highlight several examples of this capability towards the generation of new biological insights, including the resolution of methylation states in repetitive and GC-rich regions of the genome, and large-scale changes in the methylation status across a cancer genome as a function of drug sensitivity.

415T

M-QTL analysis between asthma GWAS loci and DNA methylation interactions. *A. Kumar[1,2,3], S.K. Merid[1], C. Söderhäll[1,7], C. Xu[4,5], O. Gruzieva[1], J. Kere[6,7], G. Pershagen[1], G. Koppelman[4,5], E. Melén[1,7,8].* 1) Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; 2) Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland; 3) University of Basel, Basel, Switzerland; 4) University of Groningen, University Medical Center Groningen, Department of Pulmonology, GRIAC research Institute, Groningen, the Netherlands; 5) University of Groningen, University Medical Center Groningen, Department of Genetics, GRIAC research Institute, Groningen, the Netherlands; 6) Department of Biosciences and Nutrition and Center for Innovative Medicine (CIMED), Karolinska Institutet, Stockholm, Sweden; 7) Centre for Allergy Research, Karolinska Institutet, Stockholm, Sweden; 8) Sachs' Children's Hospital, Stockholm, Sweden.

Background: Asthma is characterized as a chronic inflammation disease and has increased in prevalence over the decades. The genome-wide association studies (GWAS) have implicated several single nucleotide polymorphisms (SNPs) with varying risk estimates for asthma, but the etiology and prevalence change is still unexplained. Thus, the interaction between genetic and epigenetic variations can help us find SNPs associated with altered DNA methylation (mQTL) to explore the cis and trans-regulatory relationships that may affect the gene regulation and disease phenotypes. **Methods**: Using the Illumina Infinium HumanMethylation450 beadchip, DNA methylation was measured from peripheral blood of 464 individuals (231 patients with asthma-ever doctor diagnosis and 233 controls) up to 8 yrs age from the BAMSE study. We currently focused on cytosine-phosphate-guanine (CpG) sites in the *ORMDL3 and IL1RL1* genes, that have been implicated in multiple asthma GWAS, with genotypes derived on same samples using SNPs from Human610-Quad beadchip imputed (minimac) on 1000 Genomes reference panels (phase 1). "SNP x mQTL" interactions were examined using logistic regression models adjusting for gender, case/control status and population stratification eigenvalues. **Results**: A total of 15 CpG sites were identified in the two loci. Preliminary analyses in cases and controls separately yielded a high overlap in significant results after Bonferroni correction (BC); we thus combined cases and controls to increase power. Consequently, in the *ORMDL3* region, six local CpGs out of nine were significant after BC, with cg10909506 showing the strongest association with rs17608925 ($p=1.05x10^{-15}$) in comparison to the top GWAS SNP, rs7216389 ($p=6.499x10^{-12}$, $R^2=0.17$). Similarly, cg25869196 with rs11676124 ($p=2.78x10^{-9}$) and cg20060108 with rs887971 ($p=1.33x10^{-11}$) were found significant in *IL1RL1* gene region, out of six CpGs, while the relevant top GWAS SNP, rs1420101 was less significant, with no correlation. **Conclusion**: Our results on two well-known asthma susceptibility genes indicate that CpG sites that are significantly associated with SNPs manifest *cis*-mQTLs effects, although some *trans*-mQTLs associations were also found. The strongest mQTL effects were seen for other SNPs than the top GWAS SNPs. Thus, studying these DNA methylation changes as part of integrative genomics pipelinecan help us disentangle the molecular mechanism of the diseases better.

416F

DNA Methylation Score as a Biomarker in Newborns for Sustained Maternal Smoking During Pregnancy. *S.J. London[1], M. Wu[2], S. Zhao[3], S. Haberg[4], P. Ueland[5], R.M. Nilsen[7], O. Midttun[6], S.E. Vollset[4,5], B. Joubert[1], S. Peddada[2], W. Nystad[4], S.E. Reese[1].* 1) Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC; 2) Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 3) Biostatistics and Computational Biology Branch, NIEHS, Research Triangle Park, NC; 4) Norwegian Institute of Public Health, Oslo, Norway; 5) University of Bergen, Bergen, Norway; 6) Bevital AS, Laboratory building, Bergen, Norway; 7) Department of Research and Development, Haukeland University Hospital, Bergen, Norway.

Maternal smoking during pregnancy leads to numerous adverse outcomes in offspring. We previously identified highly reproducible methylation signals using a genome wide methylation platform (Illumina 450K) that reflect sustained, but not transient, smoking during pregnancy. Questionnaire assessment of smoking in pregnant women may not detect all smokers and timing of assessments may be insufficient to assess sustained smoking. There is no biomarker of sustained smoking during pregnancy. We used data on a short-term smoking biomarker, cotinine, measured in maternal plasma during pregnancy and Illumina 450K whole newborn cord blood DNA methylation in a pregnancy cohort (N=1279) to develop a biomarker in newborns of sustained maternal smoking during pregnancy. We used logistic least absolute shrinkage and selection operator (LASSO) regression with area under the curve (AUC) cross-validation to train a model predictive of sustained maternal smoking in pregnancy. A subset of CpGs that we determined to be most predictive of smoking from the LASSO was used to estimate the probability that a newborn's mother was a smoker. The regression coefficients of these CpGs were used to develop a smoking score that reliably predicted smoking status in the train set (N=1058; AUC=0.98 Sensitivity=85%, Specificity=98%). As expected, predictive performance was lower on the much smaller test set (N=221; AUC=0.87, Sensitivity=57%, Specificity=93%). This smoking score can be applied to other studies with Illumina 450K methylation data. It represents a promising novel biomarker in newborns of sustained maternal smoking during pregnancy. This should be useful in studies with limited time course information on maternal as well as to fill in missing individual data. It also provides a means to evaluate self-reported questionnaire data. This quantitative biomarker, which incorporates duration as well as dose, might improve the ability to detect health effects of maternal smoking during pregnancy.

417T

Whole-genome bisulfite sequencing data from multiple human tissues reveal novel CpG island loci of tissue-specific regulation. *I. Mendizabal[1,2], S.V. Yi[1].* 1) School of Biology, Georgia Institute of Technology, Atlanta, GA; 2) Department of Genetics, Physical Anthropology and Animal Physiology, University of the Basque Country UPV/EHU, Spain.

CpG islands (CGIs) are one of the most widely studied regulatory features of the human genome, with critical roles in development and disease. Despite such significance and the original epigenetic definition, currently used CGI sets are typically predicted from DNA sequence characteristics. Although CGIs are deeply implicated in practical analyses of DNA methylation, recent studies have shown that such computational annotations suffer from inaccuracies. Here we used whole genome bisulfite sequencing from ten diverse human tissues to identify a comprehensive, experimentally obtained, single-base resolution CGI catalog. In addition to the unparalleled annotation precision, our method is free from potential bias due to arbitrary sequence features or probe affinity differences. In addition to clarifying substantial false positives in the widely used UCSC annotations, our study identifies numerous novel epigenetic loci of regulation. In particular, we reveal significant impact of transposable elements on the epigenetic regulatory landscape of the human genome, and demonstrate ubiquitous presence of transcription initiation at CGIs, including alternative promoters in gene bodies and non-coding RNAs in intergenic regions. Moreover, coordinated DNA methylation and chromatin modifications mark tissue-specific enhancers at novel CGIs. Enrichment of specific transcription factor binding from ChIP-seq provides mechanistic roles of CGIs on regulation of tissue specific transcription. Consequently, the CGI catalog provides a comprehensive and integrated list of genomic hotspots of epigenetic regulation.

418F

Epigenome-wide association study suggests that SNPs in the promoter region of *RETN* influence plasma resistin level via effects on DNA methylation at neighboring sites. *M. Nakatochi[1], S. Ichihara[2], K. Yamamoto[3], K. Ohnaka[4], Y. Kato[5], S. Yokota[6], A. Hirashiki[7], K. Naruse[8], H. Asano[6], H. Izawa[9], T. Matsubara[8], M. Yokota[10].* 1) Bioinformatics Section, Center for Advanced Medicine and Clinical Research, Nagoya University Hospital, Nagoya, Japan; 2) Graduate School of Regional Innovation Studies, Mie University, Tsu, Japan; 3) Department of Medical Chemistry, School of Medicine, Kurume University, Kurume, Japan; 4) Department of Geriatric Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan; 5) Department of Registered Dietitians, Faculty of Health and Welfare, Tokai Gakuin University, Kakamigahara, Japan; 6) Department of Internal Medicine, Iwakura Hospital, Iwakura, Japan; 7) Department of Advanced Medicine in Cardiopulmonary Disease, Nagoya University Graduate School of Medicine, Nagoya, Japan; 8) Department of Internal Medicine, School of Dentistry, Aichi Gakuin University, Nagoya, Japan; 9) Department of Cardiology, Fujita Health University Banbuntane Hotokukai Hospital, Nagoya, Japan; 10) Department of Genome Science, School of Dentistry, Aichi Gakuin University, Nagoya, Japan.

Aims Human resistin is considered to be a biomarker or mediator of metabolic and inflammatory diseases. To investigate epigenetic regulation of the plasma concentration of resistin, we performed an epigenome-wide association study for this parameter and DNA methylation (DNAm) in an elderly Japanese cohort and then assessed the relation of single nucleotide polymorphisms (SNPs) associated with the plasma resistin concentration to DNAm level at identified sites.*Methods* The association of plasma resistin level with DNAm status was examined in 191 nondiabetic elderly Japanese men in Kita-Nagoya with the Illumina Infinium HumanMethylation450 BeadChip array. The association between DNAm status at specific sites in the flanking region of the resistin gene (*RETN*) and *RETN* mRNA abundance was then evaluated with a public data set for 1202 monocyte samples from a multiethnic cohort. Finally, the association of DNAm status and SNPs in the promoter region of *RETN* was assessed in two cohorts composed of 191 Japanese individuals in Kita-Nagoya and 287 Japanese individuals in Fukuoka.*Results* The plasma resistin level showed a negative genome-wide significant association with DNAm status at a CpG site located in the promoter region of *RETN* ($\beta = -0.270 \pm 0.041$, $p = 6.02 \times 10^{-10}$). Four DNAm sites in the *RETN* promoter region including the CpG site ($\beta = -0.298 \pm 0.016$, $p = 4.23 \times 10^{-70}$) showed a negative genome-wide significant association with *RETN* mRNA abundance in monocytes. Furthermore, DNAm level at the CpG site was negatively associated with the minor alleles of the *RETN* promoter SNPs rs34861192 and rs3219175 (Kita-Nagoya, $\beta = -0.396 \pm 0.043$, $p = 4.43 \times 10^{-17}$; Fukuoka, $\beta = -0.334 \pm 0.035$, $p = 3.50 \times 10^{-19}$). The SNPs rs34861192 and rs3219175 were in complete LD in the Kita-Nagoya samples.*Conclusions* Although the minor alleles of rs34861192 and rs3219175 have been reported to be positively associated with circulating resistin level in Japanese, the mechanism by which they regulate the circulating resistin concentration has remained unclear. Our results suggest that *RETN* promoter SNPs might influence the circulating resistin level through an effect on DNAm at the CpG site and on *RETN* mRNA abundance in monocytes.

419T

Epigenetic Regulation of Differential *HLA-A* Allelic Expression Levels. *V. Ramsuran[1,2], S. Kulkarni[2], C. O'hUigin[1], Y. Yuki[1], D. Augusto[1,2], X. Gao[1], M. Carrington[1,2].* 1) Cancer Inflammation Program , National Cancer Institute, Frederick, MD; 2) Ragon Institute of Massachusetts General Hospital, Massachusetts Institute of Technology, and Harvard University, Cambridge, MA.

MHC class I expression levels influence the strength of immune responses and represent another variable in determining outcome to disease beyond peptide binding alone. Identification of the *HLA* loci that vary in allelic expression levels and delineating the mechanism responsible for expression variation may provide the opportunity to modify their expression therapeutically. We have examined the expression levels of allelic lineages at the *HLA-A* locus in a sample of 216 European Americans using a real time PCR assay, which amplifies all *HLA-A* lineages specifically with equal efficiency, and observe a gradient of expression that associates with *HLA-A* allelic lineage ($R=0.6$, $p=5x10^{-25}$). DNA methylation of the *HLA-A* gene appears to contribute to the variation in *HLA-A* mRNA expression levels, as a significant inverse correlation was observed between *HLA-A* mRNA expression levels in untreated cells and the degree to which expression is increased after treatment of the cells with a DNA methyltransferase inhibitor ($R=0.6$, $p=2.8x10^{-6}$). Further, deep sequencing and chromatin immunoprecipitation assays revealed allelic lineage specific methylation patterns within the *HLA-A* promoter region where increased DNA methylation levels correlated significantly with reduced *HLA-A* expression levels ($R=0.89$, $p=3.7x10^{-9}$). These data demonstrate *HLA-A* allelic lineage specific variation in expression levels, and DNA methylation as a key factor in contributing to this variation.

420F

Epigenome-wide association analyses of healthy human skin and blood DNA methylation profiles in relation to total body nevus count. *L. Roos[1], J. Sandling[2], D. Glass[1], T.D. Spector[1], P. Deloukas[3], V. Bataille[1], J.T. Bell[1].* 1) Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; 2) Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 3) William Harvey Research Institute, Queen Mary University of London, London, United Kingdom.

The strongest risk factor for melanoma is the number of melanocytic nevi. Nevi senesce from middle age onwards, but this can be delayed in individuals at high risk of melanoma. Genes such as *MTAP, CDKN2A,* and *PLA2G6* as well as those involved in telomere function have already been associated with nevus count. We performed an epigenome-wide association study (EWAS) by analyzing DNA methylation profiles from the Illumina 450k array in relation to total body nevus count, both in skin tissue from a sun-protected area and whole blood in 327 and 404 healthy female individuals, respectively.We identified 3 CpG sites in skin tissue that were significantly associated with nevus count at a false discovery rate (FDR) of 5%. The strongest signal ($p = 2.1 \times 10^{-8}$) was a CpG site hypermethylated with increasing nevus count, located in a CpG island shore upstream of lincRNA *AC144831.1*. The second ranked CpG was hypomethylated with increasing nevus count and was located in the promoter region of *C15orf48*, which is upregulated in skin carcinoma samples. The third CpG site was hypermethylated with increasing nevus count and locates in the transcription start site (TSS) of *ARRD1*, implicated in embryogenesis and tissue maintenance. At a less stringent threshold (FDR=10%) we found a negative correlation between DNA methylation and nevus count 200bp upstream of the TSS of *RAF1.* Mutations in *RAF1* can lead to oncogenic properties in melanoma, and *RAF1* inhibitors have been developed as successful drug targets. These results did not replicate in the whole blood EWAS. There we did identify 9 CpG sites that were significantly associated with nevus count at EWAS (FDR 5%). Among the significant findings was a CpG site located within *EBF3,* involved in neural development, which is of particular relevance as melanocytes are derived from the neural crest.To conclude, this is the first EWAS for total body nevus count and is performed in a phenotype relevant tissue, skin, as well as whole blood samples. We have identified CpG sites near genes that are relevant for the formation of nevi and melanoma. Furthermore, different significant CpG sites were found across the two tissues, indicating that the signal in skin tissue might be tissue-specific and potentially have direct mechanistic involvement. Future work will aim to integrate genetic variation, DNA methylation, and gene expression to increase our understanding of the mechanisms underlying development and formation of nevi and melanoma.

421T

DNA methylation profiling of brains of Parkinson disease patients. *S.K. Sivasankaran, A. Hossein, A. Mehta, L.G. Maldonado, K. Nuytemans, W.K. Scott, E.R. Martin, J.M. Vance, J.I. Young.* John P. Hussman Institute for Human Genomics, Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL, USA.

Parkinson's disease (PD) is a chronic multifactorial neurological disorder that mostly affects people over the age of 60. It is the second most common neurodegenerative disorder with a prevalence of 1.5 million people in U.S. alone and about 60,000 additional patients newly diagnosed each year. The core pathology of PD is defined by progressive loss of dopamine-producing neurons in the substantia nigra in brain. Approximately 15-20% of PD patients have a family history of PD, and there are rare cases with Mendelian inheritance. Genetic studies have identified several chromosomal loci associated with PD. However, the vast majority of PD cases are idiopathic, with disease likely caused by other genetic factors, epistatic effects, and gene-environment interaction effects. Recently, increasing evidence suggests that epigenetic mechanisms, such as DNA methylation and histone modification may also regulate the expression of PD-related genes, and contribute to PD etiology.Here, we investigated genome-wide DNA methylation in 24 (12 PD cases and 12 controls, matched on sex and race/ethnicity) prospectively collected autopsied brains to assess the methylation state of anterior cingulate gyrus DNA in relation to PD. We used state-of-the art data processing protocols, including normalization of probe intensity and background. For differential methylation tests, we performed logistic regression, adjusting for batch and age of death as covariates. A total of 17 cytosine-phosphate-guanine (CpG) sites were differentially methylated in our analysis (adjusted p-value < 0.05). Of particular interest is a differentially methylated CpG site located within the 5'-UTR of the tripartite motif-containing protein 2 (*TRIM2*, p-value = $1.7x\ 10^{-07}$). *TRIM2* is an excellent candidate gene as it is expressed in brain, had been linked to axonal function and previous studies found it associated to peripheral neuromuscular disorders. In mice, deficiency of *TRIM2* causes neurodegeneration.Our analyses suggest that the identified DNA methylation changes may have a role in PD and will be complemented by methylation profiling in additional brain regions.

422F

Genome-wide DNA methylation changes in the dorsal and ventral striatum of individuals with chronic cocaine dependence. *K. Vaillancourt[1,2], G.G Chen[1], A. Diallo[1], R. Poujol[1], C. Ernst[1,3], D.C. Mash[4], G. Turecki[1,2,3].* 1) McGill Group for Suicide Studies, Douglas Hospital Research Center, Verdun, QC; 2) Integrated Program in Neuroscience, McGill University, Montreal, QC; 3) Department of Psychiatry, McGill University, Montreal, QC, Canada; 4) Department of Neurology, University of Miami Miller School of Medicine, Miami, FL.

Background: Cocaine dependence is a chronic relapsing disorder whose development and trajectory is impacted by multiple genetics and environmental factors. Transcriptional changes accompany the transition from recreational cocaine use to cocaine dependence and epigenetic mechanisms may mediate these effects. Recently, several studies have identified epigenetic marks that are associated with the acquisition of compulsive drug seeking in animal models, but little is known about the role of epigenetics in human cocaine dependence. Of particular interest is DNA methylation as it represents a mitotically stable mark that has been shown to be altered by environmental experience.**Methods**: We used Reduced Representation Bisulfite Sequencing (RRBS) on post mortem nucleus accumbens and caudate tissue from 25 dependent cocaine users and 25 drug-free and age-matched controls. This approach has allowed us to detect genome-wide cytosine methylation at base pair resolution. Differentially methylated regions (DMRs) were validated using targeted bisulfite sequencing and cell-type specificity will be investigated in neuronal and non-neuronal nuclear fractions that have been separated using Fluorescence Activated Cell Sorting (FACS).**Results**: All RRBS libraries contained more than 4 million reads at 10X coverage and are well aligned to the human genome. In addition, all libraries were bisulfite converted with over 98% efficiency. Our study has identified multiple clusters of hyper- and hypomethylation associated with chronic cocaine dependence, in both brain regions. Differentially methylated regions span gene bodies, including introns and exons, as well as promoters and intergenic regions. High throughput discovery of cocaine-associated networks and pathways allow us to investigate the epigenetic changes in brain that accompany the transition from cocaine abuse to chronic cocaine dependence.Funded by NIDA (DA033684).

423T

Preliminary study on Genome wide methylation profile of liver biopsies in HCV infected patients with severe hepatic fibrosis. *L.R.S. Vasconcelos[1,2], R.F. Carmo[4], V.C.S. Sousa[3], S. Cabantous[5], C. Chevillard[5], L.M.M.B. Pereira[2], P. Moura[3], A. Dessein[5].* 1) Parasitology Departament, Aggeu Magalhães Research Center - FIOCRUZ-PE , Recife, Brazil; 2) Liver and transplant institute of Pernambuco - IFP-PE, Recife, Brazil; 3) Biological sciences institute - University of Pernambuco -ICB/UPE, Recife, Brazil; 4) Federal University of São Francisco Valley- UNIVASF, Petrolina, Brazil; 5) Immunology and Genetics of Parasitic Diseases, INSERM 906, Università d'Aix-Marseille, Marseille, France.

Introduction:Hepatitis C virus (HCV) infection is a major cause of chronic hepatitis, liver cirrhosis and hepatocellular carcinoma (HCC) worldwide. It is unknown why some individuals will develop severe forms of the disease.Studies have shown the importance of DNA methylation in the liver fibrosis progression. Therefore, theexploration of DNA methylation in the stages of liver fibrosis may disclose triggering mechanismsof HCV outcome.We conducted an exploratory investigation on methylation profileof liver biopsies from patients with HCV. Materials and Methods: Liver biopsies sampleswere collectedfrom HCV-RNA positive patients before treatment (n= 12),and they wereclassified byMETAVIR score.Six patients had fibrosisF0-F1 and 6 had F3-F4.Liver DNA was extracted with Zimo research kit and Illumina HumanMethylation450 BeadChip arrays was used to identify regions of hypo- or hypermethylation influencing hepatic fibrosis. Results:Regardingbiochemical analysis, serum levels of ALT, ASTand alpha-fetoproteinwere higher intheF3-F4 group (p=0.03, p=0.01 and p=0.04, respectively) characterizing active liver disease. Top methylation analysis showed 95 CpGs sites differentially methylated in F0-F1 vs. F3-F4 ($p < 5 \times 10\text{-}15$),in which12 siteswere-hypomethylated (in 10 genes) and 83 hypermethylated (in 64 genes). We performed Global methylation analysis in the 5kb distance of promoter regions from the genes that were differentially methylated ($p < 0.05$) in F0-F1 vs. F3-F4, top 10 genes hypermethylated included genes of innate immunity, extracellular matrix and mitochondrial function and top 10 hypomethylated included genes of Ras signaling and mitochondrial function ($p < 5 \times 10\text{-}5$). The pathway enrichment analysis showed that six pathways were associated to the F3-F4 methylation profile; BSID: 83105 (pathways in cancer); 102279 (endocytosis); 373901 (HTLV-I infection); 83051 (cytokine-cytokine receptor interaction); 523016 (transcriptional misregulation in cancer) and 658418 (viral carcinogenesis) PcBonf $< 3.99 \times 10\text{-}2$. Conclusion: Global methylation analysis in the promoter regions showed an important epigenetic regulation of genes involved with mitochondrial function, extracellular matrix, cell signaling and immunity, which may be related to the liver damage caused by HCV and that may emerge as a trigger for carcinogenesis, since F3-F4 profile strongly associated to cancer development.Further studies are being conducted to confirm this hypothesis.

424F

Epigenetic suppression of VEGF in retinal pigment epithelial cells by vitamin C. *G. Wang, D. Sant, J. Chen, D. Van Booven, V. Camarena.* Dr. John T. Macdonald Foundation Dept. of Human Genetics, Hussman Inst. for Human Genomics, Univ Miami, Miami, FL.

We and others recently reported that ascorbate is critical for DNA demethylation by serving as a cofactor for ten-eleven translocation (TET) dioxygenases to convert 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) in DNA. In this study, we examined the epigenetic effect of ascorbate in cultured human retinal pigment epithelial ARPE-19 cells. Treatment with ascorbate increased the global content of 5hmC in a time- and dose-dependent manner. RNA-seq was then conducted to evaluate the influence of ascorbate treatment on the transcriptome of ARPE-19 cells. We found 993 genes that were differentially expressed comparing cells cultured with (50 mM) or without ascorbate for 7 days using three different methods (EdgeR, DESeq and BaySeq). The expression of the top 10 genes (based on fold changes and *P* values) was further verified by qRT-PCR. In particular, a significant decrease in VEGFA expression (0.26 fold) was found after ascorbate treatment. The effect of ascorbate on VEGFA expression was further verified by western blot and ELISA of the conditioned media, in which ARPE-19 cells were cultured for certain periods of time. We are now examining 5mC and 5hmC in the VEGFA promoter region. Further, RPE appears to be one major source of VEGF after comparing its transcriptome with that of primary human retinal microvascular endothelial cells (1~2% of adjusted reads). Ocular anti-VEGFA therapy has been used to treat eye disorders such as neovascular age-related macular degeneration (AMD) and diabetic retinopathy. Whether it is antibody-based or VEGF-trap, the anti-VEGFA therapy is only effective for a subgroup of patients. Our results, though preliminary, suggest that ascorbate could be an epigenetic treatment for these ocular diseases.

425T

Set-based Methods for DNA Methylation Analysis. *Q. Yan[1], R. Fan[3], D. Weeks[2], N. Boutaoui[1], G. Canino[4], J. Celedon[1], W. Chen[1,2].* 1) Division of Pulmonary Medicine, Allergy and Immunology, Department of Pediatrics, Children's Hospital of Pittsburgh of UPMC, University of Pittsburgh, Pittsburgh, PA; 2) Departments of Human Genetics and Biostatistics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA; 3) Biostatistics and Bioinformatics Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD; 4) Behavioral Sciences Research Institute and Department of Pediatrics, University of Puerto Rico, San Juan, Puerto Rico.

With advances in microarray and next-generation sequencing technologies, genome-wide epigenetic data at a high resolution are available for hundreds and even thousands of subjects. Among these epigenetic data, the investigation of DNA methylation, (the addition of methyl groups to cytosines in the CpG dinucleotide context) can complement genetic studies. Although, in rare situations, a single methylated locus is associated with a disease, researchers are more interested in the association of a set of methylated loci, such as CpG islands, CpG shores and UTRs. In genetic studies, various methods have been developed to test for association between a group of rare genetic variants and complex diseases, and these methods can be extended to analyze set-based methylation loci for their effects on diseases. In this work, we conduct set-based statistics for examining association between set-based methylated loci and diseases. We propose two set-based approaches originally designed for genetic association analysis, Sequence Kernel Association Test and Functional Linear Model, to make them compatible with methylation data by considering the correlation between methylated loci. Results from our simulation studies show that both approaches retain the desired Type I error rate, and have comparable power in all the scenarios we considered. Finally, we illustrate the use of our proposed approaches by analyzing whole-genome DNA methylation data from studies on asthma and rheumatoid arthritis (RA). We identified potential differentially methylated genes associated with asthma-related traits, and found that the HLA gene (associated with RA in genetic studies) is differentially methylated in subjects with and without RA.

426F

Comprehensive Identification of Osteoblast-Specific DNA Methylation Signatures. *F. Yu, H. Shen, H.W. Deng.* Center for Bioinformatics and Genomics, Department of Biostatistics and Bioinformatics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA.

DNA methylation is an important epigenetic modification that contributes to the lineage commitment and specific functions of different cell types. Previous studies have suggested that DNA methylation plays important roles in regulating several genes that are critical for osteoblast differentiation and activities; however, the osteoblast-specific methylation pattern at genome-wide scale has not been fully examined. In this study, we compared the genome-wide DNA methylation profiles of human osteoblast with 21 other types of untransformed human cells in order to identify osteoblast-specific hyper- and hypo-methylation events. Genome-wide DNA methylation profiles assayed by reduced representation bisulfite sequencing (RRBS) were obtained from the ENCODE dataset. By carrying out differential methylation analysis, we identified 295 significant ($q < 0.001$, absolute mean methylation difference >20%) osteoblast-specific differentially methylated regions (DMRs), including 247 (84%) hyper- and 48 (16%) hypo-methylated regions. Interestingly, significant DMRs were not enriched in CpG islands (CGIs), but enriched in CGI shores and shelves, as well as 'open-sea' areas. Also, significant DMRs were more frequently observed in gene body and intergenic regions rather than regions proximal to transcriptional start sites. The gene ontology and phenotype ontology analyses revealed that the significant DMRs were highly enriched in biological processes related to transcriptional regulation and skeletal morphology and development. By integrating the DMR data with the extensive gene expression and chromatin epigenomics data in ENCODE and NIH Roadmap Epigenomics project, we observed complex, context-dependent relationships between DNA methylation, chromatin states, and gene expression, suggesting diverse DNA methylation-mediated regulatory mechanisms. Our results also suggested a number of novel osteoblast-relevant genes. For example, the integrated evidences from DMR analysis, histone modification and RNA-seq data strongly support a highly expression of a novel isoform of *neurexin-2* (*NRXN2*) gene in osteoblast. *NRXN2* was known to function as a cell adhesion molecule in the vertebrate nervous system, but its functional role in bone is completely unknown. In summary, we identified a number of key osteoblast-specific methylation events and revealed novel epigenetic mechanisms that may be critical for osteoblast differentiation and functions.

427T

Integration of DNA methylomic and transcriptomic changes in postmortem prefrontal cortex of subjects with alcohol use disorders. *H. Zhang[1], F. Wang[1], H. Xu[1], H. Zhao[2], J.H. Krystal[1], H.R. Kranzler[3], J. Gelernter[1].* 1) Psychiatry/VAMC, Yale University School of Medicine, West Haven, CT; 2) Department of Biostatistics, Yale University School of Public Health, New Haven, CT, USA; 3) Department of Psychiatry, University of Pennsylvania Perelman School of Medicine and VISN4 MIRECC, Philadelphia VAMC, Philadelphia, PA, USA.

DNA methylomic alterations in the prefrontal cortex (PFC) may contribute to risk for alcohol use disorders (AUDs). The DNA methylome in postmortem PFC of 16 male and seven female pairs of AUD and control subjects (46 European Australians) was examined using Illumina's HumanMethylation450 BeadChip assays. 1,812 CpGs (mapped to 1,099 genes) were differentially methylated in male AUD subjects ($9.5\times10^{-9} \leq P_{nominal} \leq 7.2\times10^{-4}$ and $q < 0.05$). In female subjects, no CpGs survived multiple testing corrections for their association with AUDs ($q > 0.05$). The above 1,812 CpGs were over-presented in two AUD-associated co-methylation modules ($M_{turquoise}$: 1,048 CpGs/683 genes; M_{blue}: 429 CpGs/304 genes) in male subjects ($P_{hyper} \leq 0.001$). Biological processes enriched for genes in these modules included neuron development and transcription regulation. Moreover, genes mapped by CpGs in these two modules were enriched in GWAS-identified genes associated with substance (alcohol, cocaine, opioid, or nicotine) dependence or psychiatric disorders (attention-deficit/hyperactivity disorder, autism spectrum disorder, bipolar disorder, major massive depression, or schizophrenia). Additionally, 106 of the 1,812 CpGs were mapped to 93 genes (e.g., AUD-associated genes *GRIK3*, *GRIN2C*, and *GABRA1*) with differential expression in postmortem PFC of male AUD subjects. Our findings suggest that DNA methylation alterations in reward-related brain regions result in an increased risk of AUDs.

428F

Epigenome-wide assoication in host genome for HIV infection. *K. Xu[1,2], X Zhang[1,2], Z Wang[6], Y Hu[4], R Sutton[5], Emu B[5], J Krystal[1,2], A Justice[3].* 1) Psychiatry, Yale School of Medicine , New Haven , CT; 2) Connecticut Veteran Health System, 950 Campbell ave, building 35, West Haven, CT; 3) Yale University School of Medicine, New Haven Veterans Affairs (VA) Connecticut Healthcare System, West Haven; 4) National Institute of Cancer, Center of Bioinformatics 9609 Medical Center Drive, Rockville, Maryland 20850; 5) Section of Infectious Diseases, Department of Internal Medicine, Yale University School of Medicine, New Haven, Connecticut, USA; 6) Department of Biostatistics, Yale School of Public Health.

Background: Epigenetic control of human immunodeficiency virus (HIV) gene expression is critical for viral transcription and latency in host T cells and macrophages. Little is known, however, regarding epigenetic changes in the host genome associated with HIV infection. Here, we report an epigenome-wide association study (EWAS) in 384 patients with and without HIV infection. We further tested the relationship of HIV-associated CpG sites with viral replication.**Methods:** Subjects were from the Veterans Aging Cohort Study (VACS) of HIV positive and negative patients in care in U.S., Department of Veterans Affairs Medical Centers. DNA was extracted from whole blood. 485,578 CpG sites were profiled using Illumina HumanMethylation 450K Beadchip. Data filtering, normalization, quality control were performed using programs in the R package. Generalized Linear Model was applied for association of each CpG site with HIV infection covariate for critical confounders. P value was corrected by Bonferroni correction. **Results:** Between HIV+ and HIV- patients, we found significant differential methylation at 211 CpG sites in genes implicated in HIV replication, immune, and inflammatory processes. Pathway analysis revealed the gene enrichments on chromosome 6p21, a region hosting Major Histocompatibility Complex (MHC) (false discovery rate q=0.0002). Specifically, two genes associated with inflammatory response were dys-regulated in HIV+ patients. First, three CpG sites in the promoter region of *NLRC5* (NOD-, LRR-, and CARD-containing 5), were hypomethylated in HIV+ patients, presumably increasing its gene expression. Increased *NLRC5* expression would inhibit the expression of *NFKB* and negatively regulate type I interferon signaling pathways. Second, *TNF*-α was hypermethylated in the gene body in HIV+ patients. Consistent with immune dysfunction, methylation of *NLRC5* and *TNF*-α correlated with HIV viral load in these patients after adjusted for CD4, CD8, and WBC counts (ps < 0.00001).**Conclusion:** Genome-wide methylation changes in PBMCs in HIV infected patients may influence immune function. Alterations in DNA methylation patterns in specific genes of HIV infected patients could affect outcome and might be targeted in novel approaches to therapy. **Acknowledgement:** The project was funded by NIH grants R21AA022870, R03DA039745, K12DA000167. We appreciate the supports from National Center of PTSD, U.S. Department of Veterans Affairs and Yale Center of Genetic Analysis.

429T

Exploration of hydroxymethylation in Kagami-Ogata syndrome caused by hypermethylation of imprinting control regions. *K. Yamazawa[1,2], K. Matsubara[2], M. Kagami[2], K. Nakabayashi[3], K. Hata[3], M. Fukami[2], T. Ogata[2,4].* 1) Clinical Genetics Center, National Hospital Organization Tokyo Medical Center, Tokyo, Japan; 2) Department of Molecular Endocrinology, National Research Institute for Child Health and Development, Tokyo, Japan; 3) Department of Maternal-Fetal Biology, National Research Institute for Child Health and Development, Tokyo, Japan; 4) Department of Pediatrics, Hamamatsu University School of Medicine, Hamamatsu, Japan.

Background: 5-hydroxymethylcytosine (5hmC), converted from 5-methylcytosine (5mC) by Tet enzymes, has recently drawn attention as the 'sixth base' of DNA since it is considered to be an intermediate of demethylation pathway. Nonetheless, it remains to be addressed how 5hmC is linked to the development of human imprinting disorders. In this regard, conventional bisulfite (BS) treatment is unable to differentiate 5hmC from 5mC, thus it is hypothesized that BS conversion-derived 'hypermethylation' at imprinting control regions (ICRs) which may cause human imprinting disorders would be in fact attributable to excessively increased levels of 5hmC as well as 5mC.**Methods:** We applied the newly developed oxidative BS (oxBS) treatment to detect 5hmC in blood samples from Kagami-Ogata syndrome (KOS14) patients caused by perturbed expression of clustered imprinted genes on 14q32.2 resulting from the hypermethylation of ICRs at this locus, IG-DMR and *MEG3*-DMR.**Results:** oxBS with pyrosequencing and cloning-based sequencing revealed that there were few amounts of 5hmC at the hypermethylated IG-DMR in blood sample from KOS14 patients. oxBS with genome-wide methylation array analysis demonstrated that global levels of 5hmC were very low with similar distribution patterns in blood samples from KOS14 patients and normal controls. We also confirmed that there was a huge amount of 5hmC in brain sample from a normal control. **Discussion:** 5hmC is not a major component either at the abnormally hypermethylated ICRs or at a global level, at least in blood from KOS14 patients. Considering that brain contains a large amount of 5hmC, neural tissues from KOS14 patients are promising analysis candidates to elucidate the role of 5hmC in the neurodevelopmental context.

430F

Peripheral blood DNA methylation markers associated with abstinence among chronic injection drug users in the ALIVE study. *K.M. Bakulski[1], K.S. Benke[2], A.E. Jaffe[2, 3], S.H. Sabunciyan[4], G.D. Kirk[1], S.H. Mehta[1], B.S. Maher[2].* 1) Epidemiology, Johns Hopkins University, Baltimore, MD; 2) Mental Health, Johns Hopkins University, Baltimore, MD; 3) Lieber Institute for Brain Development, Baltimore, MD; 4) Pediatrics, Johns Hopkins University, Baltimore, MD.

Background: Liability to drug addiction has a complex etiology with substantial contributions from both genetic and environmental factors. Epigenetic status represents an intersection of these factors that has not yet been explored in drug addiction epidemiologic studies.**Objectives**: The goals of this AIDS Linked to the Intravenous Experience (ALIVE) pilot study are to identify whether chronic injection drug use (IDU+) followed by abstinence (IDU-) is associated with blood DNA methylation marks, and to test whether these marks are associated with brain tissue DNA methylation comparing opiate users to controls.**Methods:** Among an ALIVE study subsample of 23 HIV-negative African American males, DNA was isolated from peripheral blood mononuclear cells at two time points. Baseline samples were taken from individuals with at least ten years of IDU+; the second sample followed at least two years of IDU-. Genome-wide DNA methylation was assessed using the Illumina Infinium 450k HumanMethylation array. We compared methylation status at each site during chronic use to cessation using a Wilcoxon Rank Sum test. We also tested the highest-ranking sites identified in blood for their concordance in post-mortem brain tissues of opiate drug users from the Lieber Institute for Brain Development.**Results:** While no single test reaches significance after False Discovery Rate correction (minimum P-value $<10^{-6}$), many sites exhibit marked difference (after cessation – chronic use). The average effect size (Cohen's d) among the top results is roughly 3. When we examine genome-wide average methylation status in injectors before (IDU+) and after cessation (IDU-), we find a significant decrease in methylation in the IDU- versus IDU+ sample. In addition, we observed large intra-individual variation at many sites in the pre- and post- comparison.**Conclusions:** In a pilot study, we performed a genome-wide scan of longitudinal DNA methylation among continuous injection drug users. Though underpowered to reach genome-wide significance, we identified genomic locations of DNA methylation sites measured in blood associated with cessation of drug use. These sites also appear to be altered in brain tissue from injection drug users, and may provide further insight about the role of epigenetics in drug addiction. Further work is needed, including study in the larger cohort as well as replication in additional populations, to clarify drug abuse mechanisms and the potential role of DNA methylation.

431T

Identification of housekeeping genes for DNA methylation study across different human ethnicities and tissues. *D.P. Chen, S.J. Wu, Y.C. Lin, C.S.J. Fann.* Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.

DNA methylation is an important epigenetic mechanism and participates in numerous cellular processes. Methylation array data has increased rapidly; for example, the number of Illumina HumanMethylation450 datasets publicly available in Gene Expression Omnibus has increased six times since 2013. Housekeeping genes (HKGs) of DNA methylation studies are crucial, because they can be used to normalize methylation datasets across different studies and to validate the efficiency of Bisulfite conversion. Previously, few studies have reported HKGs in methylation studies. However, most of HKGs found are either fully methylated or unmethylated.In this study, we proposed a systematic method to identify HKGs that are stable and consistent in methylation level. We used Illumina HumanMethylation450 BeadChips as platform, which contains more than 485K probes. We downloaded 309 datasets from different human ethnicities, tissues and conditions. To avoid batch effects, data within a chip is normalized by median and median absolute deviation (MAD). Median and MAD are less sensitive to outliers than mean and standard deviation. After quality control, such as negative beta values, redundancy, and low call rate, etc., 183 datasets were used for the analyses.Two methods were used to find the HKGs. For the first method, 12% of total probes with smaller variances across all chips were retained. The remaining probes were examined by the "two-one-sided test" that compares the averages of one probe from two datasets. All possible pairwise comparisons were performed, and the probes with close averages in many pairs of datasets were regarded as HKGs. For the second method, all probes were ranked within each dataset according to their variances, and then the ranks of probes were combined across datasets via geometric mean. Top ranked probes were regarded as HKGs. The result showed that RBM22, ARHGEF6, ZBTB8OS and a few other genes were identified as HKGs by using both methods. To our knowledge (by using PubMed, Web of Science, etc.), no evidence has been reported that these identified HKGs are disease- or tissue-specific differentially methylated genes.

432F

DNA Methylation Changes Observed in Rheumatoid Arthritis Joint Tissue Are Detectable in CD4+ Naive T Cells from Peripheral Blood. *C. Holingue[1], B. Rhead[1], M. Cole[1], X. Shao[1], H. Quach[1], D. Quach[1], E. Sinclair[2], J. Graf[3], T. Link[5], R. Harrison[3], V. Chernitskiy[3], W. Wang[4], G. Firestein[4], L. Barcellos[1], L. Criswell[2].* 1) Genetic Epidemiology and Genomics Laboratory, University of California, Berkeley; 2) University of California, San Francisco, Division of Laboratory Medicine; 3) University of California, San Francisco, Rosalind Russell / Ephraim P. Engleman Rheumatology Research Center; 4) University of California at San Diego, La Jolla; 5) University of California, San Francisco, Department of Radiology and Biomedical Imaging.

Aberrant DNA methylation patterns have previously been associated with rheumatoid arthritis (RA) [MIM 180300]. Our study aimed to determine whether differentially methylated CpGs in synovium-derived fibroblast-like synoviocytes (FLS) of RA patients were also differentially methylated in peripheral blood samples. DNA methylation was measured by generating 371 genome-wide DNA methylation profiles for 63 RA cases (57 seropositive), and 31 controls using Illumina HumanMethylation450 (450k) BeadChips. Cells from peripheral blood were FACS-sorted and CD14+ monocytes, CD19+ B cells, CD4+ memory T cells, and CD4+ naive T cells were assayed for each individual. One-tailed Wilcoxon rank sum tests were used to analyze case-control differences in the FLS candidates within these four cell types. Receiver operating characteristic (ROC) curve analysis was employed to test the predictive power of a hypermethylation score based on the differentially methylated sites we observed. Of the 5,532 hypermethylated FLS candidates, 1,056 (19%) were hypermethylated in CD4+ naive T cells of our RA cases compared to controls (FDR $q<0.05$). No other CpG candidates achieved significance in the other cell types. A hypermethylation score was calculated based on these results and had an area under the curve (AUC) of 0.73 when predicting RA case status. This hypermethylation score was compared to having the HLA-DRB1 shared epitope (SE) (yes/no) and a continuous genetic risk score consisting of 43 non-HLA SNPs (Yarwood 2013 and Eyre 2012). The SE model had an AUC of 0.66 (0.56-0.77), and the genetic risk score had an AUC of 0.51 (0.38-0.63). A combined model of shared epitope, GRS and the methylation score had an AUC of 0.78, which was the best predictive model. Both methylation score and SE remained significant ($p<0.05$) when included in a multivariable logistic model of methylation score, SE and genetic risk score, with RA case-status as the outcome. Although the methylation score had a greater AUC than the SE (0.73 versus 0.66), the SE had a larger odds ratio in the multivariable logistic regression model (5.31 versus 1.06). Our results suggest that measurement of DNA methylation in CD4+ naive T cells of peripheral blood may have diagnostic or prognostic value and represents one of the first steps towards precision medicine in RA.

433T

Comparing variability in the SeqCap Epi CpGiant and the Illumina 450k methylation microarray for characterizing the human methylome. *A.C. Just[1], R.O. Wright[2], M.M. Tellez-Rojo[3], A.A. Baccarelli[1].* 1) Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA; 2) Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 3) National Institute of Public Health, Center for Research in Nutrition and Health, Cuernavaca, Morelos, Mexico.

Background: The SeqCap Epi CpGiant is a targeted enrichment bisulfite sequencing assay designed to maximize overlap with the popular Illumina Infinium HumanMethylation450 BeadChip microarray. No prior publications have compared these two epigenome-wide methylation assays. In comparing these platforms we characterize the precision of these approaches, particularly in overlapping CpG sites and as a function of the read depth (number of sequencing reads at a given site).Methods: WBC DNA extracted from two cord blood samples collected in the PROGRESS cohort were assayed multiple times on the SeqCap Epi CpGiant (8x each) at Roche and the Illumina 450k at Illumina (15x and 2x, respectively) each according to their manufacturers specifications. Data preprocessing for the SeqCap Epi next generation sequencing data followed the Roche protocol including quality assessment, read filtering, alignment to the hg19 reference genome for bisulfite treated DNA with BSMAP, removal of PCR duplicates, and summarization of methylation at CpG contexts in aligned reads. Preprocessing for the Illumina 450k used out-of-band background correction. SeqCap Epi CpGiant and Illumina 450k results were compared at loci overlapping Illumina 450k CpG probes according to the Illumina annotation file.Results: Across all 16 runs, the CpGiant included reads on 3.7 million CpG sites on either strand of which 3.5 million were found in all 16 runs. Comparing across strands within samples, all 16 runs of the CpGiant had at least 30% of sites with a difference of more than 10% in the estimated methylation between strands (minimum of 10 reads on both strands). There were 446,886 sites covered by the CpGiant of the possible 482,421 CpG sites in the Illumina 450k. The across-replicate (technical) variance was higher for the CpGiant at 74% of these overlapping sites, and this did not vary substantially when stratifying by mean read depth of the CpGiant replicates.Discussion: The recently developed SeqCap Epi CpGiant covers a larger proportion of the methylome but an analysis of replicate samples suggests lower precision than the popular Illumina 450k microarray, even at high read depths. This may be a limitation when expecting small differences in methylation associated with environmental exposures or subtle phenotypic changes. More work is needed to understand determinants and post-processing control of technical variation in bisulfite sequencing assays.

434F

A DNA Methylation Signature of Alcohol Consumption. *C. Liu[1], R. Marioni[2], A. Hedman[3], A. Carpenter[4], D. Zhi[5], Q. Duan[6], S. Kunze[7], C. Elks[8], L. Reynolds[9], S. Aslibekyan[10], T. Tanaka[11], P. Tsai[12], K. Conneely[13], L. Stolk[14], J. Brody[15], B. Kühnel[7], C. Boer[14], S. Ligthart[14], A. Kretschmer[7], S. Love[16], J. Bell[12], Y. Liu[9], A. Baccarelli[4], E. Ingelsson[3], K. North[16], S. London[18], K. Ong[3], D. Levy[1], M. Fornage[17], The CHARGE + epigenetic methylation-alcohol group.* 1) NHLBI, Framingham, MA; 2) Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, UK; 3) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 4) Environmental Health, Harvard School of Public Health, Boston, MA, USA; 5) Department of Biostatistics, School of Public Health, University of Alabama at Birmingham, AL, USA; 6) Curriculum in Bioinformatics & Computational Biology Department of Genetics University of North Carolina at Chapel Hill, NC, USA; 7) Institute of Epidemiology II, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 8) Medical Research Council Epidemiology Unit, University of Cambridge, United Kingdom;; 9) Department of Epidemiology and Prevention Wake Forest School of Medicine Winston-Salem, NC, USA; 10) Department of Epidemiology, School of Public Health, University of Alabama at Birmingham, Birmingham, AL, 35294, USA; 11) Translational Gerontology Branch, NIA at Harbor Hospital, Baltimore, MD, USA; 12) Department of Twin Research and Genetic Epidemiology, King's College London, United Kingdom; 13) Department of Human Genetics, Emory University of Medicine, Atlanta, GA, USA;; 14) Internal Medicine, Erasmus MC, University Center at Rotterdam, Rotterdam, Netherlands; 15) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA;; 16) Department of Epidemiology, Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, NC, USA;; 17) Molecular Medicine and Human Genetics, The University of Texas Health Science Center at Houston, Houston, TX , USA;; 18) Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA.

Emerging evidence suggests that alcohol consumption alters DNA methylation (DNAm), which in turn can impact gene expression. Several studies with small sample sizes have yielded poorly replicated results for the association of alcohol consumption with DNAm. We measured DNA methylation at 475,000 CpGs using the Illumina HumanMethylation450 BeadChip on whole blood (WB) DNA from 13,346 individuals, including 9,634 of European ancestry (EA) and 2,461 of African ancestry (AA) and DNA from monocytes in 1,251 EA subjects. A linear mixed effects model tested the association of DNAm with alcohol consumption, adjusting for age, sex, BMI, batch, and differential cell count. Inverse variance weighted meta-analysis identified 881 significant (P<10-7) CpGs in EA-WB and 241 in AA-WB, and 80% of these significant CpGs displayed reduced methylation levels with increased alcohol intake.The effect estimates for the relationship between alcohol and DNAm for the 881 probes were correlated between EA-WB and AA-WB (r =0.67), and between EA-WB and EA-monocyte (r = 0.71) samples, suggesting shared effects of alcohol consumption on DNAm across different cell types and ethnic populations. About 3% of top CpGs had their effect estimates attenuated (i.e. change in beta > 20%) after additionally adjusting for smoking. Of the 881 CpGs, 341 had at least one significant methylation quantitative trait locus (mQTL, P<0.05/195562 pairs) within 100kb (n=2,500 EA). None of the significant mQTLs was a significant SNP in prior genome-wide association studies of alcohol intake. Of the 881 CpGs, 120 displayed significant correlation to transcription levels of 143 genes whose transcription start sites was within 1Mb of the CpG (P<0.05/23532 pairs, n=2,500 EA-WB). Pathway analyses for these 143 genes pointed to viral response (FDR-p=1x10-9), interferon signaling (FDR-p=5x10-8), and T-cell activation (FDR-p=7x10-7). Replication of these findings is underway.

435T

Genome-wide analysis of DNA methylation identifies a novel locus associated with bone mineral density. *J.A. Morris[1,2], P.-C. Tsai[3], F. Gao[4], V. Forgetta[2], Y. Xia[4], W. Yuan[3], C.M.T. Greenwood[1,2,5,6], E. Grundberg[1], T.D. Spector[3], J. Wang[4], J.T. Bell[3], J.B. Richards[1,2,3,6].* 1) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 2) Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal, Quebec, Canada; 3) Department of Twin Research & Genetic Epidemiology, King's College London, London, United Kingdom; 4) BGI-Shenzhen, Shenzhen, China; 5) Department of Oncology, McGill University, Montreal, Quebec, Canada; 6) Department of Epidemiology, Biostatistics, & Occupational Health, McGill University, Montreal, Quebec, Canada.

Purpose Osteoporosis is a common, complex disease characterized by increased bone fragility, often resulting in fracture. Bone mineral density (BMD) is measured to diagnose osteoporosis and estimate fracture risk. Studying the genetic determinants of BMD has led to novel disease pathophysiology insights. We sought to study epigenetic associations with BMD, as DNA methylation levels have been found to associate with disease phenotypes. **Methods** We undertook a genome-wide association study of DNA methylation levels to assess epigenetic associations with BMD. Methylation levels were estimated using methylated DNA immunoprecipitation sequencing (MeDIP-seq) to identify differentially methylated regions (DMRs) associated with BMD. BMD was measured at the forearm, femoral neck, and lumbar spine in the TwinsUK cohort, in 890, 1,518, and 1,397 twins, respectively. **Results** An intergenic DMR on chromosome 6q21 was found to be significantly associated with forearm BMD (P = 5.3×10-9). Through the usage of functional genomics data, we have linked the DMR to an enhancer putatively interacting with PRDM1, a gene regulating osteoclastogenesis. Using multiple data repositories for 3D nuclear conformations, enhancer RNAs (FANTOM5), DNase I hypersensitivity sites (ENCODE), and epigenetic marks (Epigenome Roadmap), we find evidence that the region encompassing the DMR likely contains a regulatory element of PRDM1, in that it is in close 3D proximity to the PRDM1 promoter and the DNA at these two regions is coincidentally open in relevant cell types such as monocytes, an osteoclast precursor cell. **Conclusion** Current work is underway to validate the DMR in an independent sample set by a targeted pyrosequencing-based approach for DNA methylation quantification. This study represents the most comprehensive genome-wide analysis to date of the role of methylation in BMD and forms a basis for future studies that will refine the definitive role of epigenetic-mediated effects at this osteoclast locus.

436F

Adjusting Infinium methylation profiles to suppress signals from varying cell proportion. *G.W. Nelson[1,4], L. Jia[2,4], F. Elloumi[2,4], E. Binns-Roemer[3,4], S. Limou[3,4], C.A. Winkler[3,4].* 1) BSP-CCR Genetics Core, FNLCR, Frederick, MD; 2) CCR-IFX, CCR, NCI, Bethesday, MD; 3) Basic Research Lab, FNLCR, NCI, Frederick, MD; 4) Leidos Biomedical Corp, Frederick, MD.

Methylation data ideally are generated from a single cell type, but often for practical reasons a mixture of cells is used. A well-studied case involves the use of mixed peripheral blood mononuclear cells (PBMCs). The R package minfi includes a function estimateCellCounts—implementing an algorithm developed by Housman et al.—that takes a methylation profile from mixed blood cells, typed on the Infinium HumanMethylation450 BeadChip, and returns estimates of the frequencies of common blood cell types in the sample. These estimates are then available for use as confounding covariates in analyses of methylation changes. But using cell composition as a confounding covariate may not always be a useful strategy. In the case of HIV infection, one of the strongest markers of disease progression is the loss of CD4+ T cells. This change of cell composition will strongly confound the analysis due to a loss of methylation markers characteristic of these cells. But correcting for CD4 count will tend to obscure methylation changes associated with AIDS progression. We therefore propose and construct an alternate correction strategy. The minfi package analysis is based on an Infinium data set from multiple blood cell types. Given the estimated cell composition for each sample, we subtract the methylation profile for each cell type from each sample, scaled by the proportion of that cell type in the mixed sample. This removes the cell marker variation that confounds the association analysis due to varying cell frequencies, removing the necessity of correcting for cell composition in the association analysis. Assuming that the methylation markers of disease progression are different from the cell type markers (if they are not there is no hope of observing them with mixed cell samples) we may observe the change of these markers with AIDS disease progression, and may use CD4+ T cell count as a measure of this progression. We tested this method on samples from the DC Gay AIDS cohorts, representing different stages of infection. Methylation markers that strongly distinguish CD4+ T cells will be most likely to show spurious association with AIDS progression, unless corrected for. Our correction reduced the (predominantly spurious) significance of the AIDS progression association, for most of these markers, by factors of 103 to 1010.

437T

Improvements To Existing QC Tools For Methylation 450K Arrays. *J. Romm, E. Pugh, K.F. Doheny.* CIDR/Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality Next-Gen Sequencing (NGS), Genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. At CIDR we use automated, real-time QC reports for genotyping and sequencing projects which allows the lab group to examine and compare dozens of metrics to detect failed or problematic samples and determine what, if anything, they share in common (e.g. failed array, position on a plate, DNA source). We wanted to create a similar report for Illumina's HumanMetylation450 array to supplement the QC information available within Illumina's raw data files.To accomplish this we used the R and BioConductor package Minfito calculate metrics for each sample that are not easily available in GenomeStudio, such as mean and standard deviation of beta, intensity of the methylated and unmethylated probes, detection P value, MDS distances, the mean and standard deviation of the values for Illumina's control probes and the mean and standard deviation of a subset of methylated and unmethylated probes from the Y chromosome. We then write the metrics out as a delimited file. We combine this with information about the sample, plating and lab variables. The combined information is imported into Spotfire® and used by the lab to determine outlying samples. This information is also provided to the investigator. With the investigator, we jointly determine whether the outliers represent sample failures or a subset of samples within the study that are meant to be outliers from a larger group (e.g. different sample preparation). Sample failures are redone one time in the lab and the metrics are recalculated and released to the investigator as part of our data release.We recently used this strategy for two projects. Project 1 had ~3000 samples, which included fully methylated and fully unmethylated controls for QC. We were able to identify 3 full array failures and one sample with low overall intensity. Project 2 had ~1000 samples, which also included fully methylated and unmethylated controls for QC. 12 % of the samples less than 99% of CpG islands detected. There were also 2 distributions of study samples when looking at the beta means plotted out of Minfi, which were scattered across arrays and lab processing plates.

438F

Lifestyle and biological age in a cohort study: methylation age vs telomere length. *A. Russo[1,2], S. Guarrera[1,2], G. Fiorito[1,2], C. Viberti[1,2], L. Iacoviello[3], M.C. Giurdanella[4], R. Tumino[4], C. Agnoli[5], V. Krogh[5], A. Mattiello[6], S. Panico[6], P. Vineis[1,7], C. Sacerdote[1,8], G. Matullo[1,2].* 1) Human Genetics Foundation-Torino, Turin, Italy; 2) Medical Sciences Department, University of Torino, Torino, Italy; 3) Fondazione di Ricerca e Cura "Giovanni Paolo II", Catholic University, Campobasso, Italy; 4) Cancer Registry and Histopathology Unit, "Civile-M.P. Arezzo" Hospital, ASP7, Ragusa, Italy; 5) Department of Preventive and Predictive Medicine, Epidemiology and Prevention Unit , Fondazione IRCSS Istituto Nazionale dei Tumori, Milano, Italy; 6) Department of Clinical and Experimental Medicine, Federico II University, Napoli, Italy; 7) Epidemiology and Public Health, Imperial College London, UK; 8) Cancer Epidemiology, CPO-Piemonte, Torino, Italy.

DNA methylation levels vary across life course and are influenced by environmental factors and lifestyle such as smoking and diet. Recently, age predictors based on DNA methylation profiles (methylation age) were developed to outline the difference between methylation-predicted age and chronological age (Dage) as a possible determinant of biological aging. To evaluate the effect of smoke and Mediterranean diet on Δage, we examined the methylation levels of more than 450K CpG sites in 292 matched case-control pairs belonging to the Italian section of the EPIC cohort. EPIC healthy volunteers were followed up for myocardial infarction and other diseases. We used the Illumina HumanMethylation450 BeadChip to assess CpG methylation levels on blood DNA provided at recruitment. Data were analyzed according to standard procedures (MethyLumi, Bioconductor). All analyses were corrected for sex, BMI, season and center of recruitment, estimated cellular subtypes and batch effect. The methylation-predicted age was created on the basis of a set of 71 methylation markers as described by Hannum *et al* (Mol. Cell. 2013). In our cohort, the correlation between methylation age and chronological age was highly significant (R=0.75; $p<$ 10-16). When analyzing the trend of Δage among current, former, and never smokers, we observed statistically significant differences between current and never smokers ($p<0.005$), as well as current and former smokers ($p<0.05$). Therefore, smoking was associated with an increased deviation of methylation age from chronological age. Conversely, in our study, smoking habits did not seem to influence leukocyte telomere length (LTL) measured on a subset of subjects (412) using RT-PCR ($p>0.05$), suggesting that methylation age could be a more reliable marker of biological age. We also evaluated the effect of adherence to a Mediterranean dietary pattern, as measured by the Italian Mediterranean Index score, according to Agnoli *et al* (Int J Cancer. 2013). Subjects who mostly followed a Mediterranean diet seemed to have a lower methylation age and longer LTL, reflecting a lower biological age. Other covariates like gender, alcohol, and BMI will be assessed, and most relevant results will be replicated within an independent population of 292 EPIC controls. In the frame of this replication analysis also the suitability of LTL as a marker of biological age will be further evaluated.

439T

Estimating and accounting for cell type composition in analysis of sequencing based methylation data. *A.A. Shabalin, M.W. Hattab, K.A. Aberg, E.J.C.G. van den Oord.* Center for Biomarker Research and Precision Medicine, Virginia Commonwealth University, Richmond, VA.

Variations in cell type composition can strongly affect measurements of DNA methylation and, if not accounted for, compromise downstream analyses or reduce its power. Estimates of sell type composition of analyzed samples can also give biological insight into case-control differences and allow to point which cell types contribute to methylation changes in bulk tissues. Next generation sequencing combined with enrichment technologies allows for measurement of DNA methylation at over 28 million common CpGs in human genome. The popular array based technologies (e.g. Illumina Infinium HumanMethylation450) interrogate up to 485 thousand methylation sites. However, unlike array based technologies, sequencing based approach produces read count type data, similar to RNA-seq data for gene expression. We have developed the proposed method with a focus on analysis of MBD-seq (methyl-CpG binding domain protein-enriched genome sequencing) data and have validated it by reproducing the analysis on whole genome bisulfite data. Existing method for estimation and accounting for cellular heterogeneity in array measured DNA methylation data are not directly applicable to sequencing based methylation measurements due to the differences in produced data and the very large number of interrogated CpGs.We present a new method for estimation and accounting cellular heterogeneity in sequencing based methylation data. The method involves normalization and weighting of methylation measurements based on the average coverage and local CpG density. The weighting is critical in reducing both variance and bias the cell type proportion estimates.We have validated the method on DNA measurements from isolated blood and brain cells. For blood we have data five frequently occurring clusters of differentiation (CD) in six subjects and samples with known cell type composition. We have further applied the method to methylation wide association studies for several large data sets, including 1,469 schizophrenia case and control blood samples, 1,200 major depression disorder (MDD) case and control blood samples and 383 MDD post mortem brain samples.

440F

Determining the DNA methylation landscape of human pancreatic islets using whole-genome bisulphite sequencing to characterise Type 2 Diabetes GWAS regions. *M. Thurner[1,2], M. van de Bunt[1, 2], K.J. Gaulton[1], A. Barret[2], A.J. Bennett[2], C.G. Bell[3], R. Lowe[4], S. Beck[5], V.R. Rakyan[4], A.L. Gloyn[1,2,6], M.I. McCarthy[1,2,6].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, University, of Oxford, Oxford, United Kingdom; 3) Department of Twin Research & Genetic Epidemiology, Kings College London, London, United Kingdom; 4) Centre for Genomics and Child Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, London, United Kingdom; 5) Department of Cancer Biology, UCL Cancer Institute, University College London, London, United Kingdom; 6) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom.

Type 2 diabetes (T2D) is a genetically complex disease with >80 GWAS loci identified to date.These data have demonstrated the importance of pancreatic islet dysfunction in T2D pathophysiology, but precise molecular mechanisms are mostly unknown.DNA methylation is an important tissue-specific, epigenetic mark that has a known role in common and syndromic disease. To investigate whether changes in methylation status underlie T2D association signals, we performed whole-genome bisulphite sequencing (WGBS) in 10 primary human islet samples.Samples were sequenced at 12-17X coverage to generate a full islet methylome covering 27x106 CpG sites. The high correlation between WGBS and Illumina 450k methylation data from islets (rho=0.89, 290k overlapping sites) confirmed the quality of our data. We defined low- and ultra-low methylated regions (LMRs and UMRs), which showed strong overlap with chromatin state-derived islet enhancers (82% of LMRs) and promoters (96% of UMRs), respectively. We found significant enrichment of LMRs (n=37k)in variants in high LD (r2 > 0.8) with the lead variant atT2D (fold enrichment (FE)=1.8, P=0.02) and fasting glucose level (FG) (FE=2.8, P=0.03) GWASlocicompared to matched loci.To determine whether this enrichment was due to islet-specific methylation, we took the subset of LMRs that overlapped regions with significantly different methylation (P<2.6x10-9) in islets compared to at least 4/6 publicly available methylomes (adipose, blood, buccal, liver, pancreas, lymphoblastoid cell line). These likely islet-specific LMRs (n=7.9k) showed stronger enrichment in T2D (FE=3.4, P=0.002) and FG GWAS regions (FE=7.2, P=0.007) compared to all LMRs as well as active enhancers (T2D FE=2.3, P=1x10-4; FG FE=3.8, P=5x10-4). One such islet-specific LMR, for example, overlaps the *ADCY5* GWAS variant rs11708067 that showed highly variable methylation across the WGBS islet samples and has previously been shown to be an islet methylation QTL in 450k methylation data. In addition to identifying LMRs and UMRs, we also evaluated previously known imprinted T2D loci such as *KCNQ1*. Consistent with a previously reported hemi-methylated state in adult islets, we found *CDKN1C* to be the only imprinted gene at this locus.These results highlight the utility of WGBS DNA methylation data from human islets in identifying islet specific-regulatory regions, and demonstrate that islet LMRs are enriched at T2D and FG GWAS loci.

441T

Detection of differentially methylated regions using both mean and variance differences. *Y. Wang, S. Wang.* Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY.

DNA methylation is an epigenetics modification of DNA which plays an important role in a variety of human diseases including cancer. Studies found that DNA methylation levels between cancer and normal tissues could differ in both means and variances, and those of neighboring CpG sites are often correlated. Most existing methods on identifying differentially methylated loci (DML) or differentially methylated regions (DMRs) were developed to detect differences in mean methylation levels between two study groups. Only a few published works took variance differences into consideration to detect DML, but none for DMR detection. Here we propose 1) a generalized exponential tilt model and 2) fisher's method to detect DMRs using signals from both means and variances in methylation levels using both case-control and matched case-control designs. In both proposed methods, we smooth the site-level test statistics and consider a region as a candidate DMR when there are at least three consecutive loci whose smoothed statistics exceed a pre-defined threshold. The significance of candidate DMRs were assessed using permutation. The performances of the proposed methods that combine mean and variance signals for DMR detection were compared with parallel methods that only consider mean signals or variance signals. In simulation studies, the proposed methods are most powerful when there are weak mean and variance signals. The performances are comparable to methods that only consider mean differences or variance differences but with strong signals. We also applied the proposed methods to two DNA methylation datasets, one from an oral cancer case-control design, one from the Cancer Genome Atlas (TCGA), and identified additional DMRs that were missed by tests that only consider mean differences.

442F

Preliminary analysis of genome-wide methylation qualitative trait loci for HIV infection. *X. Zhang[1], A. Justice[1], Y. Hu[3], Z. Wang[2], J. Krystal[1], K. Xu[1].* 1) Department of Psychiatry, Yale University, New Haven, CT; 2) Yale School of Public Health, New Haven, CT; 3) Center for Biomedical Informatics and Information Technology (CBIIT) National Cancer Institute, NIH, MD.

Background: Epigenetic changes in the host genome may play a critical role in the response to HIV infection. Genetic variation can influence the epigenetic response of that hosts to viral infection. The purpose of this study is to profile genome-wide genetic architecture of cytosine modifications in DNA derived from whole blood in an African American population with and without HIV infection.**Methods**: We detected methylation using Illumina HumanMethylation 450K Beadchip and genotyped single nucleotide polymorphisms (SNPs) using Illumina HumanExpress Exome Beadchip. We first performed epigenome-wide association study (EWAS) to identify significant CpG sites for HIV infection in 384 subjects. Generalized Linear Model (GLM) was conducted, adjusted for age, gender, race, type of drug abuse, and 10 principal components of methylation data. We identified 182 significant methylation sites after Bonferroni adjustment ($p < 0.05$). Next, we tested mQTL for 182 CpG sites and 1,347 SNPs within a window of 20,000 bp in 85 HIV positive and 36 HIV negative samples that had both methylation and genotyping data. Linear regression model was applied to examine correlations of CpG and SNP, with adjustment of age and 3 principal components of genotyping. Significant p was set at 0.05. We did not correct p value for multiple testing in this pilot study because the sample size was under power to detect significant genome-wide mQTL.**Results:** We identified 102 CpG-SNP pairs located in n 39 genes. Interestingly, these CpG-SNP pairs were enriched at promoter (22), less in gene body (20) and 3' UTR (10 probes). The rest of CpG-SNPs were in the first exon or intrageneic regions. Multiple genes carrying significant mQTLs were previously reported to involve HIV pathophysiology, such as *CD4, CCR6, NHEJ1, SDK2, REM2, ARRB1, CD207*.**Conclusion**: Our preliminary findings suggest that genetic variants control DNA methylation in HIV infection. mQTL annotation can be useful to identify novel loci for complex trait, and the genetic related methylation pattern could be applied in the classification of HIV infection, , treatment response, and HIV outcomes, . Further study in a larger sample population to replicate the findings is warranted.**Acknowledgement:** The project was funded by NIH grants R21AA022870, R03DA039745, K12DA000167. We appreciate the support from National Center of PTSD, U.S. Department of Veterans Affairs and Yale Center of Genetic Analysis.

443T

Systematic identification of downstream *trans*-effects for 1,900 known disease associated SNPs. *M.J. Bonder[1], R. Luijk[2,3] on behalf of the BBMRI-NL BIOS* Consortium.* 1) Genetics, University Medical Center Groningen, Groningen, Groningen, Netherlands; 2) Department of Molecular Epidemiology, Leiden University Medical Center, 2333 ZC Leiden; 3) Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, 2333 ZC Leiden.

Genetic risk factors identified in genome-wide association studies are mostly non-coding, making it difficult to understand their functional consequences. So far, large-scale*trans*-eQTL analyses have identified such downstream functional consequences for only 233 SNPs (Westra et al, NG 2013). To increase this, we used methylation-QTL mapping in peripheral blood of 4,000 population based samples from the Dutch BBMRI-NL BIOS consortium. We observed that 1,907 different GWAS SNPs affect methylation of over 10.141 unique CpGs sites*in trans*(FDR <0.05), representing a eight-fold increase in the number of disease-associated SNPs for which downstream functional effects can be detected.

To address the question in what particular biological processes these specific CpGs are involved, we also generated RNA sequencing data for 2,000 of the samples, permitting us to empirically relate CpG methylation to gene expression effects (eQTMs) for over 12,800 CpG sites (FDR <0.05). By using different genomic annotations we could accurately predict (AUC = 0.83) whether these methylation-gene expression relationships were positive (31%) or negative (69%).

By finally integrating the*trans*-meQTLs and eQTMs and adapting pathway enrichment method DEPICT, we obtained insights in the downstream functional effects of many genetic risk factors: rs3774959 (mapping close to NFKB1 and associated with ulcerative colitis) significantly affects methylation levels of 348 different CpG sites, of which many map within genes of the NF-kappaB cascade.

These results indicate that large-scale meQTL mapping permits discovery of previously unknown downstream molecular effects for many genetic risk factors, and these effects on*trans*-methylation levels have a clear biological basis.

* BIOS stands for Biobank-based Integrative Omics Study, a BBMRI-NL Rainbow project steered by Lude Franke (UMCG), Bas Heijmans (Chair, LUMC), Peter-Bram 't Hoen (LUMC), Aaron Isaacs (EMC) Rick Jansen (VU), and Joyce van Meurs (EMC).

444F

Race Specific Differential DNA Methylation Marks of COPD. *R. Busch[1], W. Qiu[1], G. Criner[2], D. DeMeo[1].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 2) Section of Pulmonary And Critical Care Medicine, Temple University School of Medicine, Philadelphia, PA.

Background: African-Americans with COPD experience more severe airflow obstruction per pack-year of smoking than non-Hispanic Whites. Smoking-related DNA hypomethylation and resultant differential gene expression may contribute to the pathogenetic mechanisms of racial disparity in COPD susceptibility among African-American (AA) smokers. Prior research identified differential methylation signals related to tobacco smoke exposure, but not associations with COPD in AA. To address this knowledge gap, we investigated differential methylation of CpG sites associated with COPD among AA in the PA-SCOPE cohort.**Methods:**We assessed DNA methylation from peripheral blood samples in 357 AA smokers with (N=91) and without COPD (N=266) in the PA-SCOPE cohort using the Illumina Infinium 27K platform. After quality control, mixed linear models were used to assess associations of COPD (defined as FEV1/FVC<0.7) with differential methylation at CpG sites, controlling for age, sex, assay batch, and pack-years of smoking.**Results:**We observed 3536 CpG sites significantly associated (adjusted p-value<0.05) with the presence of COPD in the PA-SCOPE AA cohort. 1851 of these sites were hypomethylated, while 1685 were hypermethylated. Top associated hypomethylated sites included *RBFOX2* (p=3.6x10-11), *GRASP* (p=3.6x10-11), and *FXYD1* (p=3.6x10-11). Additional biologically plausible associations included hypomethylation of the *LPO* gene (p=3.3x10-10), which is involved in the production of free radicals in airway epithelial immune response, as well as hypomethylation of known COPD genetic associations such as *FAM13A* (p=5.0x10-9) and *SERPINA1* (p=6.9x10-8). Gene ontology-based analysis suggested that this gene set included a significant number of genes involved in immune processes, leukocyte activation, and wound healing, as well as responses to stimuli, signaling, and cell communication. We compared these results to the COPD-associated CpG sites in non-Hispanic whites described by Qiu et al in AJRCCM in 2012, and found that 2355 significantly differentially methylated CpG sites were unique in our AA set, including genes such as *FAM13A*.**Conclusions:**African-Americans may have a unique subset of differentially methylated gene-targets significantly associated with COPD when compared with non-Hispanic Whites. Future investigation will examine gene pathways of these targets and how this may explain racial disparities in COPD susceptibility and severity.

445T

Masculinizing gene expression and DNA methylation in XX neural stem cells and their differentiated progeny by a single exposure of testosterone: An *in vitro* approach of understanding hormonal organization. *M. Bramble, L. Roach, A. Lipson, T. Ngun, N. Vashist, J. Gosshcalk, E. Vilain.* Dept of Human Genetics, University of California-Los Angeles, Los Angeles, CA. 90025.

The role of gonadal hormones and their influence on sexual differentiation and brain masculinization have been highly investigated in the rodent model over the past century. However, the studies that have been conducted failed to unravel at the molecular level exactly how these hormones may alter sexual behavior. Our research group recently found that neonatal androgen exposure is capable of altering DNA methylation patterns in sexually dimorphic regions of the rodent brain. The methylation differences may be an early molecular explanation as to how gonadal hormones permanently alter the developing brain, through the process known as hormonal organization. Using sex-specific embryonic day 14 C57/B6J mouse neural stem cells, we investigated the effects of exposures to testosterone propionate (TP) and its derivatives on multipotent neural stem cells prior to their maturation into neurons or astrocytes. Early results suggest significant and long lasting changes in the expression of numerous genes as a result of TP treatment at the NSC stage. We have demonstrated that testosterone exposure on XX neural stem cells can aid eliminating sex differences in gene expression, which are able to be carried over to the final stages of differentiation, again minimizing sex differences in both neurons and glia cell types. This phenomenon is likely due to epigenetic mechanisms, as we have shown that the global DNA methylation patterns in XX NSCs that are exposed to TP are not significantly different than basal XY NSCs. Prior to TP exposure however, DNA of XX NSCs are hypomethylated compared to their XY counterpart, which suggests that basal sex differences in the early developing central nervous system may be mediated through DNA methylation and other epigenetic mechanisms prior to gonadal hormone exposures during development. Using this *in vitro* approach coupled with an *in vivo* verification, we aim to elucidate the role that DNA methylation and gene expression changes have on neural stem cells, and how these alterations affect the final differentiated cell types. .

446F

Integrative approach for functional methylation loci with next-generation sequencing. *H. Xu[1], D. Ryu[1,2], V. George[1].* 1) Dept Biostatistics & Epidemiology, Georgia Regents University, Augusta, GA; 2) Division of Statistics, North Illinois University, DeKalb, IL.

DNA methylation have been involved in cancer and many other complex diseases. Current methods for identifying disease-related DNA methylation loci are mostly based on analyzing genomic patterns of DNA methylations alone and are susceptible to high false positives. We develop an approach based on integrative analysis of RNA-seq and DNA methylation data from next-generation sequencing. Our approach combines information from the both RNA-seq and DNA methylation through Liptak's method to identifying functional methylation loci. Simulations show that it has lower false positive rate than current methods that analyze DNA methylation alone. By combining information from both types of data, our approach also has improved power compared to the methods analyzing DNA methylation alone. We applied our approach to a thyroid cancer data and identified several new functional methylation loci for thyroid cancer.

447T

Oxidative stress pathways implicated in comprehensive epigenetic and transcriptomic assessment of adult hippocampus from fetal ethanol-exposed mice. *E.J. Diehl, B.I. Laufer, C.A. Castellani, B.A. Alberry, S.M. Singh.* Biology, University of Western Ontario, London, Ontario, Canada.

Alcohol abuse during pregnancy leads to a range of neurological abnormities termed Fetal Alcohol Spectrum Disorder (FASD). Its molecular basis is poorly understood, however epigenetic dysregulation of genetic programs in the brain are involved. We have developed a mouse model of FASD showing learning and memory impairment and persistent changes in brain gene expression into adulthood. Epigenetic phenomena likely maintain these changes; however, few studies have examined these modifications comprehensively.We examined global differences in gene expression, DNA methylation, miRNA, and histone methylation in the hippocampi of ethanol-exposed mice. We hypothesized that altered epigenetic mechanisms induce gene expression changes in the hippocampus which may underlie learning and memory impairment. Mouse pups were injected with saline or ethanol on postnatal days 4 and 7, equivalent to human trimester three. At 70 days of age, hippocampus was isolated and used for gene and miRNA expression microarray, methylated DNA immunoprecipitation microarray (MeDIP-chip), and histone H3 lysine 4 trimethylation and H3 lysine 27 trimethylation ChIP-chip.The top gene expression pathway was "Free Radical Scavenging, Gene Expression"; we confirmed four of these expression changes by droplet digital PCR (ddPCR). The top pathway for genes affected by epigenetic methylation changes was "Peroxisome biogenesis"; we confirmed two of these changes by bisulfite pyrosequencing. Alteration of pathways indicates a widespread dysregulation of the oxidative stress response. Ethanol is known to induce oxidative stress in the developing brain through a variety a mechanisms including reduction of antioxidant levels and increasing reactive oxygen species (ROS). Alteration in the epigenetic regulation of oxidative stress response genes into adulthood represents a novel point of interface between the epigenetic and oxidative stress mechanisms of FASD.

448F

High SUMO E3 ligase PIAS1 protein expression in Breast Cancer Luminal A molecular subtype. *A. Mannermaa[1,2], R. Pirinen[3], R. Sironen[1,2], H. Peltonen[1,2], V-M. Kosma[1,2].* 1) Dept Pathology & Forensic Med, Univ of Eastern Finland, Kuopio, Finland; 2) Imaging Center, Pathology, Kuopio University Hospital, Kuopio, Finland; 3) Department of Pathology, Northern Karelia Central Hospital, Joensuu, Finland.

PIAS1 acts as a SUMO E3 ligase by specifying target proteins for SUMO conjugation by Ubc9. The sumoylation activity of PIAS1 suppresses TGFβ-induced Epithelial-Mesenchymal transition (EMT). Thus, SUMO E3 ligase PIAS1 has emerged as a critical regulator of EMT, which contributes to epithelial tumor cell invasiveness and metastasis. Interestingly, PIAS1 is also a chromatin-bound Androgen Receptor (AR) coregulator that functions in a target gene selective fashion to regulate prostate cancer cell growth. Here we report a PIAS1 immunohistochemical tumor staining study in an Eastern Finnish breast cancer sample set.PIAS1, Estrogen receptor (ER) and progesterone receptor (PR) were evaluated by immunohistochemical staining from paraffin embedded tissue samples. HER2 mutation status was obtained from clinical data of the patients.Statistical analyses were carried out using SPSS 21 for Windows. Differences between groups were analyzed by either non-parametric Kruskal-Wallis test, Mann - Whitney U-test or by Fishers's Exact Test (2-sided).PIAS protein expression was negative or weak in 22.7% of the tumors, moderate in 51.2% and high in 23.8%. We detected an association between the expression of PIAS1 and ER positivity (p=0.023) and PIAS1 positivity and HER2 amplification negativity (p=0.03) in invasive cases. After grouping tumours according to molecular subtypes Luminal A (ER+ and/or PR+, HER2-), Luminal B (ER+ and/or PR+, HER2+), Triple negative/basal-like (ER-, PR-, HER2-) and HER2 type (ER-, PR-, HER2+) a strong association between PIAS1 expression and Luminal A subtype was seen (p=0,01).Our results provide evidence for PIAS1 as a classifier of breast tumours and urges for additional studies on the role of this EMT and AR related regulator in breast cancer.

449T

Histone Binding Strength is Quantitatively Associated with Gene Expression Across Individuals. *K. Fletez-Brant[1,3], Y. He[2], K.D. Hansen[1,3].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Dept. of Epidemiology, Johns Hopkins Bloomberg School of Public Health; 3) Dept. of Biostatistics, Johns Hopkins Bloomberg School of Public Health.

Histone modifications are known to be associated with active and inactive parts of the genome. These modifications are mainly profiled genomewide using ChIP-seq, and for technical reasons most existing experiments are performed in cell lines with unlimited input material. Little is known about differences in histone marks profiled in the same cell type but in different individuals.We study this question using publicly available data on histone modifications H3K4me1, H3K4me3 and H3K-27ac from multiple different HapMap cell lines. We show a quantitative association between the expression of a gene and H3K4me3 binding at its corresponding promoter. By analyzing input experiments on the same samples we show that the input channel measured across a promoter show some quantitative association with gene expression; this emphasises the need for using experimental controls in this analysis. By comparing with the input channel, we conclude that H3K4me3 do have a quantitative relationship with gene expression at some, but not all, promoters. This association shows H3K4me3 binding strength is a quantitative and functionally relevant measure, which is associated with biological variation. We establish a similar quantitative relationship between gene expression and strength binding by H3K4me1 and H3K27ac at corresponding enhancers, with enhancers defined by the FANTOM 5 project. Finally we establish the quantitative relationship between promoter H3K4me3 binding and gene expression in rat samples, thereby replicating our finding in a different mammal.

450F

Inter-species comparison of endothelial cell gene regulation reveals the conserved control of vascular disease genes. *A. Medina Rivera[1,2], L. Antounians[1,3], A. Alizada[1,3], M. Liang[1,3], L. Dang[4,6], F. Gagnon[5], J. Fish[4,6,7], M. Wilson[1,3].* 1) Genetics and Genome Biology, SickKids Research Institute; 2) Laboratorio Internacional de Investigación Sobre el Genoma Humano, UNAM, Juriquilla, Queretaro, Queretaro, Mexico; 3) Department of Molecular Genetics, University of Toronto; 4) Department of Laboratory Medicine and Pathobiology, University of Toronto; 5) CIHR STAGE, Dalla Lana School of Public Health, University of Toronto; 6) Toronto General Research Institute (TGRI); 7) Heart and Stroke Richard Lewar Centre of Excellence.

Comprehensive maps of transcription factor (TF) binding and epigenetic modifications are rapidly generating new testable insights into human disease gene regulation. However, it is still not trivial to identify which gene regulatory regions have the most impact on tissue specific gene regulation. We set out to identify gene regulatory regions that are required for vascular homeostasis by making controlled cross-species epigenetic comparisons (Zoo-ChIP) in aortic endothelial cells (ECs) isolated from multiple species. We profiled the JUN transcription factor, an important member of the AP1 complex, and several histone modifications in human, rat and bovine aortic ECs cultured ex vivo. We identified shared orthologous TF binding and shared orthologous active histone modifications. We found that approximately 5% of JUN and 33% of H3K27ac are shared between human and one additional mammal (rat or cow). These shared putative regulatory DNA regions were enriched for EC pathways including angiogenesis and nitric oxide metabolism. Shared orthologous JUN binding sites also coincided with reported regulatory human disease mutations, several of which have a plausible EC component. Importantly, only a minority of these shared JUN/H3K27ac interactions could have been predicted using established measures of DNA constraint or ChIP-seq signal alone. By comparing our findings to similar studies performed in primary liver tissue, we demonstrate that comparative epigenomic profiling enriches for tissue-specific pathways and distinct human regulatory DNA mutations.

451T

Effect of natural variation in copy number on epigenetic patterning. *J. Yu, C. Li, K. Iyengar.* Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.

Background. Copy number variation (CNV), the most prevalent type of structural variation, has been associated with many human diseases, but its role in disease pathogenesis is far from clear. We hypothesized that CNVs alter disease risk through differences in epigenetic patterning, and performed a systematic evaluation of the *cis* relationships between CNVs and histone marks overlapping a called CNV in two embryonic cell lines, H1 and H9, and a lymphoblastoid cell line, GM12878; these lines are widely studied for epigenetics. **Data.** We obtained data from the ENCODE, Roadmap, and 1000 Genomes Project (1KG): CNV data for H1 and GM12878, and SNP chip data for H1 and H9 with CNVs called by PennCNV. When CNV calls were available from multiple sources, only overlapping calls were kept. We also obtained consolidated profiles for histone marks (27 for H1, 28 for H9, and 11 for GM12878) and DNase hypersensitivity as a covariate. **Analysis**. CNVs status were categorized as deletion (CNV=0,1), duplication (CNV=3,4), and normal (CNV=2). For each histone mark, we performed linear regression of log-signal intensity on CNV status, with and without DNase. We then performed two meta-analyses, one over H1 and H9, and the other over all three cell lines. **Results**. **1)** When adjusting for DNase, nominal significance was observed for duplications for 17 marks in H1, 16 in H9, and 8 in GM12878; among these, 9 were in both H1 and H9, and 4 in all three lines. **2)** Nominal significance for deletions was found for 3 marks in H9, 10 in GM12878, and 2 in both. **3)** After Bonferroni correction, we found significant *cis* effects of duplication on increased intensity of H3K9me3 ($p = 1e-3$ for with and without DNase) in meta-analysis of H1 and H9, and of deletion on increased intensity of the promoter marks H3K4me2 ($p = 2e-4$ and 1e-4) and H3K9ac ($p = 5e-6$ and 1e-5) in meta-analysis over all three lines. **4)** Among the CNVs driving our significant results, one duplication CNV straddled the olfactory receptor family on chromosome 14. Olfactory receptors are known to be epigenetically programmed. Another duplication CNV overlapped with defensins on chromosome 20, a known copy variable region. Several deletion CNVs overlapped with the SET domain on chromosome 12.

452F

Methylation alterations of LINE-1 and imprinting genes related to folate deficiency during embryo development. *L. Wang, T. Zhang, W. Zhen, S. Chang.* Capital Institute of Pediatrics, Beijing, China.

Periconceptional maternal folate deficiency is a risk factor for neural tube defects (NTDs), the mechanism underlied remains unclear. Intracellular folic acid is thought to be key player in providing an adequate source of methyl groups for methylation of DNA and proteins, and methylation modifications of repeat elements and imprinting genes is suggested to be sensitive to epigenetic regulation of folate acid nutrition in developmental programming. However, the mechanism of how folate deficiency altered methylations modification of repeat elements and imprinting genes in earlier development is still unclear. Our present studying analyzed methylation modification alternations of LINE-1 and imprinting genes using NTD cases, mESCs with folate deficiency, and also in mouse models with periconceptional folate-deficiency. The results implying that hypomethylation of LINE-1 was associated with an increased risk of NTDs. Folate insufficiency induced LINE-1 hypomethylation at the lowest levels in folate-free ESCs, compared with that in the folate-normal group. Moreover, LINE-1 methylation level was positively correlated with folate content, and negatively correlated with homocysteine content. Similar, kinds of imprinting genes also altered imprinting modifications in NTDs with folate deficiency, including Gnas imprinting cluster, DLK-MEG3 imprinting cluster, MEST imprinting and H19/IGF2 imprinting. Mouse embryos exhibited a significantly increased ratio of IUGR and developmental dysplasia of the brain in response to folate deficiency. Data shown here implying that periconceptional maternal folate deficiency altered the imprinting established and gene expression in fetal brains, which may be associated with the higher ratio of developmental dysplasia.

453T

Investigating Imprinting As A Mechanism For The Development Of Asthma and related phenotypes In Two Canadian Birth Cohorts. *A. Eslami[1], L. Akhabir[1], G. Ellis[1], AB. Becker[2], AL. Kozyrskyj[3], C. Laprise[4], PD. Paré[1], AJ. Sandford[1], D. Daley[1].* 1) Centre for Heart Lung Innovation, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada; 2) Department of Pediatrics and Child Health, Faculty of Medicine, University of Manitoba, Winnipeg, MB, Canada; 3) Department of Pediatrics, Faculty of Medicine and Dentistry, University of Alberta, Edmonton, AB, Canada; 4) Università du Québec á Chicoutimi, Saguenay, QC, Canada.

Background: Asthma is a complex disease caused by a combination of genetic and environmental factors. To date, 23 genome-wide association studies (GWAS) have been completed for asthma as a primary phenotype. Combined, these studies have identified 44 SNPs at a p value <1x10-6. The consensus is that main genetic effects of these common SNPs (with modest effects) do not fully explain the heritability of asthma. This has led to further studies of more complex models to identify the "missing heritability" in asthma. Genomic imprinting is an alternative mechanism, which may explain some of the missing heritability. Imprinting is an epigenetic phenomenon where the expression of genes depends on their parental origin (parent-of-origin effect). Imprinting effects have been reported in the development of many complex diseases including asthma.**Hypothesis**: Specific genomic imprinted regions are associated with asthma and related phenotypes (atopy and airway hyperresponsiveness (AHR)).**Methods**: To identify candidate genomic regions for imprinting we used GWAS data from two family-based studies (two parents and one offspring). These studies are: 1) the Canadian Asthma Primary Prevention Study (CAPPS), a high-risk asthma birth cohort and, 2) the Study of Asthma Genes and Environment (SAGE), a population-based asthma birth cohort. We used a likelihood-based variant of the Transmission Disequilibrium Test. Parent-of-origin effects were tested by including a modifier (the sex of parents) in the analysis.**Results**: In the joint analysis of CAPPS and SAGE with 148 asthmatic complete trios, 13 SNPs showed significant parent-of-origin effects with p<10-5. 3 SNPs remained significant after 100,000 permutations. Notably, we showed a parent-of-origin effect at a known imprinted gene, CTNNA3; this gene was previously associated with occupational asthma in a GWAS. Six SNPs were in or near Long non-coding (lnc)RNA genes. LncRNAs have been increasingly implicated in imprinting by modulating imprinting control regions or being imprinted themselves. The joint analyses for atopy (237 trios) and AHR (231 trios) yielded 3 and 2 significant SNPs after permutation, respectively.**Conclusion**: We identified several SNPs showing parent-of-origin effects in asthma and related phenotypes. To increase statistical power and confirm our results, we will perform a meta-analysis using 3 family studies: CAPPS, SAGE and the Saguenay-Lac-Saint-Jean and Quebec City, a French-Canadian founder population.

454F

Somatic cell hybrids as a model system to evaluate genome editing and genomic imprinting of pig chromosomes in the development of pre-clinical large animal models. *R.D. Nicholls[1,2], M.A. Johnson[1], F. Mompart[3], S.M. Gollin[2], K.D. Wells[4,5], R.S. Prather[4,5], M. Yerle[3].* 1) Dept of Pediatrics, Div of Medical Genetics, Children's Hospital of Pittsburgh of UPMC, Pittsburgh, PA; 2) Dept of Human Genetics, School of Public Health, University of Pittsburgh, Pittsburgh, PA; 3) INRA-INPT UMR1388 Génétique Physiologie et Système d'Elevage, Castanet-Tolosan, France; 4) Div of Animal Sciences, College of Agriculture and Natural Resources, University of Missouri, Columbia, MO; 5) National Swine Research and Resource Center, University of Missouri, Columbia, MO.

Maintenance of human chromosome epigenetic states within a rodent cell has been demonstrated by using human-rodent somatic cell hybrids (SCH), including for the developmental status of globin gene expression, the active or inactive state of X chromosomes, and genomic imprinting. We have now used a pig-rodent SCH panel [Yerle M, et al. (1996) *Cytogenet Cell Genet*. 73: 194-202] to assess imprinted gene expression and DNA methylation of candidate imprinted loci. Furthermore, we are using SCH as a cellular model to explore genome editing of pig chromosomes. The SCH lines by cytogenetic analysis contain on average 5-7 pig chromosomes, for this work including *Sus scrofa* chromosome 1 (SSC1) or SSC9. Using FISH with a whole SSC1 paint probe and with a specific *SNORD116* gene locus probe, we confirmed that five SCH have a single SSC1q18 region [the location of the Prader-Willi syndrome (PWS)-orthologous imprinted locus]. Gene expression analysis showed that the PWS-orthologous genes *SNURF-SNRPN*, *SNORD116*, *SNORD107* and *MAGEL2*, as well as the syntenic *PLAGL1* gene, are differentially expressed in the SCH lines, consistent with an imprinted pattern of monoallelic expression. Using the methylation-sensitive restriction enzyme *Hha*I followed by PCR, the promoter of the bicistronic *SNURF-SNRPN* locus was unmethylated in three SCH and methylated in two SCH, correlating with gene expression and enabling assignment of paternal or maternal origin of the pig SSC1. Similarly, we are currently analyzing SSC9 candidate imprinted genes. Transfection of SCH lines with vectors encoding transcription activator-like effector nucleases (TALENs) or the CRISPR (clustered, regularly interspaced, short, palindromic repeats)/Cas9 ribonucleoprotein system designed to target sites in the PWS-orthologous domain resulted in successful induction of double-strand breaks and repair by non-homologous end joining to generate deletions ranging from 2.05 kb to ≥1.2 Mb. Levels of genome editing by both designer nucleases were greater on the transcriptionally active, paternal allele than on the silent, maternal allele, with only some CRISPR gRNA pairs functional for the latter. In conclusion, SCH are a valuable resource for imprinting assays of candidate genes in the pig. Further, SCH provide a powerful cell model to test efficacy and mechanisms of genome editing of pig chromosomes by designer nucleases. *Supported by NIH, FPWR, PWSA and the Storr Family Foundation.*.

455T

Analysis of monoallelic expression in 1084 human individual cells revealed novel putative imprinted genes. *C. Borel[1], F. Santoni[1], M. Garieri[1], E. Falconnet[1], P. Ribaux[1], S. Antonarakis[1,2,3].* 1) Dept Genetic Medicine, Univ Geneva Medical Sch, Geneva, Switzerland; 2) Geneva University Hospitals-HUG, Service of Genetic Medicine, Geneva, Switzerland; 3) iGE3 Institute of Genetics and Genomics of Geneva, University of Geneva, Switzerland.

Genomic imprinting results in the mutually exclusive expression of either the paternally or maternally inherited allele. Imprinted genes are implicated in the etiology of rare syndromes and have been associated with common diseases such as diabetes and cancer. More than 100 human imprinted genes have been identified so far, and the predicted number is in the range of few hundreds. We aim to identify novel imprinting genes using a single-cell (SC) RNA sequencing approach.From the proband of 2 family trio and 3 unrelated individuals, 1084 individual fibroblasts were RNA sequenced and more than 770'000 heterozygous SNVs were genotyped (WGS, 25x depth). For each gene, we analyzed the allelic specific expression through our in-house pipeline (>1 reads/site and >1SC/site). We modeled the likelihood of a gene to be monoallelically expressed with a beta-binomial distribution and evaluated the significance of the aggregate monoallelic ratio (reads sum of the most frequent allele per site / total reads) with the log-likelihood test. Genes presenting with a significant (adjusted p-value < 0.05) aggregate monoallelic ratio between 0.9-1 were retained for the identification of the allelic parent of origin. We were able to validate the imprinting status of 6 known imprinted genes and to detect 31 putative imprinted genes. We will also describe the allelic expression profile of 75 known imprinted genes in our single-cell dataset. Notably the analysis of 1084 single cells definitively improved the detection of average to low expressed genes in fibroblast with respect to bulk RNAseq allowing for a deep profiling of the allelic imbalance of each gene.In conclusion, single-cell RNA sequencing proves to be an excellent approach to overcome the intrinsic heterogeneity of the tissue and hereby it is a reliable tool to address the cell-type specificity of genomic imprinting.C.B. and F.S. contributed equally.

456F

Tissue-specific maps of genomic imprinting across the mammalian phylogeny reveal causal evolutionary pressures. *T. Babak.* Biology, Queen's University, Kingston, Ontario, Canada.

Genomic imprinting is an epigenetic process that restricts gene expression to either the maternally or paternally inherited allele. Many theories have been proposed to explain its evolutionary origin, but understanding has been limited by a paucity of data mapping the breadth and dynamics of imprinting within any organism. We previously showed, from comparisons of tissue-specific maps of imprinting in humans and mouse, thatimprinted genes were enriched for coexpressed pairs of maternally and paternally expressed genes, showed accelerated expression divergence between human and mouse, and were more highly expressed than their non-imprinted orthologs in other species (Babak et al. Nat Gen. 2015). Here we further those efforts by developing a new method for mapping imprinting in any species, even in the absence of genotyping data that is traditionally used to infer allele-specific gene expression. We applied this method to 6 major phylogenetic branches in the mammalian lineage and will discuss the evolutionary implications of the patterns that emerged.

457T

Genome-wide DNA methylation profiles in twins with autism spectrum disorders. *A. Anhalt[1], L.R. Simard[1], A. Hanlon-Dearman[2,3], X-Q. Liu[1,4].* 1) Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Manitoba, Canada; 2) Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, Manitoba, Canada; 3) Children's Hospital of Winnipeg, Winnipeg, Manitoba, Canada; 4) Department of Obstetrics, Gynecology, and Reproductive Sciences, University of Manitoba, Winnipeg, Manitoba, Canada.

Introduction: Autism spectrum disorders (ASD) are attributed to genetic, epigenetic, and environmental risk factors. Epigenetic factors (e.g. DNA methylation) are implicated in ASD because autistic symptoms are observed in some Mendelian disorders with an epigenetic aetiology (e.g. Rett syndrome). Twins are the ideal research subjects for epigenetic studies because they share similar pre- and post-natal environments and are of the same age. Moreover, identical (monozygotic, MZ) twins share the same genetic variants but can have different epigenetic profiles. Using twins, especially MZ twins, allows researchers to separate epigenetic effects from genetic and environmental factors. Two published studies used ASD-discordant MZ twins to identify associations between DNA methylation levels and ASD. However, additional studies can build upon these findings, due to the limited sample sizes in twin studies.**Methods and Results:** We recruited same-sex twin pairs with at least one twin diagnosed with ASD. Using the Illumina HumanMethylation450 BeadChip, we generated genome-wide DNA methylation profiles from buccal and peripheral blood mononuclear (PBMC) cells for twins from six pairs (2 MZ and 4 dizygotic, DZ, pairs). We examined the correlations of global DNA methylation levels between twins. In the PBMC samples, the MZ pairs were more similar than the DZ pairs, which were more similar than the unrelated (UR) pairs (e.g. using the 10% most variable probes; MZ: r=0.98, SD=0.005; DZ: r=0.93, SD=0.006; UR: r=0.88, SD=0.02). However, the correlations were similar between the ASD-discordant and ASD-concordant pairs with the same zygosity (e.g. MZ discordant: r=0.986; MZ concordant: r=0.980). Additionally, we examined site-specific DNA methylation in an ASD-discordant MZ pair, which revealed several CpG sites with large differences in DNA methylation levels between the twins.**Conclusions and Future Directions:** Our preliminary results suggest that ASD is not related to changes in the global DNA methylation profile. However, there may be significant site-specific differences in DNA methylation levels between ASD-affected and unaffected individuals. We will further investigate the site-specific relationships between DNA methylation levels and ASD and will examine whether differentially methylated regions can be detected in both PBMCs and buccal cells. We will also use RNA-seq data to examine the correlation between differentially methylated regions and gene expression levels.

458F

Longitudinal differences in DNA methylation profile of twins reared apart and together. *A. Ganna[1], Y. Wang[1], C. Wärnhjelm[1], X. Shen[1], T. Morris[2], C.A. Reynolds[3], N.L. Pedersen[1], S. Hägg[1].* 1) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet; 2) Medical Genomics Group, UCL Cancer Institute, University College London; 3) Department of Psychology, University of California-Riverside.

DNA methylation can be influenced by environmental and genetic factors. Twin studies can be used to evaluate the environmental contribution by controlling for the genetic effect. A study design that compares twin pairs reared together with twin pairs reared in separate families is particularly attractive because it can be used to investigate the contribution of common family environment. Several studies have suggested a modest contribution of common environment in phenotypic variation, but no previous studies have evaluated DNA methylation. In addition, no studies have examined how such differences change over time in the same individuals.We compared differences (Euclidean distances) in the overall methylation profile (Illumina HumanMethylation450 array, 428,545 probes) of 74 twin pairs (34% monozygotic - MZ) reared apart (i.e. separated before the age of 10) and 80 pairs (54% MZ) reared together from the Swedish Adoption Twin Study of Aging. Repeated measurements (max 5, mean 2.3 measurements per twin pair) were collected over a 20-years follow-up period, resulting in 710 samples with DNA methylation data. Pre-processing was done using Dasen normalization with adjustments for cell composition in blood and batch effectOverall within-pair differences in methylation profile increased over time (P-value < 0.0001, using mixed models) and were higher in DZ compared to MZ twins (p-value=0.05). Surprisingly, we found larger differences in the methylation profile of twin pairs reared together compared to those reared apart (P-value = 0.01). This effect was stronger for MZ pairs. Probe-specific analysis identified 2 probes (cg00344178, P-value=4.5x10-8 and cg10213695, P-value=2.6x10-8) that were positively associated with larger differences in MZ twins reared together after controlling for an FDR of 5%.We performed extensive sensitivity analysis by adjusting for differences in smoking, alcohol, BMI, cardiovascular disease, lipids levels, dementia, cognitive function, but the observed effect remained significant. Additional analysis assessing potential batch effects couldn't disprove our findings.Future replications needed to confirm our findings are planned in additional N=104 MZ twins.

459T

Is familial correlation in genome-wide DNA methylation level due to sharing the womb or genes? *S. Li[1], E.M. Wong[2], C. Apicella[1], J. Stone[3], G. Dite[1], J.E. Joo[2], G.G. Giles[4], M.C. Southey[2], J.L. Hopper[1].* 1) Centre for Epidemiology and Biostatistics, The University of Melbourne, Victoria 3010, Australia; 2) Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Victoria 3010, Australia; 3) Centre for Genetic Origins of Health and Disease, The University of Western Australia, Western Australia 6009, Australia; 4) Cancer Epidemiology Centre, Cancer Council Victoria, Victoria 3004, Australia.

Genome-wide DNA methylation level is studied as a risk factor for many diseases, but the genetic and environmental causes of its variation are less well studied. We measured DNA methylation from dried blood spots for 479 Australian women (mean age 56 years) comprising 66 MZ and 66 DZ twin pairs, and 215 of their sisters (130 families from Australian Mammographic Density Twins and Sisters Study), using the Illumina Infinium HumanMethylation450 BeadChip. Genome-wide DNA methylation level was defined as the average beta value across autosomal probes. We applied multivariate normal pedigree analysis using the software FISHER to adjust for age, estimate correlations, and fit variance components models where A refers to additive genetic factors, C to environmental factors shared by relatives (equally by twin pairs and to a lesser extent $s<1$ by sibling pairs), and E to individual factors. The estimated correlation in genome-wide DNA methylation level across all 468,406 probes was 0.42 (SE=0.10; $P=2\times10-5$) for MZ pairs, 0.40 (SE=0.10; $P=3\times10-5$) for DZ pairs, and 0.01(SE=0.06; $P=0.9$) for sibling pairs (including twin–sister pairs). After removing 184,424 probes with known genomic factors (SNPs, INDELs, *etc.*), the estimated correlations were 0.34 (SE=0.11; $P=2\times10-3$), 0.35 (SE=0.10; $P=4\times10-4$) and 0.04 (SE=0.05; $P=0.5$), respectively. For both measures, there was no difference between the MZ and DZ pair correlations (both $P>0.9$) and the twin pair correlations were greater than the sibling pair correlation (all $P<0.01$). Similar results were found for methylation levels in different genomic regions and chromosomes. For genome-wide methylation DNA level across all probes, we estimated A=2% (SE=11, $P=0.9$) and $s=0$; the only significant familial component was C, which explained 40% (SE=10, $P=10-4$; SE=11 when A=0) of the variance. We also found that, for non-twin sister pairs born after twins, the correlation was 0.47 (SE=0.15; $P=10-3$), which was greater than 0.01 (SE=0.19; $P=0.5$) for non-twin sister pairs born before twins ($P=0.05$). We conclude that there is no evidence from either the twin or sibling pair correlations that genetic factors explain variance in genome-wide DNA methylation level. Instead, the data are consistent with substantial variance due to non-genetic factors shared by twins alone. One explanation is that twins share the uterus at the same time. There is evidence that twin birth might influence the genome-wide methylation DNA level of subsequent sisters.

460F

Integrative Methods to Characterize X Chromosome Inactivation Patterns in Epithelial Ovarian Cancer. *S. Winham[1], N. Larson[1], Z. Fogarty[1], M. Larson[1], S. Armasu[1], B. Fridley[2], E. Goode[1].* 1) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 2) University of Kansas Medical Center, KUMC, Kansas City, KS.

In females, X chromosome inactivation (XCI), randomly silences transcription of one of the two homologous copies of the X chromosome via DNA methylation in order to equalize levels of gene expression with males. This process is tissue- and cell-specific, and some genes escape XCI. While studies suggest that inactivation is skewed (one chromosome is preferentially activated) in lymphocytes from epithelial ovarian cancer (EOC) patients, the role of XCI in tumors is largely unknown. Better understanding of chromosome and gene-level XCI patterns will advance our knowledge of ovarian tumor biology.To examine the role of XCI in EOC, we integrated genotype, gene expression, and DNA methylation data from tumors of 113 ovarian cancer patients. We measured allele-specific expression (ASE) for 481 genes across the X chromosome by combining RNA-Seq, and 1000 Genomes Project imputed genotype data to identify the active alleles for each tumor. Chromosome-wide, 89 patients showed skewed XCI based on a likelihood ratio test ($P<0.01$). Using a beta-binomial mixture model, we estimated which genes escaped XCI for each patient, and 22% of genes showed variable escape patterns across tumors. We compared the tumor escape status to previously reported lymphocyte escape status, and found that on average, 14% of genes were discordant within individuals; 61% of the discordances were previously reported lymphocyte escape genes. In one tumor-normal ovarian tissue pair, 19 of 68 expressed genes showed evidence of somatic changes, with 7 genes undergoing XCI in adjacent normal tissue but escaping in the tumor, including *TSC22D3* (MIM: 300506), a transcriptional regulator previously implicated in EOC.Because XCI is governed by DNA methylation, we also assessed beta values of the Infinium 450K Array CpG probes in X chromosome gene promoter regions, and observed cis associations with ASE imbalance. To assess global XCI patterns, we performed cluster analyses on the ASE data, identifying two ASE clusters. We also identified two clusters based on the promoter methylation data, and observed strong association with the ASE clusters ($P=0.007$). We examined the relationship between the clusters and clinical factors, and found XCI patterns were associated with tumor histology, stage, and overall survival ($P<0.04$).Together, these results suggest that XCI may play an important role in EOC. Future studies to examine somatic changes with paired tumor-normal tissue are needed.

461T

Oxytocin and Religious Brain. *J.R. Korenberg[1], L. Dai[2], J.B. King[3], M.A. Ferguson[3], J.A. Nielson[4], D. Giangrasso[3], C.S. Carter[5], H.P. Nazarloo[5], J.S. Anderson[6].* 1) Brain Institute, Department of Pediatrics, University of Utah, Salt Lake City, UT; 2) Department of Pediatrics, University of Utah, Salt Lake City, UT; 3) Neuroscience Graduate Student, University of Utah, Salt Lake City, UT; 4) Psychiatry and Psychology, Harvard University & Massachusetts General Hospital; 5) The Kinsey Institute, Indiana University, Bloomington, Bloomington, IN; 6) Department of Radiology, University of Utah, Salt Lake City, UT.

The molecular and neural mechanisms regulating human social behaviors are fundamentally important but largelyunknown. Religious experience is one of the most powerful transformative factors in the lives of individuals andcultures, and provocative evidence showed that religion influences social behaviors in different and even controversialways. It has been suggested that religious priming serves to enhance pro-sociality for in-group members, whileincreasing aggression towards out-group. It is interesting that social peptide hormone oxytocin, has been implicatedto play a role in cooperation and competition within and between groups. In addition, recent study showed thatreligion priming and an oxytocin receptor gene polymorphism interact to affect self-control in a social context. Wetherefore hypothesized that religiouspriming will result in an increase in pro-sociality as measured by observer ratingsof social stimuli, plasma serum levels of oxytocin, and functional connectivity in social brain regions. In this study, wereport the plasma oxytocin levels before and after a thirty-minute period of religious experience in a cohort of elevenreturned Mormon missionaries. Our preliminary results showed that the plasma oxytocin and vasopressin levels arehighly associated. This is the first time to study the in vivo relationship of peptide hormones in regulating religiousbelief and will enhance our understanding of neurobiology of the effects of religious experience on social behaviors.

462F

Val158Met polymorphism in COMT affects the brain's white matter properties during second language immersion. *P. Mamiya[1], T. Richards[2], B. Coe[3], E. Eichler[3], P. Kuhl[1].* 1) Institute for Learning & Brain Sciences, University of Washington, Seattle, WA; 2) Department of Radiology, University of Washington, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, wA.

Second language emersion as an adult is becoming increasingly prevalent as students prepare for college in a foreign country. There is a great variability in the level of proficiency among these late second language learners. Previous studies have demonstrated that adults learning a second language show changes in both the grey and white matter brain structures as a function of learning. It is unknown whether genetic variations influence how an individual's brain adapts while learning a second language. To address this issue, we studied Chinese students who enrolled in an English immersion program as preparation for college and investigated whether the Val158Met polymorphism in COMT can affect second language learning. Methionine (Met) substitution at codon 158 in COMT results in elevated dopamine in prefrontal cortex, and is associated with superior executive function, a high order cognitive ability known to be important for learning. We hypothesized that this polymorphism may have an effect on second language learning. Methods: We used real-time polymerase chain reaction method to identify the variant in COMT, and DTI to characterize the brain's white matter properties. We entered the COMT genotype and white matter data into a regression model to predict second language learning. Our first finding was that the Met/Val and Val/Val subjects show a significant positive correlation between the number of days they had been in the class at the time of the brain scan and enhancement of white matter properties in the voxels located in the superior longitudinal fasciculus (SLF), a fiber track important for language processing. This enhancement of white matter properties was shown as an increased fractional anisotropy (FA) and reduced radial diffusivity (RD), assessed by two DTI measures, in the SLF. This correlation was not found in the Met/Met subjects. Furthermore, the FA values in the SLF and the COMT genotype significantly predict the grades students received in the program. Comparing data from a regression model using only the FA value in the SLF to one in which we added the genetic data, we found an over 100% increase in the total variance explained. Taken together, our findings suggest that the COMT genotype may play a role in enhancing neuroplasticity in the brain's white-matter fiber tracks, enabling heightened student learning during exposure to second language learning in adulthood.

463T

Significant transcriptional changes in 15q duplication but not Angelman syndrome deletion dental pulp stem cell derived neurons. *N. Urraca[1,2], Q. Tran[5], T.G. Belgard[9], R. Memon[4], S. Goorha[2], C. Valdez[2], S. Sanchez[7], J.M. Ramirez[8], M. Donaldson[4], D. Bridges[3,6], L.T. Reiter[2,3].* 1) Le Bonheur Children's Hospital, Memphis, TN; 2) Department of Neurology, UTHSC, Memphis, TN; 3) Department of Pediatrics, UTHSC, Memphis, TN; 4) Department of Pediatric Dentistry, UTHSC, Memphis, TN; 5) Department of Preventive Medicine, UTHSC, Memphis, TN; 6) Department of Physiology, UTHSC, Memphis, TN; 7) Cytogenetics Laboratory, Instituto Nacional de Pediatria, Mexico City, Mexico; 8) Department of Biochemistry and Molecular Biology, University of Basque Country, Bilbao, Spain; 9) MRC Functional Genomic Unit, University of Oxford, UK.

The analysis of gene expression in living neurons from Angelman (AS) and Duplication 15q (Dup15q) syndrome subjects has impeded our understanding of these disorders. Here we use dental pulp stem cells (DPSC) from AS deletion, 15q Duplication and neurotypical control subjects for mRNA-seq analysis in both DPSC and DPSC-neurons. We identified 20 genes unique to AS neurons, 120 genes unique to 15q duplication and 3 shared genes that were differentially expressed versus control neurons. Copy number correlated with gene expression for most genes across the 15q11.2-q13.1 critical region. Two thirds of the genes differentially expressed in 15q duplication neurons were down-regulated compared to controls including several transcription factors, while in AS most differential expression was in the 15q region. Pathway analyses revealed steroid hormone biosynthesis genes were slightly enriched in both 15q duplication and AS neurons. We found more changes in gene expression in Dup15q than AS cell lines, perhaps because the mechanism of AS may be primarily through the degradation of UBE3A protein substrates. Comparison of gene expression changes in Dup15q DPSC neurons correlated with published postmortem idiopathic autism brain studies, indicating that this may be a valid *in vitro* model for gene expression changes in the brain. Finally, the finding of a significant increase in both *HERC2* and *UBE3A* in Dup15q neurons and significant decrease in these two genes in AS deletion neurons may explain differences between AS deletion class and *UBE3A* specific classes of AS mutation where *HERC2* is expressed at normal levels.

464F

***FXN* promoter silencing in the humanized mouse model of Friedreich ataxia.** *Y. Chutake[1], W. Costello[1], C. Lam[1], M. Pook[2], S. Bidichandani[1].* 1) Dept of Pediatrics, Section of Genetics, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Dept of Life Sciences, College of Health & Life Sciences, Brunel University, London, UK.

Friedreich ataxia is caused by an expanded GAA triplet-repeat sequence in intron 1 of the *FXN* gene that results in epigenetic silencing of the *FXN* promoter. This silencing mechanism is seen in patient-derived lymphoblastoid cells but it remains unknown if it is a widespread phenomenon affecting multiple tissues. The humanized mouse model of Friedreich ataxia (YG8sR) carries a single transgenic insert of the human *FXN* gene containing an expanded GAA triplet-repeat in intron 1. The latter mimics the human condition, and similar to people with Friedreich ataxia tissues from the YG8sR mouse are deficient for *FXN* transcript when compared to an isogenic transgenic mouse lacking the expanded repeat (Y47R). Tissues from YG8sR mice carried similar sized expanded GAA triplet-repeats and showed evidence of increased DNA methylation at the *FXN* locus. We found that tissues from YG8sR, but not Y47R, showed deficiency of *FXN* transcript that extended both upstream and downstream of the expanded GAA triplet-repeat, suggestive of deficient transcriptional initiation. This pattern of deficiency was seen in all tissues tested, irrespective of whether they are known to be affected (dorsal root ganglia, heart, cerebellum) or spared (cerebrum, skeletal muscle, fibroblasts) in disease pathogenesis. *FXN* promoter function was directly measured by quantitative analysis of transcriptional initiation via metabolic labeling of newly synthesized transcripts in fibroblasts, which revealed significant deficiency of transcriptional initiation in the YG8sR mouse compared to the Y47R control mouse. Thus, deficient transcriptional initiation accounts for the transcriptional deficiency in the humanized mouse model of Friedreich ataxia, similar to patient-derived cells, and the mechanism underlying promoter silencing in Friedreich ataxia is widespread across multiple tissues.

465T

Chromatin accessibility in the mammalian retina: Potential impact on studies of human retinal and macular degeneration. *F. Giuste[1], R. Ratnapriya[1], Y. Hyun-Jin[1], S. Montezuma[2], T. Longo[1], A. Parker[1], A. Police Reddy[1], D. Ferrington[2], A. Swaroop[1].* 1) Neurobiology, Neurodegeneration, & Repair Laboratory, NEI, NIH, Bethesda, MD; 2) Department of Ophthalmology & Visual Neurosciences, University of Minnesota Twin Cities, Minneapolis, MN.

A majority of variants associated with common disease reside in the non-coding region of the genome. To evaluate the nature of genomic elements harboring potential susceptibility variants associated with retinal and macular degeneration, we are developing a genome-wide chromatin map of the mammalian retina. The primary cause of blindness in a vast majority of degenerative retinal diseases is the dysfunction or loss of photoreceptors. Since it is difficult to obtain photoreceptor-specific chromatin data from human retina, we have also assayed open chromatin regions in flow-sorted mouse rod photoreceptors using ATAC-seq open chromatin assay. We are beginning to integrate ATAC-seq data of mouse rod photoreceptors and of human retina with respective global transcriptome profiles. Our analyses should provide important insights into genetic and epigenetic regulation of retinal gene expression and assist in identifying the potential impact of susceptibility variants on photoreceptor function and disease pathogenesis.

466F

Locally adaptive comparison of DNase I profiles to detect fine scale differences in regulatory activity. *J. Morrison[1], R. Sandstrom[2], N. Simon[1].* 1) Biostatistics, University of Washington, Seattle, WA; 2) Genome Sciences, University of Washington, Seattle, WA.

Variation in gene-expression and regulation leads to cell-level phenotypic variation from the scale of differences in cell identity to differences in drug sensitivity among cells of the same type. DNase I digestion experiments have been used to identify active regulatory DNA in numerous cell types. Different cell types can be characterized by sets of DNase hypersensitve sites (DHSs) - short regions with very high rates of DNase I cleavage. Analysis based on called DHSs is powerful for identifying and characterizing cell type specific and constitutive regulatory regions. However, DHS based analysis may be insufficient for characterizing more subtle phenotypic variation within a single cell type. This type of variation may be tied to a combination of dramatic and fine scale variation in DNase sensitivity. Summarizing DNase sensitivity as a series of DHSs discards information about quantitative differences and about differences in regions with lower sensitivity.In this work we explore a locally adaptive method for detecting and describing differences in chromatin accessibility by applying a recently developed method for comparing genomic profiles called Joint Adaptive Differential Estimate (JADE). JADE can be used to fit smooth DNase sensitivity profiles and identify regions where these profiles differ across phenotypic classes. In order to use JADE effectively, the smoothing strategy must be biologically informed and preserve meaningful variation while reducing the complexity of the problem. We explore adaptations of trend filtering for smoothing DNase I sensitivity data. We then apply JADE using these smoothing strategies to comparative analysis of DNase I profiles of small cell lung cancer cells classified as either sensitive or insensitive to drug therapy.

467T

ChIP-seq identifies the target genomic loci of the SMARCA6 helicase in human hematopoietic neoplasms. *J.A. Welch[1,2,3], D. Esopi[2], P. Shen[3], P. Schaughency[2,4], A. Skaist[2,4], S. Yegnasubramanian[2,4], K.H. Burns[1,3], S.J. Wheelan[1,2,4].* 1) McKusick-Nathans Institute of Genetic Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Oncology, The Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Pathology, The Johns Hopkins University School of Medicine, Baltimore, MD; 4) Next Generation Sequencing Center, Sidney Kimmel Comprehensive Cancer Center, Baltimore, MD.

The SMARCA6 helicase is a chromatin regulator that is thought to be important for somatic retrotransposon silencing. Like other Snf helicases, which promote transcriptional silencing in mammalian genomes, SMARCA6 has demonstrated DNA-dependent ATPase activity *in vitro*. SMARCA6 is known to interact with DNA methyltransferases DNMT1 and DNMT3B and histone deacetylases HDAC1 and HDAC2. In mice, chromatin immunoprecipitation has demonstrated an interaction between SMARCA6 and LINE, SINE, and IAP retrotransposons. Evidence from transgenic mouse models indicates that SMARCA6 function is required for retroelement silencing. Mice that are homozygous for a null mutation show decreased methylation of LINE, SINE, and IAP elements as well as derepressive H3 and H4 acetylation marks in IAPs and exhibit perinatal lethality. Defects in retroelement silencing are phenocopied in hypomorphic mice with targeted deletions of conserved helicase-encoding exons, and while viable, these animals display growth defects and age prematurely. Whether or not SMARCA6 is involved in retrotransposon silencing and localizes to retroelements in the human genome are open questions. The availability of a commercially manufactured polyclonal antibody specific to human SMARCA6 and the development of computational strategies to identify retrotransposons in high throughput sequencing data have positioned us to investigate the latter. To this end, we performed chromatin immunoprecipitation followed by sequencing (ChIP-seq) to identify the target loci of SMARCA6 in a human Burkitt lymphoma cell line. Here, we discuss our findings from these experiments.

468W

Morpholino oligomers as a potential therapeutic option for correction of defective alternative splicing. *S. Tantzer[1], K. Sperle[1], G.M. Hobson[1,2].* 1) Nemours Biomedical Research, Alfred I. duPont Hospital for Children, Wilmington, DE; 2) Pediatrics, Thomas Jefferson University, Philadelphia, PA.

Most mutations of the proteolipid protein 1 gene (*PLP1* [MIM 300401]), including copy number changes and those affecting the coding sequence, cause the X-linked leukodystrophy Pelizaeus-Merzbacher disease (PMD [MIM 312080]). We recently showed that mutations reducing the *PLP1*/*DM20* alternative splice ratio have a distinctly different early phenotype on MRI called hypomyelination of early myelinating structures (HEMS). Although there are currently no cures for HEMS or the many other diseases that can be caused by disruption of alterative splicing, molecules that modulate aberrantly spliced products toward a more normal ratio have exciting therapeutic potential. We designed a morpholino oligomer (MO-PLP) to target the *DM20* 5' splice site, thereby increasing the *PLP1*/*DM20* ratio. Treatment of an immature oligodendrocyte cell line, Oli-neu, with MO-PLP significantly increased the endogenous *Plp1*/*Dm20* ratio as well as the *PLP1*/*DM20* ratio expressed from transfected minigene splicing constructs containing patient mutations known to reduce the *PLP1*/*DM20* ratio. Our previously published mouse model, PLP-ISEdel, carries one of these patient mutations, an intronic 19 base pair deletion of a splicing enhancer that causes a decrease in the *PLP1*/*DM20* ratio. A single intracerebroventricular injection of MO-PLP into neonatal PLP-ISEdel mice resulted in an increase in PLP/DM20 ratio of both RNA and protein in both brain and spinal cord. The effect lasted up to postnatal day 60, which is well beyond the early postnatal spike in myelination and PLP production. Our results suggest that morpholino oligomers have therapeutic potential for the treatment of HEMS and other diseases caused by misregulation of alternative splicing.

469T

Seeking drugs for Pelizaeus-Merzbacher disease using drug repositioning approach targeting a novel cellular pathology. *K. Inoue[1], H. Li[1], P.R. Mangalika[1], A. Nishizawa[1], Y. Numata[1], S. Nakamura[1], T. Morimura[1], H. Saya[2], Y. Goto[1].* 1) Dept MR & BD Res, Natl Inst Neurosci, NCNP, Kodaira, Tokyo, Japan; 2) Div Gene Regulation, Inst Advanced Medical Research, Keio Univ, Tokyo, Japan.

Pelizaeus-Merzbacher disease (PMD) is the most common, but incurable inherited hypomyelinating leukodystrophy caused by various mutations in the *PLP1* gene, which encodes a major myelin membrane lipid-protein. PLP1 amino acid substitutions often cause severe form of PMD. Although it has been thought that activation of the apoptotic pathway of unfolded protein response (UPR) triggered by accumulation of the mutant PLP1 in endoplasmic reticulum (ER) is the major cellular mechanism for oligodendrocyte cell death, attenuation of this pathway doesn't rescue the cellular phenotypes, suggesting an involvement of other underlying mechanisms. We recently identified that mutant PLP1 also interfere with intracellular transport of other normal membrane and secretary proteins, possibly affecting the normal cellular physiology. This finding has raised a hypothesis that the mutant PLP1-induced dysfunction of secretary pathway may serve as another cellular mechanism for PMD and the drugs that can enhance the transport of secretary and membrane proteins may rescue the cellular phenotype caused by mutant PLP1. To identify such drugs, we have established a screening system using HeLa cells transiently co-expressing mutant PLP1 and secretable-form luciferase reporter. We have screened a library of 1400 existing medicine, so that our findings are immediately applicable to the clinical study. We next tested if the selected compounds can also rescue the abnormal intracellular localization of KDEL receptor that is normally present in Golgi apparatus but is mislocalized in ER when co-expressed with mutant PLP1. Using these screening systems, we have identified 20 compounds that can correct the reduced intracellular transport of normal secretary and membrane proteins. These compounds may serve as potential candidates for the treatment of PMD caused by amino acid substitutions.

470W

Rapamycin as a novel treatment for vascular malformations caused by somatic mutations that activate the PI3K>AKT signaling pathway. *L. Boon[1,2], J. Hammer[1], E. Seront[3], S. Dupont[4], F. Hammer[5], Ph. Clapuyt[6], M. Vikkula[2,7].* 1) Centre Vascular Anomalies, Division of Plastic and Reconstruction Surgery, Cliniques universitaires St-Luc, Università catholique de Louvain, Brussels, Belgium; 2) Human Molecular Genetics, de Duve Institute, Università catholique de Louvain, Brussels, Belgium; 3) Department of Medical Oncology, Cliniques universitaires St-Luc, Università catholique de Louvain, Brussels, Belgium; 4) Pediatric Oncology, Cliniques universitaires St-Luc, Università catholique de Louvain, Brussels, Belgium; 5) Department of Interventional Radiology, Cliniques universitaires St-Luc, Università catholique de Louvain, Brussels, Belgium; 6) Pediatric Radiology, Cliniques universitaires St-Luc, Università catholique de Louvain, Brussels, Belgium; 7) WELBIO (Walloon Excellence in Lifesciences and Biotechnology), de Duve Institute, Università catholique de Louvain, Brussels, Belgium.

Background: Venous and lymphatic malformations (VM and LM) are congenital lesions composed of ectatic veins or lymphatic cysts, respectively. They have an incidence estimated at 1/2.000 and 1/10.000, respectively. They often cause deformity, pain, local intravascular coagulopathy or chronic infection. Inherited VMs (VMCM) are caused by germline TIE2 mutations, whereas sporadic VMs are often caused by somatic TIE2 mutations. Sporadic LMs and combined malformations are caused by somatic PIK3CA mutations.**Purpose:** The identification of the molecular bases and activation of the downstream PI3K-AKT signaling pathway opened the door for development of precision therapies. We assessed the efficacy and safety of Rapamycin, an mTor inhibitor, as a possible therapy for patients with difficult-to-treat slow-flow malformations.**Methods:** The trial is registered under VASCA-LM at clinicaltrials.gov (NCT01811667). Eighteen patients aged from 3 to 64 years with refractory-to-standard-care LM, VM and/or complex vascular anomaly were enrolled. Informed consent was obtained. Clinical symptoms included chronic daily debilitating pain, functional impairment, recurrent infections, and daily gastrointestinal bleeding. Efficacy was evaluated by anamnesis of symptoms, repeated evaluation of pain, quality of life and clinical parameters, and photographs of the lesions. Blood tests included D-dimer and fibrinogen measurements. A global self-evaluation percentage of increase/decrease in quality of life was recorded, as well as side effects, according to CTCAE version 3. Volumetric MRI was performed before initiation and at a yearly bases.**Results:** Fifteen patients reached 12 months follow-up. All patients with LMs, 6/7 patients with VMs and all patients with complex vascular anomalies experienced important relief of pain and symptoms, improved functional restraint and self-perceived quality of life. A statistically significant reduction in volume was observed with MRI with patients that reached one-year follow-up. Known side effects were not infrequent.**Conclusion:** Results from this clinical pilot study demonstrate that rapamycin can be a therapeutic option for refractory-to-standard-care slow-flow vascular malformations, which classically exhibit PI3K>AKT activation. Similar results were seen in a mouse model of VMs generated by activating TIE2 mutations. Therefore, rapamycin has great promise as the first molecular therapy for difficult-to-treat slow-flow vascular malformations.

471T

The Inflammasome: A Novel Concept in VCP Disease. *A. Nalbandian[1], A. Khan[2], R. Srivastava[2], K. Llewellyn[1], V. Kimonis[1], L. BenMohamed[2].* 1) University of California, Irvine; 2) Gavin Herbert Eye Institute.

Valosin Containing Protein (VCP) disease, discovered in 2000, is caused by mutations in the gene encoding *Valosin-Containing Protein (VCP)*. The spectrum of phenotypes associated with VCP mutations continues to expand, including disorders with overlapping pathogenesis, such as Inclusion Body Myositis (IBM) in 80%, Paget's Disease of Bone (PDB) in 50%, Frontotemporal Dementia (FTD) in 33%, and Amyotrophic Lateral Sclerosis (ALS) in 15% individuals. VCP disease patients exhibit progressive proximal limb girdle muscular weakness and eventually die from progressive muscle weakness and cardiac and respiratory failure. Histologically, patients show the presence of rimmed vacuoles and TAR DNA-binding protein 43 (TDP-43)-positive large ubiquitinated inclusion bodies in the muscles. Several studies have investigated the onset of VCP-associated myopathy, however to our knowledge, we are the first group to explore the inflammasome signaling network in this disease. We discovered induction of NLRP3-mediated inflammasome in inflammatory macrophages and neutrophils may possibly contribute to inflammation and muscle-related tissue damage observed in VCP disease. Such inflammation might be triggered by ROS, NOS or, other yet-to-be determined inflammatory mechanisms, in damaged muscle tissues; and blocking one or several of pathways of NLRP3 inflammasome checkpoints in VCPR155H/+ heterozygote mice may reduce muscle-related inflammation and tissue damage resulting in robust and sustained therapeutic effect. We also detected an increase in the IL-1b inflammatory cytokine and Caspase 1. These findings suggest that targeting the inflammasome pathway by pharmacological inhibitors may ameliorate an increasing list of disorders and can now include in VCP disease and related neurodegenerative multisystem proteinopathies.These findings suggest that targeting the mTOR pathway and restoring autophagic flux ameliorate an increasing list of disorders and can now include in VCP disease and related neurodegenerative multisystem proteinopathies. In conclusion, this pre-clinical study in VCPR155H/+ heterozygote mice may lead to a novel and powerful immunotherapy for patients suffering from VCP- and associated neurodegenerative diseases.

472W

Efficacy and Safety of Diazoxide Choline Controlled-Release Tablet in Patients with Prader-Willi syndrome. *A. Surampalli[1], M. Wencel[1], JA. Gold[1,2,3], NM. Cowen[4], V. Kimonis[1].* 1) Division of Genetics and Genomics, Department of Pediatrics, University of California Irvine, School of Medicine, Irvine, CA; 2) Loma Linda University School of Medicine, Loma Linda, CA; 3) Children's Hospital of Orange County, Orange, CA; 4) Essentialis, Inc. Carlsbad, CA.

Prader-Willi syndrome (PWS) is a complex, multi-system genetic syndrome affecting 1:10,000-1:30,000 males, females and all ethnicities. PWS is the most common syndromic form of obesity. PWS patients have a characteristic behavior disorder, mild intellectual disability, hypotonia, hypothalamic-pituitary axis dysfunction, growth hormone (GH) deficiency, hypogonadism, other hormonal deficiencies, short stature, centrally- hyperphagia, obesity, reduced lean mass and reduced energy expenditure. Diazoxide choline controlled-release tablets (DCCR) works by mimicking sensitivity to leptin and insulin in AgRP and POMC neurons, reducing de-novo fatty acid biosynthesis, stimulating β-oxidation of fat and normalizing GABA response, reducing hyperphagia, fat mass and aggressive behavior in PWS patients.Thirteen obese subjects with genetically confirmed diagnosis of PWS, between 10-22yrs were enrolled in the study PC025. Two subjects were withdrawn from the study due to a prior psychiatric illness in one and progressively compromised glycemic control at the highest dose in the other. Eleven subjects completed the open-label treatment phase, were designated as Responders (hyperphagia response rate 92%) and randomized into the double-blind phase of the study. Data was analyzed for both open-label treatment phase and double-blind phase.We found significant improvements in hyperphagia in DCCR treated subjects at the end of the open-label treatment period (-31.6%, p=0.003) and in those who continued on DCCR in the double blind phase (-29.2%, p=0.006). There was significant reduction in 'aggressive', 'threatening' , 'destructive' behavior in comparison to all other PWS associated behaviors at the end of open label treatment phase (62.5% Vs 29.8% respectively, p=0.01). The effect of DCCR on behavior seems to be independent of the effect on hyperphagia. Significant impacts were seen on fat mass (-3.8%, p=0.011), lean body mass (+5.4%, p=0.001), lean body mass/fat mass ratio (+9.8%, p=0.002) during open-label treatment. The impacts on body composition were of similar magnitude in GH treated and GH naïve subjects. DCCR treatment significantly reduced TG, Tot-C, and Non-HDL-C in both the open-label and double blind phase.The safety profile of DCCR in PWS subjects was consistent with prior studies. The most common adverse events included peripheral edema which was responsive to dose reduction and diuretics, and transient hyperglycemia.

473T

Histone Deacetylase Inhibitor Reverses Promoter Silencing in Friedreich Ataxia. *S. Bidichandani[1], Y. Chutake[1], C. Lam[1], W. Costello[1], M. Anderson[2].* 1) Dept of Pediatrics, Section of Genetics, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Dept of Biostatistics & Epidemiology, University of Oklahoma Health Sciences Center, Oklahoma City, OK.

Friedreich ataxia (FRDA) is caused by an expanded GAA triplet-repeat (GAA-TR) mutation in the *FXN* gene, and patients are typically homozygous for expanded alleles containing 100–1300 triplets. Patients have a severe deficiency of *FXN* transcript, which is predominantly caused by epigenetic silencing of the *FXN* promoter. We sought to determine if a class I histone deacetylase (HDAC) inhibitor that is currently being developed as a rational therapy for FRDA is able to reverse *FXN* promoter silencing. Patient-derived lymphoblastoid cell lines bearing a range of expanded alleles (200–1122 triplets) were evaluated for *FXN* transcript levels by quantitative RT-PCR following treatment with the 2-aminobenzamide derivative drug, 109 or a related but inactive compound, 966. *FXN* transcript levels increased ~2-fold ($p<0.001$) upon treatment with 109 (but not 966), both upstream and downstream of the expanded GAA-TR mutation, suggesting that the increase is due to improved promoter function. *FXN* promoter function was directly measured by quantitative analysis of transcriptional initiation via metabolic labeling of newly synthesized transcripts in living cells, which revealed that 109 (but not 966) resulted in a ~3-fold increase in both transcriptional initiation ($p<0.01$). *FXN* promoter structure was measured via Nucleosome Occupancy and Methylome Sequencing (NOMe-Seq), a high-resolution *in vivo* footprint assay for nucleosome occupancy in individual chromatin fibers, which revealed that 109 (but not 966) resulted in a ~3-fold increase in chromatin accessibility ($p<0.0001$) at the *FXN* promoter. Epigenetic promoter silencing in FRDA is therefore reversible, and the results implicate class I HDACs in repeat-mediated promoter silencing.

474W

Phase 3 and long-term extension study with migalastat, a pharmacological chaperone, demonstrate stable renal function, reduced left ventricular mass and gastrointestinal symptom improvement in patients with Fabry disease. *D.P. Germain[1], D.G. Bichet[2], R. Giugliani[3], D. Hughes[4], R. Schiffmann[5], W. Wilcox[6], J. Castelli[7], E.R. Benjamin[7], J. Yu[7], N. Skuban[7], J. Barth[7].* 1) Division of Medical Genetics, University of Versailles, Montigny, France; 2) Hôpital Sacré-Coeur, University of Montreal, Canada; 3) Medical Genetics Service, HCPA/UFRGS Porto Alegre, Brazil; 4) Royal Free Campus, University College London, UK; 5) Baylor Research Institute, Dallas, TX; 6) Department of Human Genetics, Emory University, Georgia; 7) Amicus Therapeutics, Cranbury, NJ, USA.

Background: Fabry disease (FD) is a progressive, devastating X-linked genetic disorder caused by the deficient activity of lysosomal alpha-galactosidase. **Objectives**: Migalastat (1-deoxygalactonojirimycin, AT1001) is an orally-administered investigational pharmacological chaperone for Fabry disease that selectively binds and stabilizes α-Gal A, leading to increased lysosomal activity. Study 011 (AT1001-011, NCT00925301) included a 6-month double-blind placebo-controlled period (Stage 1) and an 18 month open-label extension. Sixty-seven subjects were randomized; 48 subjects completed the study and continued in long term extension Study AT1001-041. **Methods**: Estimated GFR (eGFR) was assessed every 3-6 months using the CKD-EPI and MDRD equations. Left ventricular mass index (LVMi) was assessed by echocardiography every 6-12 months by a blinded central laboratory. Gastrointestinal symptoms were assessed every 6 months using the Gastrointestinal Symptom Rating Scale (GSRS). p-values are unadjusted for multiple comparisons. Efficacy results are reported for the 50 subjects with amenable *GLA* mutations. **Results**: eGFR remained stable in subjects treated up to 48 months, with mean annualized eGFR changes (±SEM) of -0.80±0.60 (CKD-EPI) and +0.70±0.80 (MDRD) mL/min/1.73m2/yr. LVMi was significantly reduced in subjects treated up to 48 months (-8.0 ±5.5 g/m2); the largest reductions were observed in subjects with abnormal LVMi at baseline (-17.2±16.5 g/m2). Improvement was observed in the diarrhea domain of GSRS during the 6-month double-blind period (-0.3 migalastat, +0.2 placebo, p=0.03), and in diarrhea, indigestion and reflux domains during the open-label extension. **Conclusions**: Treatment with migalastat for up to 48 months was associated with stable renal function, reduced LVMi and improved gastrointestinal symptoms in patients with Fabry disease who have amenable *GLA* mutations.

475T

***Hydroxyurea induces γ-globin expression through microRNAs-mediated actions in human hematopoietic and K562 erythroleukemia cells.*.** *D.G. Pule[1], S. Mowla[2], N. Novitzky[2], A. Wonkam[1].* 1) Human Genetics, Clinical Laboratory Sciences, University of Cape Town, Cape Town, Western Cape, South Africa; 2) Hematology, Clinical Laboratory Sciences, University of Cape Town, Cape Town, Western Cape, South Africa.

Background: Hydroxyurea is the only available drug treatment for Sickle Cell Disease (SCD); but the mechanism of fetal hemoglobin (HbF) induction is poorly understood. In particular, the role of microRNAs in the post-transcriptional regulation of HbF, through potent regulators such as *MYB, BCL11A* and *KLF-1,* is understudied.**Objectives:** We have investigated the role of microRNAs in the post-transcriptional regulation of HbF in erythroid cells treated with hydroxyurea.**Methods:** As models, we used 1) hematopoietic stem cells derived from umbilical cord blood (HSCs) and 2) the K562 human erythroleukemia cell line. Both primary erythroid and K562 cells were treated with hydroxyurea and changes in *BCL11A*, *KLF-1*, *GATA-1*, *MYB,* β- and γ-globin gene expression and 7 targeted miRNAs, previously shown to be modified by hydroxyurea or associated with basal γ-globin expression were also investigated. **Results:** We differentiated CD34+ stem cells into erythroid precursors using a single-phase expansion and differentiation protocol. *BCL11A* was down-regulated and there was a marked 7-fold increase ($p < 0.003$) in γ-globin expression in both primary human erythroid and K562 cells. Down-regulation of *GATA-1* was also consistent with the above findings. Similarly, *KLF-1* was down-regulated in both cell models, corresponding to the repressed expression of *BCL11A* and β-globin gene ($p < 0.04$). In both cell models, most miRNAs was significantly up-regulated by hydroxyurea with some level of variation. Down-regulation of, another potent negative regulator of γ-globin and direct target of the up-regulated miR-15a and miR-16-1, provides a post-transcriptional tier of HbF regulation. We also demonstrated that the inhibition of miR-26b; miR-151-3p and miR-451 resulted in an apparent decrease in HbF.**Conclusions:** The data provide evidence of a microRNAs-mediated post-transcriptional mechanism for γ-globin regulation under hydroxyurea treatment. These findings will add a unique piece to our understanding of HbF regulation and incite further investigation of the post-transcriptional regulatory network for HbF expression.

476W

High-content screening for small molecule inhibitors of facioscapulohumeral dystrophy. *S.J. Palmer[1], L. Caron[2], E. Eiffe[1], A. Heaton[1], U. Schmidt[2].* 1) Novogen Laboratories Ltd, 16-20 Edgeworth David Ave, Hornsby, Sydney, NSW 2077 Australia; 2) Genea Biocells, Level 3, 321 Kent St, Sydney NSW 2000 Australia.

Facioscapulohumeral muscular dystrophy (FSHD1 [MIM 158900] and FSHD2 [MIM 158901]) is the third most common form of MD with a prevalence of 1:14,000-20,000. It initially weakens the muscles of the face and upper limb girdle, and progresses to affect most skeletal muscles. It is caused by a deletion in a subset of the D4Z4 macrosatellite repeats in the subtelomeric region of chromosome 4 (FSHD1) or a mutation in the *SMCHD1* gene (FSHD2) that regulates the epigenetic status of the D4Z4 locus. Both forms of the disease result in decreased epigenetic repression of the locus and myogenic mis-expression of the *DUX4* retrogene located in the D4Z4 repeats. Evidence indicates that the DUX4 double homeobox transcription factor initiates a cascade of inappropriate gene expression that inhibits myogenesis, increases oxidative stress, induces muscle atrophy and activates autoimmune responses. In order to develop an efficient in vitro screening system to identify novel small molecule drugs combatting the disease, embryonic stem (ES) cell lines were isolated from human embryos diagnosed with FSHD1 as part of the preimplantation genetic diagnosis service for FSHD-affected families. FSHD ES cell lines were grown in myogenic differentiation media to create myotubes that model aspects of the adult disease. In comparison with controls, the FSHD ES lines demonstrated detectable *DUX4* expression, attenuated myotube morphological parameters and an altered transcriptional profile. Correction of these criteria in vitro provides a basis for high-content screening of novel small molecule therapeutics. A large body of evidence indicates that the isoflavone molecule, genistein, can permanently modify epigenetic marks in the genome, which was elegantly demonstrated in mice carrying the epigenetic reporter allele, *Avy*. Maternal genistein supplementation increased repeat-mediated gene silencing of the locus and increased the frequency of the pseudoagouti coat colour in the offspring. Therefore, genistein and related isoflavonoid structures could increase repeat-mediated gene silencing of the D4Z4 locus in stem cells within the skeletal muscle niche, thus blocking *DUX4* expression and halting the progress of FSHD. We screened a series of novel genistein-related isoflavonoids using the FSHD ES cell assay described, which led to the identification of several lead compounds that fulfil the rescue criteria and could form the basis of the first drug treatment for this currently incurable disease.

477T

The small molecule aldehyde trap NS2 represents a pharmacological approach to enzyme replacement therapy for SSADH deficiency (SSADHD). *G.R. Ainslie[1], K.R. Vogel[1], V.C. Cullen[2], S.L. Young[2], T.C. Brady[2], K.M. Gibson[1].* 1) Experimental and Systems Pharmacology , Washington State University, Spokane, WA., USA; 2) Aldeyra Therapeutics, Inc., Lexington, MA, USA.

Succinic semialdehyde dehydrogenase (SSADH; ALDH5A1) deficiency (SSADHD; MIM 271980) is the most prevalent inherited disorder of 4-aminobutyrate (GABA) degradation. The phenotype includes developmental delay, hypotonia and expressive language impairment in childhood, and various neuropsychiatric aspects, including obsessive-compulsive disorder, in adolescence and adulthood. Epilepsy occurs in 50% of patients. Accumulation of the enzyme substrate succinic semialdehyde (SSA) associates with elevations in both GABA and the related neuromodulator, gamma hydroxybutyrate (GHB), as well as a number of other potentially toxic metabolites, including the aldehyde 4-hydroxy-2*E*-nonenal (4-HNE). Current treatment approaches target GABAergic signaling, with mixed results. Aldeyra Therapeutics has developed a small molecule aldehyde trap, NS2, which has an excellent safety profile and which is currently in a clinical trial for another genetic disease of aldehyde metabolism (Sjogren-Larsson Syndrome). We hypothesized that NS2 can trap SSA and therefore prevent its accumulation, thus indirectly lowering downstream disease markers such as GABA, GHB and 4-HNE. In this way, NS2 represents a novel therapeutic approach to SSADH deficiency by targeting the disease at its root cause. Here, we show that NS2 can react *in situ* with SSA, and additionally with 4-HNE, forming the NS2-SSA and NS2-4HNE adducts, respectively, which can be detected by LC MS/MS. This aldehyde trapping was shown to occur under physiological pH conditions, and was rapid, attaining completion in 1.5-2 hours. Initial pharmacokinetic studies in mice completely deficient for SSADH demonstrated that, after a single intraperitoneal bolus administration at 10 mg/kg, NS2 attained micromolar concentration in serum, and reached all target organs tested (brain, liver, kidney). Moreover, NS2 was able to trap SSA *in vivo* in these mice, as confirmed by the detection of the NS2-SSA adduct in target tissues 8 hours post-dose. These studies form the basis for testing NS2 in a repeated-dosing paradigm (previously shown to be non-toxic in other rodent studies) for improvement of metabolic, neurological, and lifespan outcomes. If successful, these preclinical studies may lead to trials of NS2 in SSADH-deficient patients. Aldehyde trapping may thus represent a novel, pharmacological enzyme replacement therapy for SSADH deficiency and other Mendelian disorders in which reactive aldehydes accumulate.

478W

Effects of oral eliglustat on bone parameters in treatment-naïve patients with Gaucher disease type 1: results from the phase 3, randomized, placebo-controlled ENGAGE trial after 18 months. *P. Mistry[1], D.J. Amato[2], M. Dasouki[3], S. Packman[4], G.M. Pastores[5], S. Assouline[6], M. Balwani[7], A. Ortega[8], S.S. Shankar[9], M.H. Solano[10], L.H. Ross[11], J. Angell[11], M.J. Peterschmitt[11].* 1) Yale University School of Medicine, New Haven, CT; 2) Mount Sinai Hospital, Toronto, Canada; 3) University of Kansas Hospital, Kansas City, KS; 4) UCSF School of Medicine, San Francisco, CA; 5) New York University School of Medicine, New York, NY; 6) Jewish General Hospital, Montreal, Quebec, Canada; 7) Mt. Sinai Hospital, New York, NY; 8) OCA Hospital, Monterrey, Mexico; 9) Emory University, Atlanta, GA; 10) Hospital de San Jose-Fundacion de Ciencias de la Salud, Bogota, Colombia; 11) Genzyme, a Sanofi company, Cambridge, MA.

Background: Manifestations of Gaucher disease type 1 (GD1) include splenomegaly, hepatomegaly, anemia, and thrombocytopenia. Patients are also at risk of bone complications, including decreased bone mineral density, increased risk of pathogenic fractures and joint collapse, and chronic bone pain. Eliglustat is an oral substrate reduction therapy approved for the treatment of adults with GD1. Safety and efficacy of oral eliglustat were evaluated in the placebo-controlled ENGAGE trial (NCT00891202; sponsored by Genzyme, a Sanofi company). All primary and secondary endpoints were met in the 9-month primary analysis period (Mistry et al., *JAMA* 2015;311:695-706).**Methods:**After the primary analysis, 39 of 40 patients entered a 9-month, open-label extension phase in which all patients received eliglustat. Bone evaluations during both phases of the trial included changes from baseline in bone marrow infiltration, bone marrow burden (BMB) score, and lumbar spine T- and Z-scores.**Results:**At baseline, 80% (16/20) of eliglustat patients and 75% (15/20) of placebo patients had marked to severe total BMB scores. After 9 months, significant improvements (eliglustat vs. placebo) were observed for spine and femur BMB scores; spine Z-scores also improved. Among patients receiving eliglustat for 18 months who had bone data available (n=18), improvements continued during the extension period (mean changes from baseline: total BMB: -2.15; lumbar spine T-score: +0.19, lumbar spine Z-score=+0.26). For patients switched from placebo to eliglustat (n=20), mean changes from baseline after 9 months of eliglustat were: total BMB: -0.94, lumbar spine T-score: +0.03, and lumbar spine Z-score: +0.03. No new safety concerns were identified. **Conclusions:**Patients receiving eliglustat for 18 months continued to show improvement in bone parameters beyond the initial 9-month primary analysis period.

479T

Stability is maintained in adults with Gaucher disease type 1 switched from velaglucerase to eliglustat or imiglucerase: a sub-analysis of the eliglustat ENCORE trial. *R. Pleat[1], T.M. Cox[2], T.A. Burrow[3], P. Giraldo[4], O. Goker-Alpan[5], B.E. Rosenbloom[6], J. Angell[1], L. Croal[1], L. Underhill[1], M.J. Peterschmitt[1].* 1) Genzyme, a Sanofi company, Cambridge, MA; 2) Department of Medicine, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK; 3) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 4) Centro de Investigación Biomédica en Red de Enfermedades Raras, (CIBERER), Zaragoza, Spain; 5) Lysosomal Disorders Research and Treatment Unit, Center for Clinical Trials, O&O Alpan, LLC, Springfield, VA; 6) Cedars-Sinai/Tower Hematology Oncology, Beverly Hills, CA.

Introduction: Gaucher disease type 1 (GD1) is an autosomal recessive disorder caused by deficiency of the lysosomal enzyme acid-β-glucosidase resulting in accumulation of glucosylceramide. Clinical manifestations include anemia, thrombocytopenia, hepatosplenomegaly, and skeletal disease. The current standard of care is enzyme replacement therapy (ERT). Eliglustat, an oral substrate reduction therapy, was recently approved for the long-term, first-line treatment of adults with GD1 who are extensive, intermediate or poor CYP2D6 metabolizers. The 12-month ENCORE trial (NCT00943111; Genzyme) found oral eliglustat non-inferior to imiglucerase in maintaining disease stability in adults with GD1 who had achieved pre-specified therapeutic goals after ≥3 years of ERT (Cox, *Lancet* 2015:25Mar.pii:S0140-6736(14)61841-9).**Methods:**This post-hoc analysis examined safety and efficacy in the 30 ENCORE patients whose treatment at study entry was velaglucerase alfa: 22 of whom were randomized to eliglustat and 8 to imiglucerase. The composite primary efficacy endpoint was the percent of patients meeting all 4 pre-specified stability parameters in spleen (≤25% increase from baseline [BL]), liver (≤20% increase from BL), hemoglobin (≤1.5 g/dL decrease from BL) and platelets (≤25% decrease from BL).**Results:**Mean duration of velaglucerase treatment prior to study entry was 1.34 years at a mean dose of 49.9 U/kg/2weeks. After 12 months of eliglustat, 90% of velaglucerase-transitioned patients maintained all four stability goals; separately, 95% of patients maintained goals for hemoglobin and platelets and 100% of patients for spleen and liver. Mean ± SD changes from BL for patients transitioned to eliglustat were: spleen volume (multiples of normal [MN]), -1% ± 8.7; liver volume (MN), 1.6% ± 7.9; hemoglobin (g/dL), -0.42 ± 0.62; and platelets, 1% ± 25.9. Data were comparable in velaglucerase-transitioned patients switched to imiglucerase. Among all velaglucerase-transitioned patients, most adverse events were mild/moderate; there were no treatment-related serious adverse events, and 1 eliglustat patient discontinued treatment due to an adverse event (palpitations, moderate, possibly-related).**Conclusions:**Efficacy and safety results in velaglucerase-transitioned patients randomized to either eliglustat or imiglucerase were consistent with the full ENCORE trial population. In both groups, clinical stability was maintained and the treatments were well-tolerated.

480W

Gene Expression Profiling of Whole Blood from TNF Receptor-Associated Periodic Syndrome (TRAPS) Patients to Understand Response to Canakinumab Treatment. *R.I. Torene[1], M. Gattorno[2], H. Lachmann[3], L. Obici[4], A. Meini[5], V. Tormey[6], R. Caorsi[2], L. Baeriswyl[7], U. Affentranger[7], S. Starck-Schwertz[7], M. Letzkus[7], N. Hartmann[7], K. Abrams[8], N. Nirmala[1].* 1) Biomarker Development, Novartis Institutes for Biomedical Research, Cambridge, MA; 2) G. Gaslini Institute, Genova, Italy; 3) Royal Free Hospital, London, United Kingdom; 4) Amyloid Centre, IRCCS Policlinico San Matteo, Pavia; 5) Pediatric Clinic, Spedali Civili, Brescia, Italy; 6) Galway University Hospitals, Galway, Ireland; 7) NIBR, Basel, Switzerland; 8) Novartis Pharmaceuticals Corporation, New Jersey, USA.

Purpose: TRAPS is an autosomal dominant condition resulting from variants in the TNF super family receptor 1A (*TNFRSF1A*) gene. Thus far, >100 different allelic variants, including SNPs, indels, and complex variants, in this gene have been identified in TRAPS patients. It is an autoinflammatory disease causing unprovoked fevers, myalgia, abdominal pain, rash, headaches, and, in severe cases, AA amyloidosis. Canakinumab is a high-affinity, human, selective, anti-IL-1b monoclonal antibody, developed for the treatment of autoinflammatory diseases. The purpose of this study was to determine whether gene expression in whole blood can support a molecular mechanism for the activity of canakinumab in TRAPS patients.**Methods:** Twenty patients with active TRAPS received open-label canakinumab 150 mg sc/month for 4 months in an efficacy and safety study. Whole blood was collected at baseline, Day 15 and Day 113 from 19 of these patients and 1 sample each from 19 untreated age-matched healthy volunteers for analysis of gene expression levels by microarrays.**Results:** All 20 patients, regardless of *TNFRSF1A* allelic variant, improved upon treatment. Gene expression profiles of these patients shifted towards that of healthy volunteers. The disease-causing gene (*TNFRSF1A*), drug target gene (*IL-1b*), and other inflammation related genes (e.g., *MAPK14 and NFKB*) were down-regulated after treatment and several inflammation related pathways were evident among the differentially expressed genes. Many of the differentially expressed genes had expression levels that correlated with neutrophil count, however, neutrophil count alone could not account for the expression differences observed.**Conclusions**: Altogether, the gene expression data support a model in which canakinumab increases neutrophil apoptosis and reduces pro-inflammatory signals through its inhibition of IL-1b. Canakinumab is able to reverse the overexpression of several genes associated with inflammatory response, including *IL-1b*. Interestingly, IL-1b blockade normalized the overexpression of the disease-causing gene, *TNFRSF1A*, at the RNA level, suggesting a direct impact on the main pathogenic mechanism of TRAPS.

481T

Impact of NGS in rare diseases: Identification of novel biological pathways and therapeutic development in nemaline myopathy. *V.A. Gupta, A. Cankaya, A. Kondo, A. Krapf, A.H. Beggs.* Division of Genetics and Genomics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA.

Next generation sequencing has opened new avenues in understanding the genetic causes of hereditary disorders. Unbiased sequencing of the human genome/exome has resulted in identification of novel genes and pathways that were previously inconceivable using traditional genetic approaches. This is particularly evident in rare genetic disorders that suffer from the limitations of small sample size and clinical heterogeneity making gene discovery very challenging. Nemaline myopathy (NM) is a rare genetic disease that is characterized by generalized skeletal muscle weakness. NMs form a highly heterogeneous group of myopathies ranging from fetal akinesia, through severe congenital forms with death from respiratory failure during the first year of life, to a mild-childhood onset myopathy with survival into adulthood. Approximately, 75% of known genes in genetically diagnosed patients encode structural proteins constituting the sarcomeric thin filaments of skeletal muscle, limiting the prospect of developing specific therapies. Using whole exome/genome sequencing in 95 NM patients, we had identified mutations in a novel kelch protein encoding gene, *KLHL41*, in five unrelated families. Unlike previously known thin filament-associated NM genes, Kelch proteins function as substrate-specific adaptors for cullin E3 ubiquitin ligase, a core component of the ubiquitin-proteasome system to regulate protein turnover. Immunofluorescence studies in patients' skeletal muscles with *NEB* and *ACTA1* mutations showed highly reduced levels of proteasomal complexes suggesting a crucial role of KLHL41 in the proteasomal pathway in nemaline myopathy. As mutations in *NEB* affect 50% of genetically diagnosed cases in NM, we evaluated the effect of proteasome regulation in restoring muscle function in nebulin deficiency in a zebrafish model. Treating *neb* -/- zebrafish with small chemicals targeting the proteasomal pathway resulted in improved skeletal muscle function. These studies provide promising novel therapeutic targets for developing treatments for nemaline and related muscle diseases for which no treatment is available yet. This signifies the impact of next generation genomic approaches in translating gene discoveries into therapeutic developments that will have tremendous impact on improving the quality of life in patients affected with rare diseases.

482W

Systemic induced loss of NF1 in adult mouse is lethal. *A.N. Turner, K. Li, T.R. Schoeb, B.R. Korf, R.A. Kesterson.* Department of Genetics, University of Alabama at Birmingham, Birmingham, AL.

Mutations in the Neurofibromatosis type 1 (NF1) can manifest in numerous tissues as a result of loss of neurofibromin protein function(s), one of which is negative regulation of p21ras (Ras) signaling via a known GTPase activating (Ras-GAP) domain. Other potential functions of the multi-domain NF1 protein are less clear with biallelic inactivation of the NF1 gene leading to an embryonic lethal phenotype. Our lab is currently creating and characterizing mice with recurrent nonsense mutations found in NF1 patients to recapitulate the premature termination codons (PTCs) found in humans, thereby providing a model for therapeutic intervention studies (e.g. nonsense suppression therapy (NST) that enhances the insertion of an amino acid at a PTC and allows translation to proceed to produce a full-length protein). Thus, we created a novel NF1 mouse line carrying a recurrent nonsense mutation found in NF1 patients at exon 18 (c. 2041 C>T; p. Arg681Ter, Nf1st18), which when combined with a conditional knockout allele (Nf1Flox) in the proper context will develop spinal cord tumors. While our goal is to develop therapeutic interventions to inhibit or treat developing tumors associated with NF1 loss, the development of tumors in NF1 mouse models is a time consuming and costly experimental process. Therefore, we established an “acute” conditional knockout model using a tamoxifen-inducible CAGGCre-ER recombination system to gain a better understanding of the role of NF1 in the adult mouse and to develop more rapid methods of assessing neurofibromin function in response to treatment. Following inactivation of floxed NF1 allele(s), adult NF1F/F; CAGGCre-ER mice lose function of NF1 systemically and are not able to survive beyond 12 days with animals showing severe damage to mitotically active cells throughout the body, including hematopoietic cells. Similarly, mice harboring the PTC Nf1st18 allele along with a single floxed NF1 allele (NF1St18/F; CAGGCre-ER) fail to survive more than 12 days following inactivation of the floxed allele. This model allows rapid testing of nonsense suppression drugs, and the identification of NF1 sensitive cells and tissues in the adult for a more robust readout of NF1 activity to complement our current ongoing tumor studies.

483T

Caveolae-mediated uptake of a-galactosidase A in Fabry disease *in vitro* systems. *E. Changsila, O. Goker-Alpan, T. Taber, C. Martin, C. Sejpal, M. Ivanova.* Lysosomal and Rare Disorders Research and Treatment Center, Fairfax, VA, USA.

Fabry disease (FD) is a lysosomal storage disorder due to the deficiency of the lysosomal enzyme a-galactosidase A, resulting in cardiac, renal and cerebrovascular end-organ injury. Two human recombinant a-galactosidase A (rh-aGAL-A) preparations are available as enzyme replacement therapy (ERT) for FD, however only a-galactosidase beta (Fabrazyme), has been approved in the USA. Despite of ERT, there is still a significant mortality associated with FD, suggested to be a result of lack of delivery of rh-aGAL-A to different cell types with similar efficiency. Although, mannose-6-phosphate receptor (M6PR) dependent endocytosis is believed to be the major pathway for intracellular processing of rh-aGAL-A, other mechanisms mediated by megalin, sortilin, and even passive diffusion have been implicated in different cells.We investigated the uptake and intracellular transport of rh-aGAL-A *in vitro* in cell types affected in Fabry disease including; fibroblasts derived from patients with FD and healthy controls, neuronal cells (HCN2), and embryonic kidney cells (HEK293), mainly focusing on mannose-6-phosphate receptor(M6PR)-dependent endocytosis. Enzyme uptake was measured by a-galactosidase assay, using fluorescent labeling and microscopy. Inhibitors of M6PR (mannose-6-phosphatase, or M6PR neutralized antibody), caveolae inhibitor (Filipin) and the microtubule depolymerization inhibitor (nocodazole) were used to study the endocytic process. We report that the efficiency of enzyme uptake is cell type specific. The time to maximum enzyme load was 3h for fibroblasts *vs* 6h for HCN2 cells. HEK293 cells demonstrated biphasic uptake with the first maximum occurring 1h after treatment followed by return to the basal levels and the second peak 24h after treatment. Uptake of rh-aGAL-A into lysosomes during the initial hours is achieved by caveolae-mediated endocytosis, as evidenced by the inhibition in presence of Filipin. Nocodazole blocked the uptake in all cell lines including fibroblasts from patients with FD. Unexpectedly, M6PR inhibitors did not block uptake of rh-aGAL-A, indicating that this mechanism may not operate initially or alternative M6P inhibitors will need to be investigated.

484W

Development of an Intrathecal Enzyme Replacement Therapy for Sanfilippo Syndrome Type D (Mucopolysaccharidosis IIID). *D.R. Moen[1,3], X. Zhang[3], S-H. Kan[3], J. Wood[1,4], S. Ekins[1], T-F. Chou[2,3], P.I. Dickson[2,3].* 1) Phoenix Nest, Inc., Brooklyn, NY; 2) Department of Pediatrics, Harbor-UCLA, Torrance, CA; 3) Los Angeles Biomedical Research Institute, Torrance, CA; 4) Jonah's Just Begun, Brooklyn, NY.

Sanfilippo syndrome (mucopolysaccharidosis type III; MPS III) is a devastating neurodegenerative disorder with no cure or effective treatment currently available. All subtypes of MPS III have similar clinical phenotypes with onset in infancy or early childhood: progressive and severe neurological deterioration, hearing loss, and visceral manifestations. Most patients die before reaching adulthood, but some may survive to the fourth decade with progressive dementia and retinitis pigmentosa. There are, however, many therapies in early development for MPS III, including genetic, stem cell, enzyme replacement (ERT), chaperone, and substrate reduction.MPS IIID, the rarest of the MPS III subtypes, is an autosomal recessive storage disorder caused by genetic deficiency of N-acetylglucosamine-6-sulfatase (GNS), a lysosomal enzyme vital in the pathway of glycosaminoglycan (GAG) degradation. Over time, these GAG accumulate and eventually induce cellular death, with nervous tissue most severely affected. ERT has long been a strategy to treat MPS disorders, however, due to the difficulties in delivery of therapeutic agents across the blood brain barrier, treatment of the neurodegenerative pathology of the disease has proven difficult. Therefore, our strategy is to manufacture and deliver the enzyme directly to the affected tissue via an intrathecal route. Early studies of intrathecal ERT for other forms of MPS have already shown promise, with administration early in life preventing or reversing the visible signs of neurodegeneration in animal models.Using a stably transfected Chinese hamster ovarian (CHO) cell line, we have begun producing pre-clinical levels of recombinant human GNS (rhGNS) protein. rhGNS has been purified and enzymatically characterized, and we have optimized storage conditions for both longevity and safe administration. Using MPS IIID fibroblasts, we are evaluating its cellular uptake and will further demonstrate localization in the lysosome and the ability to reduce GAG storage.

485T

Comparison of taliglucerase alfa 30 U/Kg and 60 U/Kg in treatment-naïve pediatric patients with Gaucher disease. *M.P. Wajnrajch[1], A. Zimran[2], D.E. Gonzalez-Rodriguez[3], A. Abrahamov[2], P.A. Cooper[4], S. Varughese[4], A. Paz[5], E. Brill-Almon[5], D. Lewis[1], R. Chertkoff[5].* 1) Pfizer Inc, New York, NY; 2) Gaucher Clinic, Shaare Zedek Medical Center, Jerusalem, Israel; 3) Inst Privado de Hematol e Investig Clin, Asuncion, Paraguay; 4) University of the Witwatersrand & Charlotte Maxeke Johannesburg Academic Hospital, Johannesburg, South Africa; 5) Protalix BioTherapeutics, Carmiel, Israel.

Background/Objectives: Taliglucerase alfa is an enzyme replacement therapy approved for treatment of patients with Type 1 Gaucher disease (GD) and is the first approved plant cell–expressed recombinant therapeutic protein.**Patients/Methods:** Pediatric patients were randomized to receive either 30 (n=6) or 60 (n=5) U/kg of taliglucerase alfa every other week. Due to small patient numbers, there were numerical imbalances in disease parameters between the dose groups at baseline but they were clinically comparable with regard to anemia, risk of bleeding, and organ volumes. Mean percentage changes from baseline were used to compare the response between the dose groups and as a measure of control for numerical imbalances in baseline disease parameters.**Results:** Through 12 months, taliglucerase alfa 30 and 60 U/kg, respectively, increased mean hemoglobin concentration (+13.8% and +15.8%) and mean platelet count (+30.9% and +73.7%), and reduced mean spleen volume (-34.1% and -48.5%), liver volume (-14.5% and -25.0%), chitotriosidase activity (-58.5% and -66.1%), and CCL18 concentration (-50.6% and -52.6%). **Discussion/Conclusion:** Although statistical analysis was not possible due to small numbers of patients, both treatment groups demonstrated clinically meaningful improvement from baseline in these disease parameters with numerically greater improvement observed in the 60-U/kg dose group.**Disclosure:** This study was sponsored by Protalix BioTherapeutics. Pfizer and Protalix entered into an agreement in November 2009 to develop and commercialize taliglucerase alfa.

486W

Long-term safety and efficacy of taliglucerase alfa in pediatric patients with Gaucher disease who were treatment-naïve or previously treated with imiglucerase. *A. Zimran[1], D.E. Gonzalez-Rodriguez[2], A. Abrahamov[1], P.A. Cooper[3], S. Varughese[3], P. Giraldo[4], M. Petakov[5], E.S. Tan[6], A. Paz[7], E. Brill-Almon[7], R. Chertkoff[7].* 1) Gaucher Clinic, Shaare Zedek Medical Center, Jerusalem, Israel; 2) Inst Privado de Hematol e Investig Clin, Asuncion, Paraguay; 3) University of the Witwatersrand & Charlotte Maxeke Johannesburg Academic Hospital, Johannesburg, South Africa; 4) CIBERER, Hospital Universitario Miguel Servet, Zaragoza, Spain; 5) Belgrade University Medical School, Belgrade, Serbia; 6) KK Women's and Children's Hospital, Singapore; 7) Protalix BioTherapeutics, Carmiel, Israel.

Background/Objectives: Taliglucerase alfa is an enzyme replacement therapy approved for treatment of patients with Type 1 Gaucher disease (GD) and is the first approved plant cell–expressed recombinant therapeutic protein.**Patients/Methods:** This extension study of taliglucerase alfa in pediatric patients included those who were either treatment-naïve (n=10) or who were previously switched from imiglucerase (n=5). Patients received taliglucerase alfa 30 U/kg or 60 U/kg (treatment-naïve patients) or at the same dose as previously treated with imiglucerase (switch patients).**Results:** In treatment-naïve patients, taliglucerase alfa 30 and 60 U/kg, respectively, increased mean hemoglobin concentration (+19.7% and +23.3%) and mean platelet count (+23.9% and +156.6%) while also reducing mean spleen volume (-67.8% and -68.9%), liver volume (-37.0% and -34.3%), and chitotriosidase activity (-72.7% and -84.4%) from baseline through 36 total months of treatment. In patients previously treated with imiglucerase, these disease parameters remained stable through 33 total months of treatment with taliglucerase alfa. In both studies, most adverse events were mild/moderate and treatment was well tolerated.**Discussion/Conclusion:** These long-term results of taliglucerase alfa in pediatric patients with GD extend the taliglucerase alfa clinical safety and efficacy data set.**Disclosure:** This study was sponsored by Protalix BioTherapeutics. Pfizer and Protalix entered into an agreement in November 2009 to develop and commercialize taliglucerase alfa.

487T

Eteplirsen, a Phosphorodiamidate Morpholino Oligomer (PMO) for Duchenne Muscular Dystrophy (DMD): Clinical Update. *E.M. Kaye[1], L.R. Rodino-Klapac[2], Z. Sahenk[2], L.P. Lowes[2], L.N. Alfano[2], K. Berry[2], A.M. Gomez-Ramirez[2], S. Lewis[2], K.M. Flanigan[2], L.H. Cripe[2], S. Al-Zaidy[2], P. Duda[1], P. Sazani[1], J. Saoud[1], J.R. Mendell[2].* 1) Sarepta Therapeutics, Cambridge, MA; 2) Nationwide Children's Hospital, Columbus, OH.

DMD is a rare, degenerative, X-linked recessive genetic disease affecting approximately 1:3500 male births. It is caused mainly by mutations in the *DMD* gene that result in the inability to produce functional dystrophin, an essential protein for muscle fiber function. Eteplirsen, a PMO designed to enable functional dystrophin production in boys amenable to exon 51 skipping is being evaluated in on-going clinical trials.Twelve boys aged 7-13 years with eligible genotypes were randomized 1:1:1 to weekly IV eteplirsen 30 mg/kg, 50 mg/kg, or placebo for 24-weeks. All patients then transitioned into an ongoing open-label extension trial with 30 or 50 mg/kg eteplirsen. Clinical efficacy endpoints included the 6 Minute Walk Test and Pulmonary Function Testing. Safety assessments included adverse event recording, ECG, ECHO, hematology, blood chemistry, and urinalysis.A statistically significant treatment benefit of 65.4 meters ($p \leq 0.017$) on the 6MWT over 168 weeks was observed for patients in the continuous 30 and 50 mg/kg eteplirsen cohorts who were able to perform the 6MWT (n=6) compared with the placebo/delayed-treatment cohort (n=4). PFTs including MIP (+11.1%), MEP (+13.5%), and MIP/MEP %-predicted (-2.4% & -6.3%) were stable from Week 1 through 168 in all 12 subjects.No deaths, discontinuations due to AEs, treatment-related SAEs, immune activation including infusion reactions, or clinically significant abnormal laboratory findings were reported. Treatment compliance was high with ~99% of all possible doses administered.Eteplirsen's long-term tolerability and statistically significant clinical benefit suggest it could be a viable therapy for DMD boys amenable to exon 51 skipping.

488W

CRISPR-Cas9 as a potential therapeutic tool for the Bardet-Biedl Syndrome (BBS) M390R mutation *in vitro* and *in vivo*. *M.R. Cring, J.E. Garrison, V.C. Sheffield.* University of Iowa, Iowa City, IA.

Bardet-Biedl Syndrome (BBS) is a rare, autosomal recessive disease characterized by abnormal ciliary function and biogenesis. There are several phenotypes associated with BBS, including obesity, intellectual disability, hypogonadism, male infertility, and severe retinopathy. The most common mutation contributing to BBS is M390R in the BBS1 gene, which accounts for approximately 25% of known cases. Individuals with this particular mutation suffer photoreceptor death and typically go blind within their teenage years. Gene therapy approaches to correct this phenotype in the retina are attractive due to the accessibility of the eye. Previous studies have shown that overexpression of wild type BBS1 in the retina after delivery by AAV vectors leads to overexpression toxicity and continued photoreceptor loss. Therefore, gene correction strategies that restore endogenous expression levels may be a more viable therapeutic option. The CRISPR-Cas9 endonuclease has received extensive attention in recent years due to its ease of engineering and relatively high efficacy in mammalian cells. This study tests the feasibility of using the CRISPR-Cas9 platform and a homologous donor template to correct the M390R mutation *in vitro* and *in vivo*. We show that the spCas9 can target BBS1 in IMCD-3 cells, and addition of a single stranded oligo in iPSCs can introduce the M390R mutation by homologous recombination. Recent work has shown that the *Staphylococcus aureus* Cas9 endonuclease is small enough to package into AAV vectors. Therefore, we have produced several gRNAs in plasmids containing the saCas9, including an M390R allele specific gRNA. We have also constructed a homologous donor template including synonymous changes such that endonuclease activity will cease after homologous recombination events have occurred. We are testing our saCas9 constructs in IMCD-3 cells to determine their level of endonuclease efficiency. We are delivering multiple independent CRISPR-Cas9 constructs along with gRNAs and homologous donor templates to mutant mouse brains via electroporation. Excision and sequencing of brains will determine whether homologous recombination and correction has occurred. Once this has been established, we will package the Cas9 constructs that facilitate homologous repair into an AAV2/5 vector and deliver it to the retinas of 2 week old mutant mice. These mice will be monitored and tested for photoreceptor survival by histology and ERG.

489T

Promoterless gene-targeting using adenoassociated viral (AAV)-mediated homologous recombination to treat methylmalonic acidemia. *R.J. Chandler[1], A. Barzel[2], L. Lisowski[3], M.A. Kay[2], C.P. Venditti[1].* 1) National Human Genome Research Institute, NIH, Bethesda, MD; 2) Departments of Pediatrics and Genetics, Stanford University, Stanford, CA; 3) Gene Transfer, Targeting and Therapeutics Core, Salk Institute, San Diego, CA.

Methylmalonic acidemia (MMA) is a lethal inborn error of metabolism most typically caused by mutations in methylmalonyl-CoA mutase (*MUT*). Affected patients suffer from frequent and potential lethal bouts of metabolic instability that can be treated by orthotopic liver transplantation. Adeno-associated viral (AAV) gene therapy has shown great promise as the treatment for MMA in a murine model of the disease. However, a majority of the AAV-treated mice developed hepatocellular cancer, which was determined to arise from AAV-mediated insertional mutagenesis. In an attempt to create a safer gene therapy platform for the treatment of MMA, we created a novel vector, AAV-2A-MUT, to target site-specific gene addition of human *MUT* into the mouse albumin (*Alb*) locus without the use of nuclease. A promoterless AAV backbone contained a 2A-peptide coding sequence proximal to a codon-optimized human *MUT* gene and was flanked by arms of homology immediately upstream of *the Alb* stop codon. Site-specific integration of the vector into the *Alb* locus results in the expression of *Alb* and *MUT* as a single transcript. Following transcription the 2A-peptide causes ribosomal skipping and allows Alb and MUT to be translated as separate proteins. We delivered 2.5e12 GC of AAV8-2A-MUT by intraperitoneal injection to mice with MMA at birth. At one month post-injection, we observed increased hepatic expression of the MUT by western blot, improved growth and a significant reduction of plasma methylcitrate, a disease related metabolite, in the treated MMA mice. This gene delivery approach is anticipated to provide permanent hepatic transgene expression while reducing the risk of off-target integration and vector mediated insertional mutagenesis.

490W

Choroid plexus-targeted viral gene therapy in an alpha-mannosidosis mouse model increases brain LAMAN activity globally and eliminates brain pathology. *E. Choi[1], J. Wolfe[2], S. Kaler[1].* 1) Section on Translational Neuroscience, Molecular Medicine Program, NICHD/NIH, Bethesda, MD; 2) Children's Hospital of Pennsylvania, Philadelphia, PA.

The choroid plexuses are vascularized structures that project into the cerebrospinal fluid (CSF). The specialized polarized epithelia of choroid plexuses produce CSF by transporting water and ions into the ventricles, and are post-mitotic, i.e., do not undergo turnover. We hypothesized that remodeling these epithelia to secrete missing lysosomal enzymes by one-time CSF administration of a recombinant AAV (rAAV) could be an attractive strategy for long-term treatment of lysosomal storage diseases (LSDs). Lysosomes function as digestive units of cells and specific enzymes within lysosomes break down nutrients. Patients with LSDs cannot metabolize certain nutrients, resulting in diminished lifespans and reduced quality of life. CSF-directed recombinant enzyme replacement has shown promise for several LSDs but requires repeated instillations due to short recombinant enzyme half-lives. In contrast, rAAV-mediated gene transfer to the choroid plexus could enable continuous synthesis of lysosomal enzymes and steady delivery to the brain globally. To evaluate this hypothesis, we obtained an alpha-mannosidosis mouse model generated by targeted disruption of the lysosomal acid-mannosidase (LAMAN). We cloned the human (hu) LAMAN cDNA into a rAAV shuttle plasmid and generated high titer rAAV5 expressing huLAMAN. We administered viral particles to the CSF of homozygous mutant mice by brain lateral ventricle injection on day 3 of life. We documented dose-dependent transduction and huLAMAN mRNA expression confined to the choroid plexuses of rAAV5-treated animals. Brain biochemical analyses at 1, 2 and 6 months post-treatment documented sustained increases of LAMAN enzyme activity globally across the brain. By 8 months of age, untreated mutant mice showed prominent lysosomal vacuoles in hippocampal neurons, in contrast to rAAV5-LAMAN treated mutants for which brain histopathology was comparable to wild-type. If choroid plexus-targeted viral gene therapy approach were similarly successful in larger animals and human subjects, the most significant current barriers to health for patients with LSDs could be circumvented.

491T

Development of a Stem cell/gene therapy approach to treat Sanfilippo syndrome type B. *D. Clarke, S. Kan, S. Le, Q. Bui, V. Sanghez, P. Dickson, M. Iacovino.* Pediatrics department, Los Angeles Biomedical Institute at Harbor UCLA, Torrance, CA.

Mucopolysaccharidosis type IIIB (MPS IIIB, Sanfilippo syndrome type B) is a lysosomal storage disease characterized by profound intellectual disability, dementia, and a lifespan of about twenty years. The cause is mutation in the gene encoding α–N-acetylglucosaminidase (NAGLU), deficiency of NAGLU, and accumulation of heparan sulfate. Obstacles to treatement include the absence of mannose 6-phosphate on recombinant human NAGLU, the blood brain barrier, and degradation of the enzyme. We will address these issues using a stem cell/gene therapy approach. A fusion protein of recombinant NAGLU and a fragment of insulin-like growth factor II (IGFII) was prepared for endocytosis by the mannose 6-phosphate/IGFII receptor. The fusion protein sequence was cloned into a lentiviral vector and used to transduce NAGLU -/- induced pluripotent stem cells (iPSCs) derived from NAGLU -/- mouse embryonic fibroblasts (MEFs). The iPSCs were differentiated to Neural Stem Cells (iNSCs). NAGLU activity assays were performed to test in vitro whether the NAGLU and NAGLU-IGFII were secreted and uptaken from deficient cells in a comparable manner. Here we show that the while the secretion of NAGLU was higher compared to NAGLU-IGF2, the uptake showed opposite trend with NAGLU-IGFII more efficiently uptaken in -/- iNSCs. Our next step will be transplanting corrected iNSC expressing NAGLU or NAGLU-IGFII into the neonatal brain of NAGLU -/- mice. We will assess the distribution of transplanted cells via the expression of NAGLU and lysosomal storage markers to determine whether corrected neuronal progenitor cells will be a long lasting and effective therapy forSanfilippo syndrome type B.

492W

Rescue of Diamond-Blackfan Anemia haploinsufficiency by knock-up of the deficient protein. *I. Dianzani[1], A. Aspesi[1], E. Pavesi[1], S. Parrella[1], S. Macri[1], D. Cotella[1], U. Ramenghi[3], S. Gustincich[2], A. Follenzi[1], C. Santoro[1].* 1) Health Sciences, Università del Piemonte Orientale, Novara, Italy; 2) International School for Advanced Studies (SISSA/ISAS), Trieste, Italy; 3) Department of Public Health and Pediatric Sciences, Turin, Italy.

Diamond Blackfan Anemia (DBA) is a rare erythroid aplasia often associated with malformations. In about 70% of cases heterozygous LOF mutations are found in ribosomal protein (RP) genes and point to haploinsufficiency. DBA is considered the paradigm for ribosomopathies, a class of diseases caused by ribosome dysfunction and often characterized by bone marrow failure. Mutations in RPs either of the small or the large subunit impair rRNA maturation and ribosome biogenesis. As a consequence of the activation of ribosomal stress abnormal amounts of rRNA precursors accumulate and the P53 pathway is activated in affected cells.The aim of our project is to correct the effects of the haploinsufficiency observed in DBA by increasing the level of the deficient RP by favouring translation of the WT allele transcript. As a proof of concept precursor rRNA and the P53 pathway will be first evaluated to assess the rescue. Some of us discovered a new class of antisense non-coding RNAs (SINEUPs) that promote translation of partially overlapping sense coding mRNAs with no effects on mRNA levels. This “knock-up” activity depends on two functional domains: a 5’ overlap with the coding mRNA that confers target specificity and an embedded inverted SINEB2 element that functions as activator of translation. Synthetic transcripts with these features increase the translation of target mRNAs. We have decided to employ this technology to knock-up the synthesis of deficient RPs in cells from DBA patients, that carry a null mutation. We designed SINEUP molecules specific to RPL5 and RPS19, the two most commonly mutated DBA genes, and prepared lentiviral vectors to efficiently express them in primary cells.We show that the SINEUP technology is able to increase the level of the target RP in HeLa and erythroid cell lines without affecting the mRNA levels. Preliminary experiments on DBA patient activated lymphocytes suggest a rescue of the rRNA phenotype.

493T

CSF-directed AAV9 gene therapy plus subcutaneous copper provides superior rescue in a mouse model of Menkes disease. *M.R. Haddad[1], D. Martinelli[1], E.Y. Choi[1], P.M. Zerfas[2], P.H. Sullivan[3], D.S. Goldstein[3], L.R. Brinster[2], D. Abebe[4], J.A. Centeno[5], S.G. Kaler[1].* 1) Section on Translational Neuroscience, Molecular Medicine Branch, National Institute of Child Health and Human Development, NIH, Bethesda, MD; 2) Office of Research Services, Division of Veterinary Resources, National Institutes of Health, Bethesda, MD; 3) Clinical Neurocardiology Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD; 4) Research animal management branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD; 5) Joint Pathology Center, Andrews Air Force Base, MD.

Menkes disease is a lethal infantile neurodegenerative disorder caused by mutations in a copper transporter, ATP7A. Untreated patients typically die by 3 years of age. Currently, the sole available treatment involves daily subcutaneous (sc) copper injections, which however are ineffective in ≈75% of affected patients even when commenced before the symptoms occur. To develop a more complete treatment for this illness, potentially suitable even for subjects with complete loss-of-function ATP7A mutations, we tested adeno-associated virus (AAV) vectors carrying human ATP7A cDNA in the mottled-brindled (mo-br) mouse. The mo-br mouse recapitulates the symptoms and early demise of human Menkes disease. We previously demonstrated that a viral gene therapy approach using AAV serotype 5 (AAV5) in combination with copper chloride administered to the cerebrospinal fluid (CSF) was able to significantly prolong survival in mutant mice. However, growth and neuromotor defects persisted in long-term surviving mice. In the current study, we report superior outcomes (n=40 mice) using CSF-directed AAV9 in combination with sc clinical grade copper histidine. We evaluated both AAV9 and AAVrh10 vectors given their broad tropism (choroid plexus epithelia, neurons and glial cells) after CSF administration compared to AAV5 (selective choroid plexus tropism). Individual treatments with AAV alone or copper alone did not significantly enhance survival. We also compared different AAV9 and rh10 doses: 1.6x1010; 5.0x109 and 1.6x109 vg, in combination with 15 µg of sc Cu, and found that intermediate and high AAV9-ATP7A doses were most effective. In correlation with the greatest survival benefit, CSF-directed AAV9 plus sc Cu normalized growth and neurobehavioral outcomes. This synergistic treatment effect markedly improved biomarkers of brain copper metabolism in comparison to untreated mutant mice, and correlated with viral genome copy number. Brain copper measurements in combination-treated mutants were not statistically different from wild type indicating restored copper transport to the brain. Electron micrographs and H&E stain of brain regions reflected significant improvements in brain pathology in the combination-treated animals. Compared to our previous study with AAV5, the present findings strongly suggest that CSF-directed gene therapy with AAV9, in combination with sc copper, may have clinical utility for this difficult illness.

494W

CRISPR-Cas9 Mediated Genome Editing of Myocilin in Hereditary Glaucoma. *A. Jain[1], G. Zode[2], K. Bugge[1], C. Searby[1], F. Zhang[3], A. Clark[2], V. Sheffield[1].* 1) Pediatrics, University of Iowa, Iowa City, IA; 2) North Texas Eye Research Institute, UNT Health Science Center at Fort Worth, Texas; 3) McGovern Institute for Brain Research at Massachusetts Institute of Technology, Cambridge, Maryland.

Glaucoma is a leading cause of irreversible blindness worldwide. Elevated intraocular pressure (IOP) remains the most important risk factor for glaucoma. Mutations in myocilin (*MYOC*) have been reported in >16% of Juvenile Open Angle Glaucoma and about 4% of Primary Open Angle Glaucoma patients. Mutations in *MYOC* lead to myocilin misfolding and are associated with endoplasmic reticulum (ER) stress in the trabecular meshwork (TM), the tissue that maintains aqueous humor outflow and regulates IOP. ER stress and/or death of the TM cells lead to ocular hypertension and glaucoma. Genetic depletion of *MYOC* expression does not lead to a discernable phenotype suggesting that *MYOC* glaucoma is a gain-of-function mutation. We propose to relieve ER stress in human and mouse TM cells by targeting the human *MYOC* gene using state of art CRISPR-Cas9 technology.We generated and characterized human TM cell lines overexpressing wild type and mutant human MYOC and looked for ER stress markers using quantitative PCR, Western immunoblotting and immunocytochemistry. *MYOC* was targeted by either transient transfection of px330-CRISPR plasmids or adenovirus Ad5- mediated transduction of guide RNAs targeting exon 1 of the human *MYOC* gene. Human primary TM cells were treated with Ad5-cr*MYOC* virus and evaluated for dexamethasone (DEX)-mediated MYOC accumulation and ER stress. Intravitreal injection of Ad5 viruses in transgenic (Tg)-h*MYOCY437H* ocular hypertensive mice was performed to study the effect on IOP.The results revealed that TM cell lines overexpressing mutant MYOC have increased accumulation of myocilin in the ER and increased ER stress markers compared to wild type. Both transient transfection and Ad5-mediated transduction reduced levels of MYOC and ER stress in these cells. Primary TM cells have reduced levels of DEX-mediated ER stress when pretreated with Ad5-cr*MYOC* as compared to Ad5-cas9 virus controls. Ad5-cr*MYOC* virus significantly reduced IOP as compared to Ad5-cas9 control virus in the Tg-h*MYOCY437H* mouse model. This is a proof of principle study indicating that targeting the *MYOC* gene can relieve ER stress in TM cells, rescuing their function and health, and preventing glaucoma.

495T

Spell-Checking Nature: Versatility of CRISPR/Cas9 for the Treatment of Inherited Disorders. *D.U. Kemaladewi[1], D. Wojtal[1,2], Z. Malam[1], S. Abdullah[1], T.W.Y. Wong[1,2], E. Hyatt[1], S. Pereira[1,3], J. Stavropoulos[4], V. Mouly[5], K. Mamchaoui[5], F. Muntoni[6], T. Voit[5], H. Gonorazky[1,4], J.J. Dowling[1,2,4,7], R. Mendoza-Londono[4,7], E.A. Ivakine[1], R.D. Cohn[1,2,4,7,8].* 1) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Canada; 2) Department of Molecular Genetics, University of Toronto, Toronto, Canada; 3) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada; 4) Department of Paediatrics, University of Toronto, Toronto, Canada; 5) Sorbonne Universites, UPMC Univ Paris 06, INSERM UMRS974, Paris, France; 6) Dubowitz Neuromuscular Centre, Institute of Child Health and Great Ormond Street Hospital, London, United Kingdom; 7) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Canada; 8) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, Canada.

Recent estimates suggest that about 400 million people worldwide are affected by orphan diseases, mostly caused by primary genetic abnormalities. While orphan-drug development has made some progress over the last few years, most genetic disorders lack therapeutic options. The recent discovery and advancement of genome engineering tools, such as the CRISPR/Cas9 system, have opened up avenues for the development of novel treatment strategies. Here we explore the versatility of CRISPR/Cas9 to alter expression of disease-modifying genes as well as target disease-causing mutations in a variety of genetic disorders. First, we used the CRISPR/Cas9 system to upregulate utrophin, a compensatory protein in Duchenne muscular dystrophy (DMD). We coupled a catalytically inactive Cas9 fused to transcriptional transactivator VP16 (dCas9-VP160) with guide RNAs targeting either *UTRN A* or *UTRN B* promoters. We observed 4.1–14.6 fold upregulation of utrophin in DMD myoblasts, leading to restoredβ-dystroglycan expression.We next interrogated whether the system can be used for allele-specific targeting, such as in achondroplasia patients carrying pathogenic dominant-negative, gain-of-function mutation in fibroblast growth factor receptor 3 (*FGFR3*: c.1138G>A, p.Gly380Arg). We used *S. pyogenes* Cas9 (*Sp*Cas9) paired with a protospacer adjacent motif (PAM)-discriminating guide RNA, and detected 12 times higher indels frequency in the mutant allele compared to the wild-type by means of deep sequencing technology.Subsequently, we used the CRISPR/Cas9 system to remove large genomic rearrangements. Using *Sp*Cas9 coupled with a single RNA guide targeting the duplicated region, we removed a large 278kb tandem duplication on the X-chromosome including the *MECP2* gene. Moreover, using similar strategy, we successfully removed a 139kb region in DMD fibroblasts carrying exons 18-30 duplication in *DMD*, leading to restoration of full-length dystrophin and α-dystroglycan expression. Importantly, we found predicted off-target cutting to be minimal and not affecting any critical regions within the human genome.Taken together, our data indicate that the CRISPR/Cas9 system is a versatile approach to correct pathogenic mutations and alter expression of disease-modifying genes. Proof of concept studies utilizing patient cells as outlined here are critical in laying the foundation for further research into the application of these therapeutic strategies for numerous inherited disorders.

496W

Broad Therapeutic Window for Retinitis Pigmentosa. *S.F. Koch[1], Y.T. Tsai[1], J.K. Duong[2], L. Bonet-Ponce[3], C.W. Hsu[1], W.P. Wu[1], W.W. Wu[1], C.S. Lin[4], S.H. Tsang[1].* 1) Ophthalmology, Columbia University, New York, NY; 2) Department of Biostatistics, Mailman School of Public Health, Columbia University Medical Center, New York, NY; 3) Department of Physiology, School of Medicine and Dentistry, Catholic University of Valencia, Valencia, Spain; 4) Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, NY.

Retinal gene therapy has yet to achieve sustained functional rescue after disease onset – perhaps because transduction efficiency is insufficient ("too little") and/or the disease is too advanced ("too late"). To test the latter hypothesis, we developed a novel mouse model for retinitis pigmentosa (RP) that allowed us to restore the mutant gene in *all* diseased photoreceptor cells, thereby ensuring sufficient transduction efficiency. We then treated mice at early, mid or late disease stages. At all three time points, degeneration was halted and function rescued for at least 6 months. In addition to being the first demonstration of gene therapy efficacy after onset of degeneration, our study is the first to demonstrate a broad therapeutic time window. The results suggest that RP patients are treatable, despite most being diagnosed after significant cell loss. To maximize clinical impact, our work suggests that gene therapy research must focus on improving transduction efficiency.

497T

Pre-clinical development of a genetically-modified human dermal fibroblast (FCX-007) for the treatment of recessive dystrophic epidermolysis bullosa (RDEB). *M.P. Marinkovich, MD[1], N. Ehsani-Chimeh, MD[1], N. Nguyen[1], S. Moncrief, PhD[2], V.K. Dailey[2], M. Chakiath[2], A. Elayadi, PhD[2], S. Krishnan, MS[2], J. Maslowski, MS[3].* 1) Stanford University School of Medicine, Stanford, CA; 2) Intrexon Corporation, Germantown, MD; 3) Fibrocell Science, Exton, PA.

Recessive dystrophic epidermolysis bullosa (RDEB) is an autosomal recessive, inherited skin disease caused by null mutations within the type VII collagen gene (*COL7A1*). The mutations cause an absence or reduction of functional collagen VII (COL7), which make up anchoring fibrils that maintain binding of the epidermis to the dermis. The disease is characterized by a mechanical fragility and repeated blister formation in the sub-lamina densa, at the level of the structurally defective anchoring fibrils. Currently, there is no effective therapy for this disease, and death is usually the result of aggressive squamous cell carcinoma, sepsis or malnutrition.We are developing an autologous, genetically-modified fibroblast cell therapy that is anticipated to improve skin function in RDEB patients through restoration of collagen levels. A patient's fibroblasts will be harvested, genetically modified *ex-vivo* with a functional *COL7A1* gene, and expanded in culture (FCX-007). *Ex vivo* transduction will occur through the use of a replication-defective, self-inactivating (SIN) lentiviral vector. After expansion, the fibroblasts are administered back to the patient as a local intradermal injection into target wound margins. The resulting increase in anchoring fibrils is anticipated to stabilize the connection between skin layers and reduce blistering tendency.*In vitro* product development data indicates that cGMP scale FCX-007 cells express full-length COL7 collagen exhibiting the proper trimeric form, size and binding functionality. The integrated transgene copy number and COL7 expression is dependent on viral dose. We present results from a pre-clinical animal model evaluating FCX-007 in RDEB and normal human skin xenografts implanted onto immunodeficient SCID mice. The goals of the study are to confirm persistence, distribution and localization of COL7, and to evaluate any potential for product toxicity or vector biodistribution. The data presented are intended to support an Investigational New Drug (IND) filing.We also present the study design for a proposed Phase I/II clinical trial to treat RDEB subjects with FCX-007. The primary endpoints for the study are to evaluate safety and to confirm the formation of anchoring fibrils at the basement membrane zone (BMZ) in biopsies collected from subjects after product administration.

498W

Gene therapy for a mouse model of glucose transporter-1 deficiency syndrome. *S. Nakamura[1], H. Osaka[1], S. Muramatsu[2,6], N. Takino[2], S. Aoki[1], E. F. Jimbo[1], K. Shimazaki[3], T. Onaka[4], S. Ohtsuki[5], T. Yamagata[1].* 1) Department of pediatrics, Jichi Medical University, Shimotsuke-shi, Tochigi, Japan; 2) Division of Neurology, Jichi Medical University, Shimotsuke-shi, Tochigi, Japan; 3) Department of Neurosurgery, Jichi Medical University, Shimotsuke-shi, Tochigi, Japan; 4) Division of Brain and Neurophysiology, Department of Physiology, Jichi Medical University, Tochigi, Japan; 5) Department of Pharmaceutical Microbiology, Faculty of Life Sciences, Kumamoto University, Kumamoto, Japan; 6) Center for Gene and Cell Therapy, The Institute of Medical Science, The University of Tokyo, Japan.

(Background) Glucose transporter 1 deficiency syndrome (GLUT1DS, OMIM #606777) is an autosomal dominant disorder caused by haplo-insufficiency of *SLC2A1*, the gene encoding GLUT1. The classical phenotypes of GLUT1DS are intractable seizures, intellectual disability, and cerebellar ataxia, starting in infancy. We investigated if gene delivery of GLUT1 using an adeno-associated virus (AAV) vector can correct the phenotype in a mouse model of GLUT1DS.(Materials and Methods) We generated an AAV9 vector in which *SLC2A1*-myc-DDK was expressed under the neuron-specific synapsin I promoter. Transduction was evaluated in GLUT1+/- mice after systemic administration of AAV9-SLC2A1; 1.8 x 1011 vg per mouse into the peritoneum at 7 days old. After 5 weeks, total RNA was extracted from the cerebrum, and vector-specific RT-PCR was performed. Also the brain was subjected to immunofluorescence staining using anti-myc antibody (MBL CO., Ltd, Nagoya, Japan). To evaluate motor performance, rotarod and footprint analyses were performed to the age of 20 weeks mice.(Results) In GLUT1+/- mice after injection of AAV9-SLC2A1, vector-derived RNA was detected in the brain. Exogenous GLUT1 was mainly stained neural cells in the cerebral cortex and hippocampus. Cerebellar Purkinje cells were also immune-reactive to the anti-myc antibody. In rotarod analyze, the time to falling off the rotating cylinder was longer in injected GLUT1+/- mice than untreated GLUT1+/- mice.(Discussion) We confirmed that intraperitoneally administered AAV9-SLC2A1 can cross the blood-brain barrier and express exogenous GLUT1. Although the level of exogenous GLUT1 expression was low (~1% of neurons in the total visual field), motor function was improved. In general, systemic administration has advantage in its invasiveness and disadvantage in transduction efficiency to the CNS. We are comparing intraperitoneal or intra-cerebroventricular route as a way of administration.

499T

Engineering Recurrent, Reciprocal Genomic Disorders using CRISPR/Cas9 in Human iPS Cells. *D. Tai[1,2,3], A. Ragavendran[1,3], A. Stortchevoi[1,3], C. Seabra[1,2,3,4], P. Manavalan[1,3], S. Erdin[1,3], R. Collins[1,3], I. Blumenthal[1], C. Zhen[5], C. Lee[5], J. Gusella[1,6,7], M. Talkowski[1,2,3,6].* 1) Molecular Neurogenetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Department of Neurology, Harvard Medical School, Boston, MA; 3) Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research and Department of Psychiatry, Massachusetts General Hospital, Boston, MA; 4) Institute of Biomedical Sciences Abel Salazar, University of Porto, Portugal; 5) The Jackson Laboratory for Genomic Medicine, Farmington, CT; 6) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT; 7) Department of Genetics, Harvard Medical School, Boston, MA.

Recurrent genomic disorders involve large copy number variations (CNVs) generated from non-allelic homologous recombination (NAHR) mediated by segmental duplications. These reciprocal microdeletions and microduplications are a major cause of human disease, often producing phenotypically distinct syndromes. However, investigation of their impact at the molecular levels have been hampered by the size of the recurrent lesion, which can encompass many genes, as well as the diverse genetic backgrounds of patients and lack of accessible tissues. The capacity to generate large reciprocal CNVs in an otherwise isogenic human induced pluripotent stem cells (iPSCs) could overcome these obstacles and provide an invaluable tool for modeling these recurrent genomic disorders. Here, we used CRISPR/Cas9 genome engineering methods to achieve this goal for the 16p11.2 microdeletion/microduplication syndrome, a common genetic cause in autism spectrum disorder, as a proof-of-principle for NAHR-mediated disorders. Using a dual-guide RNA strategy, we first ablate the 575 kb unique genic segment spanning the 16p11.2 microdeletion region. We then developed a single-guide RNA approach that uniquely targets the flanking segmental duplications. The single guide RNA strategy was able to accurately mimic the NAHR-mediated *in vivo* mechanism in the cells, generating microdeletion of the unique genic segment as well as one copy equivalent of the segmental duplications. Remarkably, the single guide method was also able to reproducibly generate microduplications, including duplication of one copy equivalent of the segmental duplication. We performed multiple replication experiments and confirmed genome and transcriptome copy number alterations using copy number analysis (qPCR), genome-wide chromosomal microarray, and RNAseq, all of which revealed that both methods were specific (no off-target CNVs were observed from array) and relatively efficient (CNV was observed in 14.8% of all clones screened). Moreover, gene-expression patterns from RNAseq recapitulated those observed in our previous human and mouse 16p11.2 microdeletion RNAseq studies. These data suggest that our genome engineering approach provides an efficient method to model recurrent, reciprocal genomic disorders in human iPSCs, which can then be differentiated to any cell type of relevance. With further optimization and development, these methods may also permit efficient correction of these defects.

500W

Chimeric U2 small nuclear RNA-DMPK trans-splicing molecule reverses pre-mRNA splicing defects in myotonic dystrophy type 1. *P.S. Lai[1,2], H.Y. Chen[3], B.X. Qian[3], S. Udayappan[3], P. Kathirvel[3,4].* 1) Dept Pediatrics, National Univ Singapore, Science Drive 2, Singapore 119228; 2) Khoo Teck Puat - National University Children's Medical Institute, National University Health System, Kent Ridge Road, Singapore 119228; 3) Dept of Clinical Research, Singapore General Hospital, Outram Road, Singapore 169608; 4) Institute of Molecular Biology, Biopolis Drive, Singapore 138673.

RNA-based therapeutic approaches have been previously explored in a number of disease models for *in situ* rescue of genes. The use of an exogenous artificial pre-mRNA splicing molecule is an attractive strategy for modifying cis-splicing processes using the endogenous spliceosome machinery and has been shown to restore loss-of-function mutations in disorders such as hemophilia, cystic fibrosis, dysferlinopathy, etc. For gain-of-function diseases, it is important that the wild type mRNA is not removed when an artificial trans-splicing molecule (ATM) is introduced to correct the mutant pre-mRNA. Myotonic dystrophy type 1 (DM1) is an autosomal dominant disorder caused by CTG repeat expansions in the 3' end of the untranslated region of the dystrophia myotonica protein kinase (*DMPK*) gene which is encoded by 15 exons spanning 13 kb of genomic DNA. In this study, we show the efficacy of correcting DM1 using an artificial trans-splicing molecule (ATM) designed to target intron 14 of the *DMPK* gene. We constructed a chimeric ATM molecule carrying pre-mRNA binding and splicing domains, coding region of *DMPK* exon 15, reporter green fluorescence protein (*GFP*) gene and specific U2 snRNA sequences. Trans-splicing efficiency of the chimeric ATM with the U2 snRNA sequences (U2-ATM250) and the ATM without U2 snRNA sequences (ATM250) was tested in myosarcoma cells (CCL-136). Compared to the ATM250, the U2-ATM250 showed increased trans-splicing efficiency of endogenous DMPK pre-mRNA from 11.85% to 30.33%. Expression of U2-ATM250 in two immortalized DM1 fibroblast clones which were differentiated by inducible *MyoD* expression showed *DMPK* pre-mRNA trans-splicing efficiencies of 36.0% and 22.61% respectively. In addition, a significant reduction ($p<0.0001$) in mutant *DMPK* mRNA aggregates (foci) was observed by *in situ* hybridization and confocal microscopy in the nuclei of the differentiated DM1 fibroblasts. These findings demonstrate that chimeric U2 snRNA–ATM can increase trans-splicing efficiency of *DMPK* transcripts and induce phenotype change in intact DM1 fibroblast cells suggesting that this strategy may have broader applications to other diseases which are amenable to 3' trans-splicing therapeutic strategy.

501T

Silencing of Important Molecules Having Roles in Pathogenesis of Idiopathic Pulmonary Fibrosis via RNA interference and Development of New Therapeutic Modalities. *O.F. Hatipoglu[1], E. Uctepe[1], K.O. Yaykasli[2], H. Oksuz[1], S. Cigdem[1], M. Gunduz[1].* 1) Department of Medical Genetics, Turgut Özal University Faculty of Medicine, Ankara, Turkey; 2) Department of Medical Biology, Kahramanmaras University Medical Faculty, Kahramanmaras, Turkey.

Idiopathic Pulmonary Fibrosis (IPF) is a disease that is characterized by the deposition of an excessive degree of myofibroblast cells and extracellular matrix components in the lower respiratory tract and lung interstitium. Median survival is 3 years after initial diagnosis. Prevalence rate varies from 14 to 43 per 100 000 people. Today, it is thought that recurrent epithelial damage and aberrant wound healing are the basis of IPF pathogenesis, resulting in the accumulation of fibroblasts in the lung. In addition, coagulation, apoptosis, angiogenesis pathway disorders, oxidative damage, and most recently epithelial–mesenchymal transition (EMT) are implicated in the pathogenesis of this disease. Major aim of our study is to show the fundamental role of osteopontin, Twist and Wnt5a genes in IPF pathogenesis which these genes have been implicated by several studies in EMT and to show whether the suppression of these genes could be effective for IPF treatment examined. First of all, we treated A549 cell line with different dose of TGFb to create EMT. We showed that after TGFb treatment, Ecadherin expression is markedly decreased, on the contrary, Vimentin expression increased showing EMT. Then, changes in the expression of genes responsible for EMT formation and fibrosis examined at the level of mRNA and protein performing siRNA knockdown for Osteopontin, Twist and Wnt5a gene transcripts in lung alveolar cell lines.**Acknowledgements:** This study was supported by TUBITAK (The Scientific and Technical Research Council of Turkey), (Project Number: SBAG/ 113S947).

502W

Potential benefit of *CSF1R* mosaicism in a family with Hereditary Diffuse Leukoencephalopathy with Spheroids. *F.S. Eichler[1,2], P. Caruso[3], A.C. Bjonnes[2,4,5], J. Pan[1], R. Saxena[2,4,5].* 1) Dept of Neurology, Massachusetts General Hospital, Boston, MA; 2) Center for Human Genetic Research, MGH, Boston, MA; 3) Dept of Radiology, MGH, Boston, MA; 4) Dept of Anesthesia, Clinical Care and Pain Medicine, MGH and HMS, Boston, MA; 5) Medical and Population Genetics, Broad Institute, Cambridge, MA.

Mutations in the colony stimulating factor 1 receptor (*CSF1R*) have recentlybeen discovered as causal for Hereditary Diffuse Leukoencephalopathy withAxonal Spheroids (HDLS). We identified a novel, heterozygous missensemutation in the tyrosine kinase domain of CSF1R (E664K) by exomesequencing in five members of a family with HDLS. Three affected siblingshad characteristic white matter abnormalities and presented with progressiveneurologic decline. In the fourth affected sibling, early progression halted afterallogeneic hematopoetic stem cell transplantation (HSCT) from an unaffectedsibling over 15 years ago. Blood spot DNA from this subject displayed 15%chimerism in *CSF1R* acquired after HSCT. Interestingly, both parents wereunaffected but the mother's blood and saliva were ~20% mosaic for the CSF1Rmutation. Our findings suggest that increased levels of wild-type CSF1R relative to the mutant version, whetherachieved by mosaicism or chimerism, may confer benefit in HDLS and suggestthat HSCT might be considered for this disorder.

503T

Gluten-free diet is a cornerstone of Ehlers-Danlos syndrome management. *K. Angione, J.D. Milunsky, J.M. Milunsky.* Center for Human Genetics, Inc., Cambridge, MA.

Ehlers-Danlos syndrome (EDS) is a heterogeneous group of heritable connective tissue disorders divided into six major types. The most well recognized include "classical" (type I and II), "hypermobility" (type III), and "vascular" (type IV). EDS is characterized by arthralgia, joint and skin hypermobility, and vascular manifestations. Gastrointestinal (GI) issues and autonomic dysregulation are also common among individuals with EDS.EDS patients typically present with a myriad of symptoms calling for implementation of various management strategies. These may include both holistic approaches and medical interventions. We proposed that implementation of a gluten-free (GF) diet as one of several management strategies will be effective in improving the symptoms of EDS patients, particularly those suffering from multiple gastrointestinal ailments. The GF diet, while primarily used for the treatment of individuals with Celiac disease, has also been utilized as a treatment option for patients with a variety of other disorders, including irritable bowel syndrome, rheumatoid arthritis, diabetes mellitus, and chronic fatigue syndrome.Consumption of a GF diet has the potential to alleviate a wide variety of gastrointestinal symptoms, including chronic nausea/vomiting, abdominal pain, bloating, and IBS symptoms (frequent diarrhea and constipation). GF diets also yield additional benefits beyond the GI tract. Patients have reported decreased joint pain, improvement in autonomic symptoms (including fatigue, "brain fog," dizziness, and paresthesias), and an overall improvement in well-being.Out of 100 patients with a clinical diagnosis of EDS seen at our Center, 51 (51%) implemented a GF diet as part of their management plan. 38 of these patients (74.5%) reported improvement of symptoms with a "strict" GF diet. One additional patient (2.0%) reported improvement with a limited gluten diet. 12 patients (23.5%) reported that they had not observed any significant improvement of symptoms after a trial period of at least 6 weeks. In cases where multiple members of the same family were evaluated, only the proband was considered for these analyses.Adherence to a gluten-free diet significantly improves self-reported symptoms among individuals with EDS, contributing to improvement of autonomic symptoms, lowered pain scores, decreased gastrointestinal symptoms, and an overall improvement of quality of life. A GF diet should be considered an essential component of EDS management.

504W

Transcriptome Comparison of Rat and Human Schwann Cells. *D. Sant[1], P. Monje[2], D. Van Booven[1], G. Wang[1].* 1) John P. Hussman Institute of Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) The Miami Project to Cure Paralysis and Department of Neurological Surgery, University of Miami Miller School of Medicine, Miami, FL.

Schwann cells (SCs) are the myelinating glial cells in the peripheral nervous system. Transplantation of SCs has been shown to improve functional recovery of spinal cord injury in rats, suggesting a similar cell therapy for human treatment. However, human experiments have failed to yield the same success as rat treatments. The failure to achieve the same success as in rat cells is hypothesized to be due to inherent differences between rat and human SCs rather than differences in experimental design. To investigate the differences between human and rat SCs that have led to key differences in the ability to remyelinate neurons, we performed whole transcriptome sequencing (RNAseq) on human and rat SCs grown under identical conditions. Statistical significance in the difference of levels of transcripts were determined using three separate programs: BaySeq, DESeq, and edgeR. We found that 7284 of the 16483 homologous genes (44%) were expressed differentially. Further pathway analysis using MetaCore revealed two important pathways to be deficient in human SCs, both of which are known to mediate the SC's response to axon contact and differentiation into myelinating cells: (1) in the adenylate cyclase/cAMP signaling pathway, 15 of the known 31 genes were found deficient in human SCs; (2) in the extracellular matrix (ECM)-dependent, adhesion-dependent control of cytoskeletal organization pathway, 37 of the known 111 genes were found deficient in human SCs. Knowledge of these key deficiencies can be used to predict the ability of SCs to myelinate in cell culture and after transplantion, and help in the design of human Schwann cell-specific potency assays for clinical trials of spinal cord injury.

505T

The NIGMS Human Genetic Cell Repository at the Coriell Institute: A research participation opportunity for individuals with inherited diseases and chromosomal abnormalities. *N. Turan.* Coriell Institute for Medical Research, Camden, NJ.

The NIGMS Human Genetic Cell Repository was established at the Coriell Institute for Medical Research, an independent non-profit research institution, in 1972 by the National Institute of General Medical Sciences (NIGMS). The NIGMS Repository provides a readily accessible, centralized resource for cell lines and genetic material from individuals with genetic disorders, including heritable diseases and chromosomal abnormalities, and from individuals from various geographical locations for studies on human diversity. Samples are also collected from unaffected, first-degree relatives of individuals with genetic disease, and from apparently healthy individuals for the study of human variation.The primary goal of the NIGMS Repository is to stimulate and facilitate biomedical research, teaching, and the diagnosis and prevention of human genetic diseases by establishing and maintaining a repository of high quality, uncontaminated, well characterized, and clinically well-documented cell lines and DNA and making these resources available to investigators throughout the U.S. and abroad.The NIGMS Repository contains more than 11,300 cell lines, including human induced pluripotent stem cell (iPSC) lines, and 5,600 DNA samples, representing over 800 Online Mendelian Inheritance in Man (www.omim.org/statistics/entry, accessed 9/24/14) phenotypes.The NIGMS Repository is an invaluable resource and samples have been used for genetic testing assay development, validation and proficiency, identification of novel genes, exploration of gene function and expression, and generation iPSC lines. To date, there have been over 6,000 publications citing NIGMS Repository samples. The NIGMS Repository is also an important resource for clinicians and genetic counselors as it offers an opportunity for patients and families to contribute to the research process in addition to, or in the absence of, the availability of clinical testing or disease-specific research. Sample donation requires a blood or tissue specimen and completion of an informed consent, a submission form, and a clinical data summary.

506W

Accelerate R&D in genetic diseases for diagnostic and therapy: initiatives from the International Rare Disease Research Consortium (IRDiRC). *S. Ayme[1], A. Mills[1], L. Lau[1], P. Lasko[2], International Rare Diseases Research Consortium.* 1) US14-Orphanet, INSERM, Paris, ile de France, France; 2) CIHR Institute of Genetics, McGill University, Montreal, QC, Canada.

Despite considerable advances in the scientific field, genetic disease patient needs are far from covered, both in terms of diagnostic measures and in terms of effective therapies. To ensure that data generated by research is optimally used for the benefit of patients, an international consortium was set up as an initiative of the European Commission and the United States' National Institutes of Health. The consortium conducted in-depth discussions with all stakeholders on the obstacles to overcome in efforts to accelerate R&D in rare diseases. This brainstorming period resulted in recommendations for funding agencies and researchers to optimize the use of shared data through database and knowledge accessibility and interoperability. Moreover, five actions were launched. The first consists in promoting the use of ontologies to describe phenomes, essential for clinical databases to become interoperable. The second is to create a data exchange platform, enabling recognition of clinically similar cases according to clinical features or genomic data in efforts to identify new clinical entities, jointly with the Global Alliance for Genomics and Health. The third is to reach an international agreement on acceptable alternatives to randomized clinical trials when they are not applicable, and increase acceptability of new methods by regulatory agencies. The fourth is to accelerate the development of criteria to measure treatment effects that are relevant to patients. The fifth is to coordinate efforts to develop the use of scientific and clinical data by using dispersed resources, including natural language, for selecting drugs as potential treatment options for rare diseases. Initiatives, both academic and commercial, have bloomed recently, targeted at identifying new therapeutic targets and to repurpose drugs. They leverage on developments in Computational Linguistics and Graph Theory, to build a representation of knowledge which is automatically analysed to discover hidden relations between any drug and any disease, representing possible Modes of Action for any given pharmacological compound. Their efficacy for selecting drugs as treatment options for genetic diseases is already documented. All these initiatives constitute IRDiRC's roadmap which will be presented. .

507T

The ketogenic diet rescues defects of hippocampal neurogenesis in a mouse model of Kabuki syndrome. *J.S. Benjamin[1,2], G.A. Carosso[1,2], G. Pilarowski[1,2], L. Zhang[1], H. Cho[1], K.D. Hansen[1,3], H.J. Vernon[1,4,5], H.T. Bjornsson[1,5].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins Sch Med, Baltimore, MD; 2) Predoctoral Training Program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland; 3) Department of Biostatistics, Bloomberg School of Public Health, Baltimore, Maryland; 4) Kennedy Krieger Institute, Baltimore, Maryland; 5) Department of Pediatrics at the Johns Hopkins University School of Medicine, Baltimore, Maryland.

Kabuki syndrome (KS) is a genetically determined intellectual disability syndrome associated with distinctive facial features, immune deficiency, and growth retardation. KS is caused by mutations in either *KMT2D* or *KDM6A*, components of the histone machinery which both participate in the process of chromatin opening. If an inability of chromatin opening plays a role in the pathogenesis of KS, agents that favor chromatin opening such as the histone deacetylase inhibitors (HDACi) may provide therapeutic benefit. To test this hypothesis we previously characterized a mouse model of KS (*Kmt2d+/bGeo*), with craniofacial abnormalities, decreased H3K4me3 in the granule cell layer (GCL) of the dentate gyrus, and defects of neurogenesis and hippocampal memory. We saw rescue of all observed defects after treatment with an HDACi (AR-42) for two weeks. However, as a cancer drug, it may be difficult to transition AR-42 to the clinic, necessitating other possible treatments. Recently beta-hydroxybutyrate (BHB), a naturally produced ketone body, was shown to be an HDACi. We confirmed HDACi activity of BHB *in vitro* by a dose dependent increase in activity in an H4Ac reporter assay. To explore potential benefits of BHB *in vivo*, *Kmt2d+/bGeo* and *Kmt2d+/+* littermates were placed on a BHB promoting ketogenic diet (KD), which led to an increase of BHB in both genotypes. After two weeks on a KD we observed rescue of the neurogenesis deficiency as measured by the relative amount of doublecortin and EdU positive cells ($P<0.01$, $P<0.001$) as well as a normalization of open chromatin marks (H3Ac, H3K4me3) in the GCL ($P<0.05$, $P<0.05$) and gene expression abnormalities in the hippocampus in *Kmt2d+/bGeo*. On a KD, *Kmt2d+/bGeo* mice also demonstrated functional recovery in the number of platform crossings in a Morris water maze ($P<0.05$). In addition, exogenous administration of BHB ameliorated the neurogenesis defect supporting a mechanistic role for BHB for the observed therapeutic effect. Interestingly, while both genotypes showed increased BHB in response to a KD, *Kmt2d+/bGeo* mice demonstrate a natural predisposition towards elevating BHB compared to acetoacetate as well as an elevation of lactate compared to pyruvate ($P<0.05$), but both systems are under the control of the NADH/NAD+ ratio, an indicator of the cellular redox state. These data suggest that BHB elevation either through a ketogenic diet or exogenous administration may have therapeutic potential for KS and related disorders.

508W

Genetic Modifiers in Women with Turner Syndrome and Bicuspid Aortic Valve: Chromosomal Microarray and Whole Exome Sequencing. *P.S. Kruszka[1], S.I. Berger[1], N.K. Banks[1], A.F. Martinez[1], Y.A. Addissie[1], L.J. Olivieri[2], C.A. Bondy[3], M. Muenke[1].* 1) Medical Genetics Branch, NHGRI/NIH, Bethesda, MD; 2) Children's National Medical Center, Washington, D.C; 3) National Institute of Child Health and Human Development, National Institutes of Health, Bethesda Maryland.

INTRODUCTION: Biscupid aortic valve (BAV) occurs at a much higher frequency in individuals with Turner syndrome (>30%) than the general population (~1%). Bicuspid aortic valve is thought to result from haploinsufficiency of a putative gene or genes on the short arm of the X chromosome. This study investigates genetic modifiers of the aortic valve phenotype in women with Turner syndrome. **METHODS:** Women with monosomy X by 50 metaphase karyotype were enrolled. Aortic valves were evaluated with cardiac MRI. SNP array and whole exome sequencing completed evaluation. **RESULTS:** Twenty-five women with Turner syndrome, average age 44 years (range 27 to 68) were enrolled; 7 women had BAV (28%), 14 women had tricuspid aortic valve (TAV) (56%), and 4 women were classified as partial fusion based on cardiac MRI. Six patients had 7 copy number variations (CNVs), each of which contained genes involved in cardiac development or known to be associated with congenital cardiac malformations, including a deletion involving *CTNNA3*, a gene associated with arrhythmogenic right ventricular dysplasia. Exome sequencing yielded a total of 1,146,906 variants. Variants were then filtered for premature stop variants, splicing variants, frameshifts, and missense variants predicted to be damaging by PolyPhen-2, leaving 27,807 variants. Using these variants, a case-control evaluation was done on participants with bicuspid aortic valve versus tricuspid aortic valve using a threshold of $P < 0.01$ (uncorrected). There were 10 variants in 7 genes that were predicted to be deleterious by PolyPhen-2 with a $P < 0.01$ (uncorrected). The most significant gene was *DYNC2H1* ($P = 7.2 \times 10\text{-}4$), which was the only gene known to be associated with CHD. A candidate gene approach was conducted with 542 genes involved in cardiac development in humans and animal models. A total of 332 variants in 139 genes were found, with an equal distribution in BAV (14.7 variants/individual), TAV (12.7 variants per individual), and PF (12.8 variants per individual) ($P > 0.8$). Of the genes know to be associated with BAV (*NOTCH1, GATA6, FBN1, TGFBR2, ACTA2, HOXA1, KCNJ2*, there were two deleterious variants in *NOTCH1* found in patients, one in BAV (p.R621H) and one in TAV (p.A2044T). **CONCLUSION:** This study is the first whole exome sequencing evaluation of women with Turner syndrome. Our results suggest that the bicuspid aortic valve phenotype is a complex trait that may involve multiple modifier genes.

509T

A comprehensive sequence analysis of 22 tRNA genes identified a critical role of mitochondrial defects in hypertension. *M. Wang[1], P.P. Jiang[1], L. Xue[2], M.X. Guan[1].* 1) Institute of Genetics, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, China; 2) Attardi Institute of Mitochondrial Biomedicine, Wenzhou Medical University, Wenzhou, Zhejiang, China.

Hypertension is one of the major global public health problems, affecting approximately 1 billion worldwide, including 265 million adults in the China in hypertension.

Mitochondria can regulate multiple aspects of vascular function and be critical for pathogenesis of hypertension. However, only limited insight into the pathogenesis of hypertension has come from the genetic defects of mitochondrial genomes. In this report, we performed the first comprehensive study to investigate the spectrum and incidence of mutations in 22 mitochondrial tRNA genes in a cohort of 2070 Chinese subjects with hypertension. This analysis identified 165 nucleotide changes occurred at variable positions among 22 tRNAs. The numbers of variants varied from 3 in the tRNASer(UCN) to 24 in the tRNAThr. Then all the variants were evaluated for pathogenicity using following criteria: evolutionarily conserved nucleotides among 41 vertebrates, the absence of 512 controls, and potential structural and functional alterations. This analysis showed that 117 variants were polymorphisms and 47 variants including 4 known hypertension-associated mutations were putative pathogenic mutations. These putative mutations may destabilize the structures of tRNAs such as altering Watson-Crick (WC) base-pairings and disrupt tRNA functions included improper maturation of tRNA molecules, reduced ability to be charged by the cognate amino acid, or inability to decode corresponding codons. Consequently, a failure in tRNA metabolism may cause mitochondrial dysfunction for the development of hypertension. Our data provide the evidence that mitochondrial tRNA mutations are the important cause of hypertension, particularly accounting for 3.9% cases of 2070 Han Chinese subjects with hypertension, which may provide new insights into the pathophysiology and valuable information for management and treatment of hypertension.

510F

***PNPLA3* rs738409 is associated with APRI, a clinical predictor of hepatic fibrosis that shows association with left ventricular mass: the Strong Heart Family Study.** *S. Cole[1], K. Haack[1], J. Kent, Jr.[1], H. Göring[2], L. Best[3], E. Lee[4], R. Loomba[5], J. Umans[6], J. Weinsaft[7], R. Devereux[7].* 1) Dept Gen, Texas Biomedical Research Inst, San Antonio, TX; 2) South Texas Diabetes and Obesity Institute, University of Texas Health Sciences Center at San Antonio, Brownsville, TX; 3) Missouri Breaks Industries Research, Inc. Eagle Butte, SD; 4) Dept of Biostatistics and Epidemiology, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 5) Division of Gastroenterology and Epidemiology, University of California at San Diego, La Jolla, CA; 6) MedStar Health Research Institute, Hyattsville, MD; 7) Division of Cardiology, Weill Cornell Medical College, New York, NY.

Increased fibrosis is known to contribute to both hepatic disease and CVD. We tested these relationships in American Indians (AI), investigating whether hepatic fibrosis (HF), as assessed by the aspartate aminotransferase (AST) to platelet ratio index (APRI), is associated with left ventricular mass (LVM), a predictor of adverse cardiac outcomes. We also tested whether APRI is associated with *PNPLA3* rs738409, a known genetic risk variant for non-alcoholic fatty liver disease (NAFLD [MIM613282]). **Methods:** We studied 3,645 AI multigenerational family members from the Strong Heart Family Study (SHFS), from Arizona (AZ), North/South Dakota (DA) and Oklahoma (OK). Measures include alcohol consumption (ETOH), echocardiographically-measured LVM, alanine aminotransferase (ALT), AST (UI/L) and patatin-like phospholipase domain-containing protein 3 (*PNPLA3*) rs738409 [PNPLA3 I148M] genotypes, collected using standard methods. APRI was calculated as [((AST/AST upper limit of normal)/platelet count(109/L)) x 100]. APRI was tested for association with LVM while stratifying by ETOH (no vs. high [>3 drinks/day men, >2 drinks/day women]). Measured genotype analysis in SOLAR was used to test center-specific association between rs728409 genotype and APRI after adjustment for age, sex, BMI and ETOH, while accounting for pedigree relationships, with significance at a Bonferroni-corrected p-value=0.004. **Results:** We observed an APRI >1.5 (indicating significant HF) in 3.0%, 1.4% and 1.8% of the AZ, DA and OK participants, respectively, suggesting undiagnosed fibrotic disease in this AI population. An APRI > 1 was significantly associated with higher LVM in participants reporting either no (p=0.008) or high ETOH (p=0.01), indicating that HF may be correlated with cardiac remodeling and impaired function. Allele frequencies for rs738409 (C reverse strand; G coding missense) were 0.21, 0.55, 0.47 in AZ, DA and OK. Rs738409 was significantly associated with ALT, AST and APRI measures across all centers (p = 4x10-4 to 3.7x10-8), but unlike other populations, the G coding allele is associated with lower levels of the liver function and HF measures, suggesting additional genetic effects at this locus marked by this allele in AI. Rs738409 was not associated with LVM likely due to fibrosis being only one of many determinants of LVM. In conclusion, for the first time, and in AI, we demonstrate the association of *PNPLA3* rs738409 with APRI, a measure that links HF to CVD risk.

511W

Functional variants in a clinical setting: an example using *APOC3* R19X and extreme triglyceride levels extracted from electronic health records. *D.C. Crawford[1], K.E. Diggins[2], N.A. Restrepo[2].* 1) Epidemiology and Biostatistics, Institute for Computational Biology, Case Western Reserve University, Cleveland, OH; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN.

In a personalized or precision medicine setting, patients with extreme triglyceride (TG) levels may be flagged for further evaluation for cardiovascular disease risk assessment (TG ≥200 mg/dl) or for the presence of hyperthyroidism, malnutrition, or malabsorption disease (TG ≤10 mg/dl). We hypothesize that the addition of functional genetic variants such as *APOC3* R19X, a variant (rs76353203) associated with low TG levels and cardioprotection, can further facilitate the triage process in assessing disease risk among flagged patients. To test this approach, we surveyed BioVU, the Vanderbilt University Medical Centers's biorepository linked to de-identified electronic health records (EHRs), for adult European American patients (>45 and >55 years of age for men and women, respectively) with the lowest percentile of TG levels. The initial search identified 262 patients with the lowest TG levels in the biorepository; among these, 184 patients with sufficient DNA and the lowest TG levels were chosen for Illumina ExomeChip genotyping. Almost half the genotyped patient sample was female (43.5%) and the average body mass index was 24.8 kg/m2. As expected based on the sampling criteria, this was an older patient population with 28% born in the 1940s and 14.7% born in the 1930s (range: 1910s to 1960s). The first mentioned TG levels for these patients ranged from 15 mg/dl to 161 mg/dl with an average first TG level of 39.3 mg/dl. More than half of the genotyped patients (60.3%) had more than one TG level in the EHR. A total of two patients were identified as heterozygotes of *APOC3* R19X for a minor allele frequency (MAF) of 0.5% in this patient population of low TG compared with a recent estimate of 0.08% in a general American population ascertained by the National Health and Nutrition Examination Surveys. Both heterozygous patients had only a single mention of TG in the EHR (31 and 35 mg/dl, respectively). In this patient population, the inclusion of *APOC3* R19X genotypes in the EHR did not assist in assessing hyperthyroidism, malnutrition, or malabsorption given no TG levels ≤10mg/dl were identified in the EHR. However, at least two patients were identified as carriers of a null variant strongly associated with cardioprotective lipid profiles and decreased risk for coronary artery disease highlighting the possible utility of inclusion of functional genetic variation in clinical risk assessment.

512T

Striatin is a novl risk gene for human dilated cardiomyopathy. *M. Dasouki[1], N. Muiya[1], M. Abouelhoda[1], E. Alsolme[2], M. Nader[1,2], N. Dzimiri[1].* 1) Genetics, King Faisal Specialist Hospital & Research Center , Riyadh, Saudi Arabia; 2) Department of Physiology,College of Medicine, Al Faisal University, Riyadh, Saudi Arabia.

Introduction: Striatin (*STRN*) is a WD-repeat calmodulin-binding protein which binds to caveolin-1 and PP2A. It is localized in the dendritic spines of the brain striatum, motor and olfactory systems as well as the intercalated disc regions of the cardiac myocytes along with the three desmosomal proteins, Plakophilin-2, Plakoglobin and Desmoplakin which are known to cause dilated cardiomyopathy (DCM). In the Boxer dogs, spontaneous mutations in *STRN,* especially an 8-bp deletion in the 3' untranslated region, have been associated with arrythmogenic right ventricular cardiomyopathy (ARVC) and DCM, manifested by ventricular systolic dysfunction, dilatation and tachyarrhythmias, syncope, and even sudden death. More recently, association studies of single nucleotide polymorphic variants in mice and humans revealed a role for striatin in QRS duration and cardiac ventricular conduction as well as modulating the interaction between salt intake and blood pressure via the nongenomic mineralocorticoid receptor pathway. However, cardiac related mutations and expression in human *STRN* are still unknown.**Hypothesis**: Genomic variants in *STRN* may have a pathogenic role in human DCM.**Methods:** The full coding region of *STRN* was Sanger sequenced in gDNA samples from 96 Saudi Arabian adult patients with idiopathic DCM, and Western blotting performed in cardiac left ventricular tissue (9 DCM versus 5 controls).**Results**: A total of 277 novel heterozygous variants were identified in the entire cohort with two hot spots (50% in exon 17 and 33% in exon 2). Within the coding sequence, 7 novel exonic missense, 1 stop-gain and 1 frameshift SNVs were found, in addition to 28 variants in the 3'UTR of the gene. The novel SNVs identified in patients with idiopathic DCM include 2 potentially damaging (*STRN*:NM_003162:exon3: c.363C>G: p.Y121X and exon18:c.2318dupA:p.D773fs) truncating variants. Western blotting results indicated that the expression of STRN is lower in DCM left ventricles compared to control myocardium.**Conclusion**: This data suggests that *STRN* is a novel risk gene for DCM. Functional characterization of the effect of these novel variants in a Zebrafish model is underway.

513F

Testing population-specific quantitative trait associations for clinical outcome relevance in a biorepository linked to electronic health records: *LPA* and myocardial infarction in African Americans. *L. Dumitrescu[1], K.E. Diggins[1], R. Goodloe[1], D.C. Crawford[2].* 1) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 2) Institute for Computational Biology, Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.

Previous candidate gene and genome-wide association studies have identified common genetic variants in *LPA* associated with the quantitative trait Lp(a), and emerging risk factor for cardiovascular disease. The associations are population-specific and many have not yet been tested for association with the clinical outcome of interest. To fill this gap in knowledge, we accessed BioVU, the Vanderbilt University Medical Center biorepository linked to de-identified health records (EHRs), including billing codes (ICD-9-CM) and clinical notes, to test population-specific Lp(a)-associated variants for an association with myocardial infarction (MI) among African Americans. We performed electronic phenotyping among African Americans in BioVU ≥40 years of age using five case definitions for MI (#1: 1 ICD-9 code 410.* on three consecutive days; #2: 1 ICD-9 code 410.* on two consecutive days; #3: ≥3 ICD-9 code 410.* ever; #4: ≥2 ICD-9 code 410.* ever; #5: ≥1 ICD-9 code 410.* ever) and a single control definition (exclude MI: ICD-9 code 410.* and exclude ischemic heart disease: codes 411.*, 412.*, 413.*, 414.*). After review of the EHRs using keyword searches of the clinical notes (for "myocardial infarction", "MI", "infarc", "STEMI", "NSTEMI"), a total of 311 MI cases were confirmed, resulting in a positive predictive values ranging from 87.6% (definition #5) to 99.4% (definitions #2 and #3). Among the 5,883 controls identified, 344 were frequency matched for age and sex and chosen for genotyping. A total of five known Lp(a)-associated genetic variants in African Americans were genotyped using Sequenom: rs1367211, rs41271028, rs6907156, rs10945682, and rs1652507. *LPA* rs3798220 (I4399M) previously associated with increased levels of Lp(a), MI, and coronary artery disease in European Americans was also genotyped. Tests of association using logistic regression assuming an additive genetic model revealed no significant associations ($p<0.05$) for any of the five *LPA* variants previously associated with Lp(a) levels in African Americans. Interestingly, I4399M rs3798220 was not associated with MI in African Americans (odds ratio = 0.51; 95% confidence interval: 0.16 – 1.65; $p=0.26$) despite strong, replicated associations with MI and coronary artery disease in European American genome-wide association studies. These data highlight the challenges in translating quantitative trait associations to clinical outcomes in diverse populations.

514W

***APOL1* risk allele is associated with early diagnosis of hypertension and a 2-3 mmHg increase in systolic blood pressure in young African American adults.** *G. Galarneau[1], G.N. Nadkarni[1], S.B. Ellis[1], R. Nadukuru[1], S.A. Scott[1], T.A. Manolio[2], R. Li[2], L.J. Rasmussen-Torvik[3], A.N. Kho[3], M.G. Hayes[3], J.A. Pacheco[3], R.L. Chisholm[3], D.M. Roden[4], J.C. Denny[4], E.E. Kenny[1], E.P. Bottinger[1], eMERGE network.* 1) Icahn School of Medicine at Mount Sinai, New York, New York, USA; 2) National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA; 3) Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA; 4) Vanderbilt University Medical Center, Nashville, Tennessee, USA.

Apolipoprotein L1 (*APOL1*) risk allele is associated with end-stage renal disease in African American (AAs). To determine whether *APOL1* was associated with blood pressure (BP) traits as predisposing factors for kidney disease, we performed a retrospective analysis of longitudinal patient health records in 3 biobanks in the Electronic Medical Records and Genomics (eMERGE) network.The discovery cohort consisted of 5,213 AAs from Mount Sinai Bio*Me* Biobank. Replication cohorts consisted of an independent set from Bio*Me* (n=1,655), Vanderbilt BioVu (n=2,889) and Northwestern NUgene (n=613) biobanks. Median follow-up was 5, 3, 6 and 8 years, respectively. We genotyped SNPs comprising the *APOL1* G1 and G2 renal risk alleles; rs73885319, rs71785313 and rs60910145 in Bio*Me*and imputed the genotypes in BioVu and NUgene with Illumina 1M data. The *APOL1* risk allele (2 copies of G1 or G2) frequencywas 14%-16%. We performed Cox proportional hazards and linear and logistic regression to test association of the *APOL1* risk allele with time to hypertension (HTN) onset, BP as a continuous outcome, and hypertensive end organ damage as dichotomous outcomes. We used METAL for meta-analyses of replication cohorts.In 457 individuals with the risk allele and diagnosed HTN, median documented onset of HTN was 2-5 years younger with age, sex and BMI as covariates in discovery (Pdis=0.04) and replication cohorts (Prep=3x10-4).To test if there were age-specific associations between the *APOL1* risk allele and systolic (SBP) or diastolic (DBP) blood pressure, we divided the cohorts in 20-year age groups (20-39, 40-59 and 60-79). In the 20-39 yo AAs without BP medication, the risk allele accounted for 2-3 mmHg SBP elevation (Ndis=946; Pdis=0.04; Prep=6x10-4 with sex and BMI as covariates; Ndis=822; Pdis=0.07; Prep=0.04 with sex, BMI and eGFR as covariates). AAs with the risk allele in Bio*Me* were more likely to manifest hypertensive organ damage including concentric left ventricular hypertrophy by echocardiogram (OR [95% CI]=1.57[1.18-2.08] P<0.01) and hemorrhagic cerebrovascular accidents (OR [95%CI]=2.66[1.31-5.38] P=0.01) with sex, age and BMI as covariates.**Conclusion:** Compared to other known BP loci, *APOL1* is a large effect size locus for SBP in young AAs that may account in part for well-documented higher BP and health disparities associated with HTN in younger AAs.

515T

MyBPH acts as a modifier of hypertrophy in patients with Hypertrophic Cardiomyopathy. *C.J. Kinnear[1], JC. Moolman-Smook[1], A. Goosen[2], M. Revera[3], PA. Brink[2], L. van der Merwe[4], JM. Mouton[1].* 1) SAMRC Centre for Tuberculosis Research, Department of Biomedical Sciences, Stellenbosch University, Cape Town, Western Cape, South Africa; 2) Department of Medicine, Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa; 3) Department of Cardiology, IRCCS San Matteo Hospital, Pavia, Italy; 4) Department of Statistics, Faculty of Natural Sciences, University of the Western Cape, South Africa.

Hypertrophic cardiomyopathy (HCM) is considered a model disease to study molecular factors underlying isolated cardiac hypertrophy. HCM manifests with various phenotypes and clinical symptoms, even in families with the same genetic defects, suggesting that additional factors contribute to the disease phenotype.Recently MyBPH, a sarcomeric protein, was shown to be involved in cardiomyocyte contraction. MyBPH is structurally similar to cardiac myosin binding protein C (cMyBPC), of which the encoding gene, *MyBPC3*, is one of the most frequently implicated genes in HCM. Given the similarity between cMyBPC and MyBPH, and the critical role MyBPH plays in sarcomere contraction, we propose that MyBPH may be involved in the modulation of cardiac hypertrophy.A family-based genetic association analysis approach was used to investigate the hypertrophy modifying effects of seven single nucleotide polymorphisms (SNPs) and haplotypes in *MyBPH*, in 256 individuals from 27 families, in which three unique South African HCM-causing founder mutations (p.R403W and pA797T in β-myosin heavy chain gene (*MYH7*) and p.R92W in the cardiac troponin T gene (*TNNT2*)) segregate. Each participant underwent extensive clinical screening which included assessments of ventricle wall thickness, maximal posterior wall thickness and maximal intraventricular septum thickness by 16 2D-echocardiographic measurements taken at the mitral valve-, papillary muscle- and supra-apex levels. SNP genotyping was done using the PCR-based KASPTM genotyping assays. Genetic association between each variant and haplotype and hypertrophy traits chosen to represent the HCM hypertrophic phenotype was analyzed with a specialized mixed-effects linear model.Significant evidence for association between three SNPs and hypertrophy traits was observed when stratifying the cohort according to HCM-founder mutations. Interestingly, we observed associations between SNPs *rs2250509*, *rs3737872* and *rs762625* and hypertrophy traits only in the p.A797TMYH7 HCM founder mutation group. Furthermore, we found that several haplotypes had markedly different significant effects in two of the HCM founder mutation groups (p.A797TMYH7 and p.R92WMYH7) and in the total cohort.Our findings suggest that the hypertrophic phenotype of HCM is modulated by the compound effect of a number of variants and haplotypes in *MyBPH*. These novel results therefore provide a basis for future studies to investigate the risk profile of hypertrophy development.

516F

Functional genomic analysis of blood and cardiac tissue during ischemia reperfusion (IR). *D. McDaniel[1], D. Rigney[1], X. Zhou[2], A. Simeone[1], L. McDaniel[3].* 1) Dept Surgery, Univ Mississippi Med Ctr, Jackson, MS; 2) Dept Pathology; 3) Dept Microbiology/Immunology.

Ischemia/reperfusion (IR) during organ procurement contributes significantly to tissue injury and may cause early organ dysfunction after transplantation. Genes associated with innate immunity are prime activators of early inflammatory responses to an allograft that lead to host-induced inflammation and organ rejection. We hypothesized that Allograft inflammatory factor 1 (AIF-1)/Daintain could promote activation of innate immune response through the activation of cardiac TLRs released during organ procurement and may contribute to allograft dysfunction. We have investigated the impact of IR injury in an *in vitro* study of human heart tissue during the process of transplantation. In addition, we investigated expression levels of AIF-1 and TLRs during cardiac IR in a rat model of the left anterior descending artery (LAD) occlusion which generates ischemia in left ventricle (LV). The mRNA expression levels of AIF-1 isoform 2 and 3 were significantly increased after 30 min ischemia/60 min reperfusion ($p<0.01$ and $p<0.005$ respectively). TLR-2 expression was minimal but the TLR-4 was non-specific. In LAD study, AIF-1 and TLR mRNA transcripts were significantly increased in a time dependent- manner after IR. These markers were upregulated as early as 10 minutes after reperfusion and further they were increased several-fold after 60 minutes of reperfusion in tissue and peripheral blood cells as compared to the control group. The TLR-2 levels were greater in blood samples, whereas, the TLR-4 levels were greater in cardiac LV after 30 min LADI/60 min reperfusion. In summary, the data supports role of the AIF-1 and TLRs in early activation of innate immune response, thus the control of these genes may have therapeutic potential for strategies preventing allograft dysfunction early on after transplantation.

517W

Rare coding variants associated with blood pressure in ~13,000 individuals of African ancestry. *P. Nandakumar[1], D. Lee[1], F. Tekola-Ayele[7], B. Tayo[6], E. Ware[4], C.C. Gu[5], M. Fornage[2], S. Kardia[4], C. Rotimi[7], R. Cooper[6], D.C. Rao[5], A. Morrison[2], G. Ehret[1,3], A. Chakravarti[1].* 1) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD; 2) Institute of Molecular Medicine and Human Genetics Center, University of Texas Health Science Center, Houston, TX; 3) Department of Specialties of Medicine, Geneva University Hospitals, Geneva, Switzerland; 4) Department of Epidemiology, University of Michigan, Ann Arbor, MI; 5) Division of Biostatistics, Department of Medicine, Washington University School of Medicine in St. Louis, St. Louis, MO; 6) Departments of Public Health Sciences and Medicine , Loyola University Stritch School of Medicine , Maywood, IL; 7) Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Hypertension is a major risk factor for cardiovascular disease. Several genome-wide association studies (GWAS) have implicated common variants at ~80 loci in systolic and diastolic blood pressure (BP) regulation across multiple populations, but the specific genes at these loci are unknown. This study focuses on identifying these and other specific BP genes associated with BP traits in individuals of African ancestry using six cohorts (GenNet, GENOA, HyperGEN, ARIC, HUFS, LUC) genotyped on the Illumina HumanExome BeadChip v1.0 (n=247,870 variants) or v1.1 (n=242,901). Data cleaning yielded a final set of 12,922 individuals and 162,702 variants for analysis. Fixed-effects meta-analyses were performed with the seqMeta R package for SBP and DBP, each adjusted for age, age-squared, sex, BMI and the first 10 principal components of ancestry from each cohort. These analyses include single variants analyzed by frequency class (common: $MAF\geq0.05$, low frequency: $0.01\leq MAF<0.05$ and rare: $MAF<0.01$), gene-based analysis with SKAT and the T1 burden test on variants with $MAF<0.01$, and those considered "damaging" (defined as intronic splice, NMD-compatible stop-gain, and conserved missense (phyloP>4)). Results were multiple test corrected by the Bonferroni method.**Results**: Associations of rare variants in *AFF1*, encoding an mRNA transcription elongation protein, (rs142319329, DBP, P=3.65x10-7) and *XPOT*, encoding an exportin, (rs140705696, DBP, P=2.58x10-7) were identified, with six and four copies across the six cohorts, respectively. Further, associations of rare variants with two to four carriers in *KIF21B*, *KRBA1* and *OR51B2* were identified for SBP. *XPOT* was also found to be significant in the SKAT test for both SBP (P=1.37x10-6) and DBP (P=3.47x10-7) and in the T1 test on damaging SNPs (P=4.33x10-6). A previous study identified an association between a locus containing *Xpot* and atrial weight index in atenolol-treated mice; the group further demonstrated expression of this gene in cardiac tissue. Additionally, for DBP, *TEC*, encoding a tyrosine kinase, was associated in the T1 test on all variants with MAF<0.01 (P=2.31x10-6), and *QSOX1*, encoding a sulfhydryl oxidase enzyme, was associated in the T1 test on damaging variants (P=3.74x10-6). Recent experiments support cardiovascular function for *QSOX1*, including its induction of vascular smooth muscle cell migration and proliferation *in vitro*. These results suggest credible candidate genes for BP regulation.

518T

The contribution of common and low-frequency/rare variants in ATP-binding cassette A1 *(ABCA1)* to lipoprotein-lipid traits. *V. Niemsiri[1], X. Wang[1], Z.H. Radwan[1], D. Pirim[1], J.E. Hokanson[2], R.F. Hamman[2], M.M. Barmada[1], F.Y. Demirci[1], M.I. Kamboh[1].* 1) Human Genetics, Graduate School of Public Health, Pittsburgh, PA; 2) Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO.

Coronary heart disease (CHD) is the leading cause of deaths globally. Risk factors for CHD include abnormal lipoprotein-lipid levels (i.e., high low-density lipoprotein cholesterol [LDL-C] levels, low high-density lipoprotein cholesterol [HDL-C] levels and high triglycerides [TG]). ATP-binding cassette A1 (ABCA1), a transmembrane transporter, plays an important role in the reverse cholesterol transport, specifically in the delivery of free cholesterol and phospholipids to form nascent HDL-C particles and in the cholesterol efflux. In humans, defects in *ABCA1* cause low levels of HDL-C, including Tangier disease (TD; MIM: 2054000). In addition to HDL-C deficiency, an increased risk of CHD has been observed in TD. In this study, we aimed to identify common (minor allele frequency [MAF] ≥5%) and low-frequency/rare variants (MAF<5%) and to evaluate the genetic contribution of *ABCA1* variants to plasma lipoprotein-lipid levels. We resequenced 50 exons and exon-intron boundaries of *ABCA1* in 95 individuals with extreme HDL-C levels (≤10th and ≥90th %tile) followed up by genotyping of selected variants in the entire sample of 623 US Non-Hispanic Whites. Subsequently, we examined the association between genotyped *ABCA1* variants with four lipoprotein-lipid traits (HDL-C, LDL-C, total cholesterol [TC], and TG). We identified 404 sequence variants (including 37 novel variants), and then selected 250 variants (216 from sequencing stage and 34 from HapMap-CEU database) for genotyping. A total of 182 bi-allelic variants (MAF≥5%, n=116; MAF<5%, n=66) passed genotyping quality controls and were advanced in downstream analyses. Gene-based test showed evidence of association between 182 *ABCA1* variants and TG (P=0.0108). Single-site association analyses revealed 26 common *ABCA1* variants nominally associated (P<0.05) with at least one lipid trait, of which 17 (16 with TG; best SNP: rs2066716 [p.Thr1427Thr], P=0.0016 and 1 with TC; rs4743763, P=0.0005) remained significant after adjustment for multiple testing. Haplotype association analyses identified 60 haplotype windows and 16 regions significantly associated with at least one lipid trait (global P<0.05). Rare variant analyses demonstrated evidence of lipid-association for multiple sets of low-frequency/rare *ABCA1* variants. Our findings support the role of *ABCA1*–common and low-frequency/rare variants—contributing to plasma lipoprotein-lipid levels in the general population.

519F

Diastolic blood pressure and intraocular pressure gene variants in relation to primary open-angle glaucoma. *L.R. Pasquale[1], H. Aschard[2], J.H. Kang[3], J.C. Bailey[4], R.R. Allingham[5], J. Haines[4], P. Kraft[2], P. Hysi[6], M. Hauser[5], J.L. Wiggs[1], NEIGHBORHOOD Consortium.* 1) Ophthalmology, Harvard Medical School, Boston, MA; 2) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 3) Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 4) Department of Epidemiology and Biostatistics, Institute of Computational Biology, Case Western Reserve University School of Medicine, Cleveland, OH; 5) Department of Ophthalmology Duke University Medical Center, Durham, NC; 6) Department of Twin Research and Genetic Epidemiology, King's College London, UK.

Purpose: Alower ocular perfusion pressure, derived as diastolic blood pressure (DBP)minus intraocular pressure(IOP),is regarded as a strong risk factor for primary open-angle glaucoma (POAG). We evaluated whether single nucleotide polymorphisms (SNPs) associated with IOP orDBP show association with POAG using data from the NEIGHBORHOOD consortium,whichincludes3,853POAGcases and 33,495 controls. Methods: We selected from two pivotol meta-analyses, 7 independent SNPs associated with IOP and 27 associated with DBP at the genome-wide significance level.We tested the association of those SNPs with POAG and also generated a genetic risk score(GRS),defined as the weighted sum of IOP-increasing and DBP-decreasing alleles. We considered weights equal to the effect estimates from the original DBP and IOP studies, and equal weights. We also considered agender-specific effect, which we evaluated through an interaction analysis.Results: Based on the original reports, the DBP SNPs and IOP SNPs collectively explain 2.2% and 1.2% of the variance in the respective traits. Overall, IOP SNPs showed strong association with POAG -fiveof seven SNPs weresignificantat the Bonferroni correction levelof 0.0015(strongest association equals OR=1.41, 95% CI: 1.28, 1.56; P=5.9x10-13). The weighted andunweightedGRS were also highly significant (OR=2.38, 95% CI: 2.02-2.81; P=3.8x10-26and OR=1.19, 95% CI: 1.15-1.23;P=2.9x10-27, respectively). In contrast, only one DBP SNP demonstrated significance at the P=0.0015 level(OR=1.17, 95% CI: 1.09-1.25; P=6.2x10-6). Neither DBP GRSwassignificantly associated with POAG (P>0.28).No gender interaction effect was identified between any GRS or any of the SNPs. Conclusion: In agreement with the existing literature, we confirmed that POAG shares genetic basis withIOP. Conversely, it appears thatthemost establishedDBP SNPshave only limited contribution to POAG. .

520W

Sequencing of *PEAR1* to identify rare and novel genetic determinants of platelet aggregation. *M.A. Taub[1], L.R. Yanek[2], K. Iyer[3], I. Ruczinski[1], D. Vaidya[2], N. Faraday[2], D.M. Becker[2], L.C. Becker[2], R.A. Mathias[2].* 1) Biostatistics, Johns Hopkins, Baltimore, MD; 2) The GeneSTAR Program, Johns Hopkins School of Medicine, Baltimore, MD; 3) Johns Hopkins School of Medicine, Baltimore, MD.

Platelet aggregation is heritable, and GWAS have found strong replicated associations of platelet aggregation to multiple agonists with a common intronic variant (rs12041331 (SNP1)) of the platelet endothelial aggregation receptor1 gene (*PEAR1*) in African Americans (AA) and European Americans (EA). Here, we used sequencing to refine the GWAS locus and identify additional rare/novel determinants of platelet aggregation in AA and EA subjects.Whole genome sequence (WGS) data on 214 unrelated GeneSTAR subjects (44% AA) was subset to chr1:156853490-156896226 to include *PEAR1*. These genomes were phased, and imputation was done in additional GeneSTAR subjects (N=1203 AAs and N=1990 EAs) using ~1M framework GWAS SNPs and the 214 GeneSTAR genomes as reference haplotypes. Tests for association for 28 platelet aggregation phenotypes were done in (1) the set of 214 subjects with WGS; and (2) the set of 2193 subjects with imputation, stratified by race. Phenotypes included platelet aggregation at baseline and post aspirin intervention (81mg/day for 14 days) across a range of doses of three agonists (collagen, epinephrine and ADP).The WGS data showed a strong association of a novel rare variant (MAF=1%) at chr1:156872148 in AAs ($p = 1.15 \times 10\text{-}9$) with aggregation to collagen. Interestingly, it is located ~2kb away from SNP1 also in intron 1, but is not in linkage disequilibrium with it. In the larger imputed dataset, SNP1 gave the strongest signal for AAs in 27 phenotypes ($p< 5.58 \times 10\text{-}7$ for all, MAF=0.36). SNP1 was also significant in EAs across multiple agonists ($p < 7.92 \times 10\text{-}6$, MAF=0.09). While in AAs, no imputed common exonic variants showed strong signals, in EAs we observed three common exonic variants (MAF=0.13-0.17) with strong association (rs822442, rs41273221 and rs56260937, $p < 8.63 \times 10\text{-}6$, D' with SNP1 = 0.4) for aggregation to epinephrine and ADP. rs822442 is missense; rs41273221 is missense and labeled "possibly damaging" by PolyPhen.This study confirms prior work by our group which demonstrated on an independent sample (N=104) that in reference to *common* variants (1) rs12041331 is a robust determinant of platelet aggregation across all agonists in AAs and EAs, and in AAs is the single strongest determinant; and (2) in EAs, there may be additional exonic variants that determine platelet aggregation. We show for the first time a rare novel variant plays a role specific to aggregation to collagen in AAs, also found in intron 1 of *PEAR1*.

521T

Exome sequencing identifies a possible new candidate gene associated with thoracic aortic aneurysms and dissections. *Y. Wan[1], J.A. Aragon-Martin[1], M. Simpson[2], N. Morsy[1], L.J. Collins[1], D. Osborn[3], J. Bharj[3], A. Saggar[4], D. Milewicz[5], M. Jahangiri[6], E.R. Behr[1], A. Child[1].* 1) Cardiovascular and Cell Sciences Research Institute, St. George's Hospital, University of London, United Kingdom; 2) Genomic Medicine Group, Division of Genetics and Molecular Medicine, Kings College London, United Kingdom; 3) Human Genetics Research Centre, St. George's Hospital, University of London, United Kingdom; 4) Clinical Genetics Unit, St George's University of London, United Kingdom; 5) Department of Internal Medicine, University of Texas, USA; 6) Department of Cardiothoracic Surgery, St. George's Healthcare NHS Trust, London, United Kingdom.

In the absence of syndromic thoracic aortic aneurysms and dissections (TAAD) such as Marfan, Loeys-Dietz and Ehlers-Danlos syndromes, ~20% is inherited with variability in penetrance and severity as well as male gender. Currently, 18 genes have been reported which account for ~25% of all cases and include; *ACTA2*, *COL3A1*, *EFEMP2*, *FBN1*, *FLNA*, *FOXE3*, *MAT2A*, *MYH11*, *MYLK*, *NOTCH1*, *PRKG1*, *SKI*, *SLC2A10*, *SMAD3*, *TGFB2*, *TGFBR1*, *TGFBR2* and *TGFBR3*. Therefore, a large proportion of cases do not have a disease-associated gene.The SGUL database has collected 100 TAAD probands over a 25 year period, which do not have an associated disease-causing gene. One of the probands had a strong TAAD family history and was analyzed by Sanger sequencing for the most common TAAD associated mutations in *ACTA2*, *FBN1*, *SMAD3*, *TGFBR1* and *TGFBR2* genes. A synonymous rare variant was detected in the *ACTA2* gene (c.246C>T) but did not segregate with affected members in this family. The three most distantly related affected individuals in this family were processed for exome sequencing. All reported TAAD genes were analyzed and no causative mutation in all affected individuals was found. The complete variant list was further filtered by removal of all intronic and intergenic polymorphisms (except splice site-disrupting variants), homozygous variants, heterozygous variants that did not segregate with the disease and MAF of >0.1% in 900 in-house controls, 1000 Genomes and EVS databases. This filtered the candidates to six genes. Confirmation via Sanger sequencing and segregation analysis provided one non-synonymous candidate variant that was found in all affected individuals. This gene is highly expressed in smooth muscle cells, suggesting a role in cardiovascular function. Sanger sequencing of this gene in the remaining 99 TAAD probands found four further variants, one for each of four probands. *In-silico* analyses (SIFT, Polyphen-2, Mutation Taster, UMD-Predictor and Align GVGD) indicated two of the four variations are disease-causing and further testing is being performed. An *in-vivo* knockdown model for this gene using the Zebrafish (*Danio rerio*) was developed by co-injecting two non-overlapping splice-directed morpholino oligonucleotides. Comparison of five days post fertilization injected fish with uninjected controls showed a large pericardial edema, moderate to severe tail curvature and reduced eye and body size, consistent with the recently published paper on a new TAAD gene *MAT2A*.

522F

Discovery of a deleterious variant in *TMEM87B* in a patient with a hemizygous 2q13 microdeletion, uncovers a recessive condition characterized by severe congenital heart defects. *H. Yu[1], C.R. Coughlin II[1], E.A. Geiger[1], E.R. Elias[1], J.L. Cavanaugh[2], S.D. Miyamoto[2], T.H. Shaikh[1,3].* 1) Department of Pediatrics, Section of Genetics, University of Colorado School of Medicine, Aurora, Colorado; 2) Department of Pediatrics, Division of Cardiology, University of Colorado School of Medicine, Aurora, Colorado; 3) Colorado Intellectual and Developmental Disabilities Research Center (IDDRC), University of Colorado School of Medicine, Aurora, Colorado.

Congenital heart defects (CHD) have an estimated incidence ranging from 4/1,000 to 50/1,000 and remain one of the leading causes of infant mortality and childhood morbidity worldwide. Despite significant advances made in medical research, the etiology of CHD is still not completely understood. Here, we describe a patient with CHD including atrial septal defect and restrictive cardiomyopathy along with craniofacial anomalies and developmental delay. Initial screening using chromosomal microarray analysis (CMA) identified a maternally inherited 2q13 microdeletion (chr2:111,406,838-113,102,594; hg19). The subject in this study had many of the features reported in previous cases with recurrent 2q13 deletion. However, the inheritance of the microdeletion from an unaffected mother combined with the low incidence (10%) and milder forms of cardiac defects in previously reported 2q13 microdeletion cases made the clinical significance of the CMA results unclear. We next used whole exome sequencing (WES) in a trio-based analysis of the patient and his unaffected parents. WES identified a paternally inherited *TMEM87B* mutation (c.1366A>G, p.Asn456Asp) in the patient. TMEM87B, a highly conserved, transmembrane protein of currently unknown function, lies within the critical region of the 2q13 microdeletion syndrome. Furthermore, a recent study had demonstrated that depletion of TMEM87B in a vertebrate model organism affected cardiac development and led to cardiac hypoplasia. Thus, by combining molecular testing results from CMA and WES, we were able to uncover an autosomal recessive disorder characterized by a severe cardiac phenotype and caused by compound heterozygous mutations in *TMEM87B*. This study not only expands the spectrum of phenotype associated with 2q13 microdeletion, but it also further confirms the role of *TMEM87B* in the etiology of 2q13 microdeletion syndrome, especially the cardiac pathology. Our results further suggest that patients with 2q13 deletion who also have signs of cardiac defects should be subject to TMEM87B mutation screening.

523W

Rare variants in primary cilium genes may contribute to the risk of AVSD in children with Down syndrome. *H. Corbitt[1], S. Sherman[2], E. Feingold[3], R. Reeves[4], M. Zwick[2], C. Maslen[1].* 1) Knight Cardiovascular Institute, and Department of Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR; 2) Department of Human Genetics, Emory University, Atlanta, GA; 3) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 4) Department of Physiology and McKusick Nathans Institute for Genetic Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD.

Down syndrome (DS) is the most common aneuploidy in humans, 1 of every 700 live births, yet the etiology of the comorbidities associated with DS is poorly understood. Congenital heart defects are the most frequently occurring birth defect seen in children with DS. In particular, the risk of atrioventricular septal defect (AVSD) is 2000-fold higher in children with DS compared to the general population. Atrioventricular septation is the mechanism that forms the heart's valves and septum in fetal development. About half of those with DS have abnormal septation events during development leading to AVSD, whereas the other half have healthy hearts. The partial prevalence of AVSD exhibits that trisomy 21 acts as a significant risk factor, but other modifiers must contribute to reach the disease state.A whole exome sequencing study of a cohort composed of cases, DS with an AVSD, and controls, DS without a heart defect, was performed to identify the genetic variants that influence this risk. We used PLINK to identify the single rare variants that are associated with AVSD. We also implemented a kernel-based method, SKAT-O, to take a gene-based approach to test the cumulative effects of rare variants. Computational tools, Mutpred and CADD, were used to predict deleteriousness of each variant.Our study coincides with a recent paradigm shift that postulates the second heart field (SHF) is responsible for the septation process, instead of remodeling of endocardial cushions. Multipotent progenitor cells from the SHF migrate and differentiate into cells that populate specific regions of the heart, such as the AV canal. The differentiation of the SHF is mediated by Sonic hedgehog (Shh) signaling through the primary cilium. We discovered four case-specific rare variants in genes that are crucial components of the ciliome (*CROCC, GLI3, CEP290*, and *ZDHHC20*). These variants are recurrent and found in 28% of the cases in this study. All four variants result in a missense change and are predicted deleterious by multiple computational algorithms. Notably, GLI3 is a transcriptional activator and a repressor of the Shh pathway and its target genes control the specification of the SHF cell fate**.** CEP290 is a major component of the primary cilium and is required for ciliogenesis as well as ciliary membrane composition. These preliminary findings imply that genes within the ciliome are a promising candidate group that may play a significant role in the pathogenesis of AVSD.

524T

Prediction of causal regulatory DNA variants regulating the QT interval phenotype. *D. Lee[1], A. Kapoor[1], A. Chakravarti[1,2].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Center for Complex Disease Genomics, Johns Hopkins University School of Medicine, Baltimore, MD.

The QT interval (MIM 610141), a measure of cardiac repolarization, is a heritable quantitative trait that has major implications for cardiovascular diseases, such as sudden cardiac death (MIM 115080) and arrhythmias (MIM 115000). Genome-wide association studies (GWASs) have identified 68 common variants at 35 loci significantly associated with QT interval variation in individuals of European ancestry. However, identification and characterization of the causal variants remains a key challenge because most of these variants are in noncoding DNA. To directly address this problem, we adopted a new computational method, deltaSVM, which can accurately predict the effect of regulatory variation from DNA sequence (Lee D†, Gorkin DU†, Baker M, Strober BJ, Asoni AL, McCallion AS*, Beer MA*. A method to predict the impact of regulatory variants from DNA sequence. *Nat Genet* (*in press*)). After removing less common variants (MAF < 0.1), we performed an in-depth analysis of the 59 QT interval associated GWAS SNPs at 32 loci. Starting with all common variants (MAF > 0.1) within LD blocks containing the GWAS SNPs, we focused on SNPs in high LD (|D'| > 0.8 and r2> 0.2) with the sentinel SNPs, resulting in 4,968 potential candidates. We demonstrated that these SNPs have three significant features: 1) They are significantly enriched in DNaseI hypersensitive sites (DHSs) in cardiac tissue, identified by DNaseI-seq on adult heart from the NIH Roadmap Epigenomics project (Pbin(X>=662 | N=4,968, p=0.05) < 2.2×10-16). 2) Significant number of the 662 candidate SNPs in heart DHSs have larger deltaSVM scores than expected by chance (Pbin(X>=103 | N=662, p=0.1) < 5.25×10-6). Notably, when each locus was individually evaluated, *SCN5A* (MIM 600163), *KCNH2* (MIM 152427), *ATP1B1* (MIM 182330) with 10% FDR and *NOS1AP* (MIM 605551), *TTN* (MIM 188840), *FADS2* (MIM 606149) with 15% FDR are predicted to have multiple significant SNPs suggesting that the combinatorial effect of multiple regulatory variants often drives the GWAS signals. 3) Many of these predicted regulatory SNPs are predicted to disrupt transcription factor binding sites (TFBSs) important for cardiac tissue, such as MEF2 and GATA, as well as generic TFBSs (AP1, ETS, CTCF, and ZEB1), demonstrating the biological relevance of our prediction. For validation, we are currently focusing on ~100 SNPs at the *SCN5A* locus using reporter assays to assess them for regulatory activity in cardiac cell lines.

525F

Familial Combined Hyperlipidemia Is a Genetically Heterogeneous Disorder. *J.T. Rämö[1], P. Ripatti[1,2,4], S. Söderlund[3], I. Surakka[1], N. Matikainen[3,8], M. Pirinen[1], A-P. Sarin.[1,5], S.K. Service[6], R.K. Wilson[7], A. Palotie[1,2,4], N.B. Freimer[6], M-R. Taskinen[3], S. Ripatti[1].* 1) Institute for Molecular Medicine Finland FIMM, University of Helsinki, Helsinki, Finland; 2) Program in Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 3) Heart and Lung Centre, Helsinki University Hospital and Research Program's Unit, Diabetes & Obesity, University of Helsinki, Helsinki, Finland; 4) Psychiatric and Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, USA; 5) National Institute for Health and Welfare, Helsinki, Finland; 6) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California, USA; 7) The Genome Institute, Washington University School of Medicine, St. Louis, Missouri, USA; 8) Abdominal Center, Helsinki University Hospital, Helsinki, Finland.

Familial combined hyperlipidemia (FCHL), a complex and common familial dyslipidemia defined by premature coronary heart disease (CHD) in the proband and elevated total cholesterol and/or triglyceride levels in at least twofamily members, substantially increases cardiovascular disease risk (OR 5.1). The underlying genetic architecture for this disorder remains unclear, and no clearly causal variants have yet been identified. We collected and carefully phenotyped 53 Finnish FCHL families with 1161 individuals. We genotyped and imputed 9 million genetic variants in a subset of 715 family members, selected to be most informative for genetic analyses. We first tested the enrichment and effect size of common, low-frequency and rare variants in the families. Next, we constructed weighted gene scores for LDL-C and TG based on 212 previously reported SNPs across the whole variant frequency spectrum, for which previous studies have reported association to these phenotypes. We used linear regression with the continuous scores to assess the fraction of polygenic elevated LDL-C and TG in the families.In a case-control setting, 12 mostly low-frequency variants were over1.5-fold enriched in probands. For 9 of these variants, this is the first reported association to FCHL. Affected subjects had elevated gene scores compared to non-affected family members. The scores explained up to 70 % of family-specific LDL-C or TG variation. In total, the known common lipid SNPs explained 27 % and 19 % and low-frequency SNPs 3.3 % and 4.1 % of LDL-C and TG variation, respectively.Our study shows that FCHL is highly polygenic, supporting the hypothesis that variants across the whole allele frequency spectrum contribute to this complex familial trait. Our results support the clinical tenet that FCHL is a cluster of overlapping genetic defects and metabolic dysfunctions instead of an etiologically homogenous disease entity.

526W

Pathogenetic study at the intersection of Marfan syndrome and autosomal dominant polycystic kidney disease. *D. Schepers[1], C. Golzio[2], E. Davis[2], C. Claes[1], E. Reyniers[1], A. Raes[3], N. Katsanis[2], L. Van Laer[1], B. Loeys[1].* 1) Center for Medical Genetics, University of Antwerp and Antwerp University Hospital, Antwerp, Edegem, Belgium; 2) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC, USA; 3) Department of Pediatrics, Ghent University, Ghent, Belgium.

Thoracic aortic dissections are among the most life threatening forms of cardiovascular disease. Thoracic aortic aneurysm, preceding dissection, is a prominent clinical feature of several heritable connective tissue disorders, including Marfan syndrome (MFS). MFS is caused by mutations in *FBN1*, which encodes fibrillin-1, an important extracellular matrix protein. Mutations in *PKD1* or *PKD2*, two polycystin encoding genes, are responsible for autosomal dominant polycystic kidney disease (ADPKD). Aortic and arterial aneurysms also occur in ADPKD. *Vice versa*, kidney cysts have also been observed in MFS. This clinical overlap suggests a mechanistic link between ADPKD and MFS.Here we describe a four generation family with nine affected individuals presenting with both thoracic aortic aneurysm and mild cystic kidneys. *FBN1*, *PKD1* and *PKD2* were excluded as disease causing genes by linkage analysis and/or sequencing. Subsequently, whole genome linkage analysis resulted in the delineation of a unique linked region on chromosome 16q21-q24.1. Exome sequencing was performed but no putative causal variants were found in the linked region. Copy number variation analysis identified two duplicated regions in the linkage interval, one (chr16: 86862531-870228808) gene-less and the other (chr16: 86357163-86725305) containing seven genes, including three genes encoding transcription factors of the FOX gene family (*FOXC2*, *FOXF1*, *FOXL1*), one gene encoding a methenyl tetrahydrofolate synthetase containing domain protein (*MTHFSD)* and three long non coding RNAs (*LOC732275*, *FENDRR, FLJ30679)*. The presence of the first duplication and its segregation in the family was confirmed using Multiplex Amplicon Quantification analysis. By overexpressing these genes separately and in combination with each other in zebrafish, we will explore the pathogenic mechanisms underlying not only MFS and ADPKD, but aneurysm and cyst formation in general. Preliminary results of these experiments, revealing a potential role for *FOXF1* in cyst formation, will be presented.

527T

***STEM9*, a novel non-coding RNA on chromosome 9p21, is downregulated in coronary artery calcification.** *S.K Sen[1], C. Oguz[1], J.D. Vargas[1], JJ. Barb[2], S.Y. Gebreab[1], S.K. Davis[1], J.J. Johnston[2], L.G. Biesecker[2], G.H. Gibbons[1], A.R. Davis[1].* 1) Cardiovascular Disease Section, National Human Genome Research Institute, Bethesda, MD; 2) National Institutes of Health, Bethesda, MD.

BackgroundFor almost a decade, multiple GWAS have validated a region of chromosome 9p21.3 as being very likely to have an etiological connection with cardiovascular disease (CVD). However, the GWAS lead SNPs are located in a gene desert and the mechanistic link connecting risk alleles to CVD remains a mystery. Here, we report a new transcript in this region with possible links to coronary artery calcification (CAC), a definitive and heritable CVD morphological marker.ResultsTo explore unannotated transcripts in 9p21.3, we used RNA-Seq to compare peripheral blood transcriptomes of eight advanced CAC subjects and eight matched controls, focusing on a 100Kb region surrounding the rs1333049 CAC GWAS lead SNP. A single ~350 bp transcript was discovered to be expressed above background noise levels, which was not included in any current gene or lncRNA annotation databases. ENCODE strand-specific RNA-Seq data showed this transcript to be located on the plus strand. Within the discovery set of CAC cases and controls, the transcript was significantly downregulated in cases (p=0.002). Analysis of RNA-Seq data from individual leukocyte populations in the GEO database revealed that the transcript is highly expressed in monocytes. By integrating Illumina 2.5M SNP chip data with RNA-Seq expression levels, we located an eQTL genomic variant located 18Kb from the transcript (p=0.03). Computational analysis of the transcript sequence revealed a high-confidence binding site for the *MZF1* transcription factor. Currently, we have applied for the name *STEM9* (Sense Transcript Expressed on Monocytes in 9p21) for this transcript.ConclusionsWe report the finding of a new transcript in a locus of extreme interest for CVD, along with preliminary evidence for its downregulation in CAC. Although our results reflect early stages of a work in progress, they merit deeper investigation, given the missing causality that persists despite repeated GWAS validation of the 9p21 locus.

528F

Novel MYLK gene mutation in a Large Family with Fatal Aortic Aneurysm and Dissection: Delineation of the Clinical phenotype. *A. Shalata[1,2], Z. Borochowitz[1], M. Mahroom[1], V. Adir[1], N. Assy[3], R. Fell[4], H. Cohen[5], M. Nashashibi[6], Y. Sova[6], A. Libov[5], M. Azab[1], D. Gieger[7], G. Habib[8], W. Nsier[9].* 1) The Simon Winter Institute for Human Genetics, Bnai-Zion Medical Center, P.O.Box 4940, Haifa,; 2) Genetic Unit Ziv Medical Center, Safed, Israel; 3) Liver Disease Unit, Ziv Medical Center, Safed Israel; 4) Research Unit, Galilee Medical Center, Naharia, Israel; 5) Pathology Dep. Galilee Medical Center, Naharia, Israel; 6) Pathology Dep. Carmel Medical Center, Haifa, Israel; 7) Science Dep. Technion Inst. of Technology, Haifa Israel; 8) Internal Medicine Dep. Carmel Medical Center Computer; 9) The Internal Medicine Dep. The Nazareth English Hospital.

Herein we describe the complexity of the MYLK gene mutation effect on the clinical presentation of thoracic aortic aneurysm and dissection (TAA, TAAD) in a large multi-generational consanguineous Arab family. The patients developed splanchnic, iliac and renal arteries dissection and subclavian and innominate arteries aneurysm.The disease presents in two distinguished clinical groups (G-A and G-B). G-A: apparently healthy young family members and at the 4th decade they suffered of severe overwhelming chest pain and dyspnea that rapidly deteriorated to semi-comatose state. Most of them died within hours to few days after onset. The mode of inheritance of this devastating phenotype fits very well with the autosomal recessive pattern. G-B: Sixth to seventh decade of age asymptomatic individuals, who presented with chronic and compatible with life chest pain and effort induced breathlessness and weakness. They developed aourtic aneurysm and dissection. The mode of inheritance fit very well with the autosomal dominant pattern with incomplete penetrance.Genotyping and gene sequencing study was performed in the available patients and family members. We identified a missense mutation, (c.4471G>T, p.Ala1491Ser) located in exon 27 of the Myosin Light Chain Kinase encoding gene (MYLK gene). This change converts a codon for alanine (GCA) to a codon for serine (TCA) within the protein-kinase domain. Comparison of amino acid sequence we found that the Ala1491 is conserved in 35 out of 35 tested species. For phenotype/genotype correlataion we examined the entire available family members (no. 45) including the known patients, their parents and apparently healthy siblings. The mutation segregation fit very well with the clinical phenotype. Hence, the group A patients were homozygous while, the group B patients where heterozygous for the c.4471G>T (p. Ala1491Ser) mutation.We used the immunohistochemistry staining to study the mutation effect on the MLCK protein expression. We used aortic dissected tissue samples from 2 patients from the family and 2 other unrelated controls. Compared to the controls, the MLCK protein in our family was not detected in one patient and very weak in the other. Both patients were homozygous for the MYLK gene mutation. This finding showed that the mutation we found lead to the MLCK protein instability and to its degradation. We conclude that the c.4471G>T mutation demonstrates very well the loss of function effect in the protein level.

529W

Genetic variants near *IRS1*, body fat percentage and metabolic traits in Hispanics/Latinos from the Hispanic Community Health Study/ Study of Latinos (HCHS/SOL). *Q. Qi[1], S. Gogarten[2], C. Laurie[2], C. Isasi[1], R. Loos[3], R. Kaplan[1].* 1) Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY; 2) Department of Biostatistics, University of Washington, Seattle, WA; 3) Department of Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York City, NY.

A previous genome-wide association study (GWAS) identified rs2943650, a single nucleotide polymorphism (SNP) near *IRS1* (insulin receptor substrate 1), for which the major allele (frequency = 0.64) was associated with lower body fat percentage (BF%) (b= -0.16 ± 0.11% per allele), but with less favorable metabolic profile. The major alleles of SNPs in high LD with rs2943650 (r2>0.8 in Europeans) have also been associated with insulin resistance and dyslipidemia. So far, SNPs near *IRS1* have been predominantly studied in populations of European ancestry, whereas few studies have investigated whether SNPs near *IRS1* affect BF% and metabolic traits in a similar way in US Hispanics. We examined associations of four previously reported SNPs (rs2943650, rs2972146, rs2943641, and rs2943634) with BF% and multiple metabolic traits among up to 12,730 US Hispanics from the HCHS/SOL. Participants were aged 18-74 years from a four-city population sample. In this Hispanic population, the four SNPs were in moderate-to-high LD (r2 = 0.51~0.84), with minor allele frequencies of 20~30%. The major alleles of all four SNPs were associated with lower body fat percent, with the SNP rs2943634 (r2 = 0.65 with rs2943650) showing the strongest result (b= -0.37 ± 0.11% per major allele; P =0.001). The effect size on BF% is greater than twice stronger than that reported among Europeans (b= -0.37% vs -0.16%; P for heterogeneity =0.06). The major allele of rs2943634 was also significantly associated with lower BMI and waist circumference ($P \leq 0.003$), but with higher levels of fasting insulin HOMA-IR, LDL cholesterol and triglycerides ($P \leq 0.05$), and lower HDL-cholesterol (P=0.002). In conclusion, our study suggests a stronger association between genetic variants near *IRS1* and BF% in US Hispanics compared with Europeans. Consistent with previous results, the body-fat-decreasing allele near *IRS1* was associated with unfavorable levels of metabolic traits.

530T

Inter-individual variation in DNA methylation levels in *PLAT* and *STX2* genomic regions contribute to variation in Tissue Plasminogen Activator plasma levels. *N. Zwingerman[1], I. Kassam[1], V. Truong[1], D. Aïssi[2], J. Dennis[1], M. Wilson[3], P. Wells[4], PE. Morange[5], DA. Trégouët[2], F. Gagnon[2].* 1) Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 2) UMR_S 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, INSERM, Paris, France; 3) Genetics and Genome Biology Program, SickKids Research Institute, Toronto, Ontario, Canada; 4) Department of Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada; 5) UMR_S 1062, Nutrition Obesity and Risk of Thrombosis, INSERM, Marseille, France.

Tissue Plasminogen Activator (tPA) is a serine protease that mediates the conversion of plasminogen to plasmin, the major enzyme responsible for endogenous fibrinolysis. In some populations, elevated tPA levels are associated with thrombotic outcomes. tPA plasma levels are highly heritable, with estimates as high as 67%. A meta-GWAS identified variants in three genes involved in tPA regulation: *PLAT*, *STXBP5*, and *STX2*. Collectively, these genetic variants explain less than 1% of tPA variance. DNA methylation in these regions can regulate gene transcription and may account for variation in tPA levels. In the genomic regions identified by the meta-GWAS, we aim to identify DNA methylation levels associated with tPA plasma levels. In discovery sample of 140 individuals from 5 pedigrees ascertained on a single proband with venous thromboembolism, a linear mixed effects regression model adjusted for relatedness matrix and covariates was used, to test for association between DNA methylation levels and tPA plasma levels. DNA methylation sites located within 150kb of *PLAT*, *STXBP5*, and *STX2* gene regions were tested. Sites with FDR<0.1 were tested for replication in a study sample of 186 venous thromboembolism cases. A total of 304 DNA methylation sites were tested for association. In the discovery sample, 13 significant associations were identified (5 within *PLAT*, 1 within *STXBP5*, and 7 within *STX2*). Two associations, 1 in *PLAT* and 1 in *STX2* regions, were replicated in an independent study sample (p<0.03). *PLAT* encodes tPA, the replicated DNA methylation site in the *PLAT* region is located upstream of the gene and in an active genomic regulatory region. The second replicated methylation site is located in the intergenic region of the *STX2* gene, which has been previously shown to be associated with tPA release, and was heritable in our study sample (p<0.001). DNA methylation associations with tPA were robust to adjustment for the identified genetic variants in *PLAT, STXBP5*, and *STX2* from the meta-GWAS. Collectively, the two DNA methylation sites explain 5.7% and 1% of tPA level variation in the discovery and replication samples, respectively. Inter-individual variation in DNA methylation levels in *PLAT* and *STX2* genomic regions contribute to the variation in tPA plasma levels, providing new mechanistic insights on tPA regulation. DNA methylation may have larger effect sizes on intermediate traits than genetic variants because of their more direct relationship.

531F

Novel homozygous BMP9 nonsense mutation causes pulmonary arterial hypertension. *G. Wang, R. Fan, R. Ji, W. Zou, D. Penny, N. Varghese, Y. Fan.* Department of Pediatrics, Baylor College of Medicine, Houston, TX.

Pulmonary arterial hypertension (PAH) is a rare progressive fatal vascular disorder. Genetic predisposition plays vital roles in the development of both familial and idiopathtic PAH with most mutations identified in genes involved in the transforming growth factor beta (TGF-b) signaling pathways including BMPR2 (bone morphogenetic protein receptor type 2), ACVRL1 (activin receptor-like kinase type 1), ENG (endoglin), SMAD8 and SMAD4 genes. As a member of the bone morphogenetic proteins (BMPs), BMP9 (also known as GDF2) plays crucial role in angiogenesis and maintenance of endothelial function. Defects in the BMP9 gene have been documented in hereditary hemorrhagic telangiectasia (HHT), the most common inherited vascular disorder that is often associated with PAH.We report the case of a 5 year-old Hispanic boy who was diagnosed with severe PAH and right heart failure at 3years of age. During the admission in the pediatric intensive care unit, treatment was started with inhaled nitric oxide and intravenous epoprostenol; he was subsequently transitioned to treprostinil, sildenafil, and prophylactic enoxaparin. Now, two years later, the child is asymptomatic on sildenafil, bosentan, treprostinil subcutaneously and warfarin. Genetic screening for all the known PAH related genes revealed a novel homozygous nonsense mutation in the BMP9 gene (c.76C>T; p.Q26X). The child did not have any telangiectasias or arteriovenous malformations; family history also negative. Subsequent parental testing showed that both parents were heterozygous for the same mutation, indicating that the child inherited the BMP9 mutant allele from each parent.To our knowledge, this is the first report about BMP9 mutation in patients with PAH. The homozygous nonsense mutation may account for the early onset and severity of PAH in this patient. The absence of clinical symptoms for PAH in the parents may be due to incomplete penetrance or various expressivity of the BMP9 mutations. Our study expanded the spectrum of phenotypes related to BMP9 mutations.

532W

Association of Glutathione S- transferase M1 and T1 gene polymorphisms in South Indian stroke patients. *Usha. P[1], S. Sultana[1], K. Venkata Kolla[1], P. Vidyullatha[1], P.P. Reddy[2].* 1) Institute of Genetics and Hospital for Genetic Diseases, Osmania University, Begumpet, Hyderabad, Telangana State; 2) Genetics, Bhagwan Mahavir Medical Research Centre, Hyderabad, Telangana State, India.

ABSTRACT Stroke is a multifactorial disease caused by the interaction of several genetic and environmental factorsand a variety of risk factors such as blood pressure, diabetes, advanced age and smoking. Activecigarette smoking has been established as a major risk factor for coronary heart disease and for bothischemic and hemorrhagic stroke and the risk is particularly elevated in younger people. Glutathione S-transferases (GSTs) are the phase II enzymes that catalyze the conjugation of glutathione to a widerange of electrophiles and represent a protective mechanism against oxidative stress. So far noinformation has been provided regarding the role of GSTM1 and GSTT1 gene polymorphisms andischemic stroke in Indian populations. So the present study is carried out to investigate the associationbetween the GST M1 and T1 gene polymorphisms and the risk of stroke in a South Indian population.We genotyped 198 ischemic stroke patients and 162 age matched controls using multiplex polymerasechain reaction. Statistical analysis showed that the frequency of both wild and null genotypes of GSTM1(OR = 0.91, p = 0.68) and GSTT1 (OR = 0.60, P = 0.077) did not differ significantly between control andstroke patients. Further analysis of GSTM1 and GSTT1 genotypes among male smokers and nonsmokers category of stroke patients and control subjects showed a protective role of GSTM1 wild type(OR = 0.44, p = 0.007) and GSTT1 wild type (OR = 0.36, p = 0.01) in smokers.

533T

Sequencing and Functional Validation Identifies Low Frequency Noncoding Variants at Chromosome 4q25 Underlying Atrial Fibrillation. *S. Lubitz[1,2], L. Weng[1,2], N. Tucker[1,2], H. Lin[3,4], C. Roselli[2], I. Christophersen[1], E. Benjamin[3,4], K. Lunetta[3,4], P. Ellinor[1,2].* 1) Massachusetts General Hospital, Boston, MA; 2) Broad Institute, Cambridge, MA; 3) Boston University, Boston, MA; 4) Framingham Heart Study, Framingham, MA.

Background: Atrial fibrillation (AF) is a heritable and prevalent cardiac arrhythmia associated with substantial morbidity. The top AF susceptibility locus is a noncoding region on chromosome 4q25 near *PITX2* that contains 4 distinct susceptibility signals tagged by common variants. We sought to identify causal alleles at the locus.**Methods**: We sequenced a 750kb region on chromosome 4q25 (median read-depth 230X) in 462 individuals of European ancestry from the Massachusetts General Hospital with early onset AF and 464 referent individuals from the Framingham Heart Study. We tested associations between overlapping rolling windows (ranging in size from 2.5kb–15kb) and AF using SKAT. We examined the contribution of variants in different allele frequency groups in stratified analyses. We tested individual variants within regions for association with AF. We tested the allele-specific enhancer activity of the variant regions with a modified luciferase reporter assay performed in a myoblast cell line.**Results**: Of the 7378 variants observed, 5988 (81%) occurred with a frequency of <1%, and 1390 (19%) with frequency ≥ 1%. Using overlapping rolling windows, we identified a specific region associated with AF that contains a known common variant association signal for AF. We refined the peak association to an ~1.5kb region and found that the association was driven by low frequency variants (minor allele frequency 1-5%), of which 3 were nominally associated with AF after adjustment for one another and associated variants at the locus (rs115779909, P=0.05; rs62339085, P=0.006; rs188425751, P=0.007). None of the 3 variants were in linkage disequilibrium with one another (all pairwise r2≤0.002). Conditioning on these variants nullified the association between the primary common variant signal in the region and AF (rs6838973, P=0.006 before adjustment, P=0.13 after adjustment). Expression of constructs containing these low frequency variants demonstrated allele-specific differences in enhancer activity as measured by luciferase activity. Of the identified low-frequency variants, rs188425751 is predicted to disrupt a *PITX2* binding site.**Conclusions**: Low frequency noncoding variants exhibiting enhancer activity explain 1 of 4 known susceptibility signals underlying AF on chromosome 4q25. Sequencing is critical for the identification of causal variation, and rolling window approaches may facilitate the discovery of causal variants in noncoding regions.

534F

Identification of wide-spread allele-specific expression at lipid GWAS loci. *M. Alvarez[1], A. Ko[1,2], E. Nikkola[1], T. Nguyen[1], R. Raulerson[3], Z. Miao[1], M. Civelek[4], F.S. Collins[5], A.J. Lusis[1,4], J. Kuusisto[6], M. Boehnke[7], R.M. Cantor[1], K.L. Mohlke[3], M. Laakso[6], P. Pajukanta[1,2].* 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA; 2) Molecular Biology Institute at UCLA, Los Angeles, USA; 3) Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA; 4) Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, USA; 5) National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA; 6) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 7) Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA.

Recent GWAS have identified 157 loci for serum lipids. However, the actual causal variants and genes remain unknown in most of these regions. We hypothesized that some of the GWAS loci influence lipids via allele-specific transcriptional changes in adipose, a key metabolic tissue in humans. To test our hypothesis, we collected genotype, adipose RNA sequence, and phenotype data from 582 Finnish males participating in the METSIM study, and tested the lipid GWAS loci for allele-specific expression (ASE). For ASE, we counted reads (>30 reads, with >6 at the lower expressed allele) at heterozygous transcribed sites that did not show a mapping bias. Using a paired t test between reference and alternate allele counts and correcting for multiple testing, we identified 33,340 ASEs. We analyzed lipid GWAS SNPs that were either within the reported GWAS gene or in linkage disequilibrium (LD) (r2>0.8) with the reported GWAS SNP. We found an enrichment of genes demonstrating ASE genes in the lipid GWAS regions with 198 (35%) ASE SNPs in 62 (53%) adipose expressed genes in the GWAS regions, compared to 5,842 (18%) SNPs in 3,158 (40%) genes in the non-GWAS regions (P=0.004), respectively. ASE SNPs in the lipid GWAS regions showed an enrichment of H3K4m3 (P=0.0028), and a higher RNA polymerase 2 binding (P=0.017), compared to the non-GWAS ASE regions, suggesting increased transcriptional activation. To find the likely functional SNP(s) at each GWAS locus, we searched for ASE SNPs associated with lipids (P<5.3x10-8). We found 10 such GWAS genes with 54 lipid-associated ASE SNPs (*JMJD1C*, *FADS2*, *FADS1*, *ZNF664*, *APOE*, *HLA-DRB1*, *HLA-DQB1*, *MLXIPL*, *LPL*, and *PLEC*). To identify additional, separate lipid-associated ASE SNPs in the GWAS regions, we explored the ASE SNPs not in tight LD (r2<0.2) with the reported GWAS SNP for lipid associations. We discovered 18 additional ASE SNPs (within 500 kb from the reported SNP) in 15 genes that showed a genome-wide significant lipid association. We also investigated whether the expression of the ASE genes correlated with lipids, and found *APOE*, *CPNE1*, and *KANK2* to be correlated with the reported lipid trait (P<0.05). Taken together, our study in adipose transcriptomes furthers the functional characterization of 19 of lipid GWAS loci with 72 lipid-associated ASE variants in 25 genes. Accordingly, we discovered widespread ASE at the lipid GWAS loci, suggesting ASE as a first clue towards their functional consequences.

535W

Identification of a homozygous mutation in *PLXND1* in patients with a complex phenotype featuring truncus arteriosus and abnormal pulmonary venous return with predisposition to Hirschsprung disease and neuroblastoma. *C. Gordon[1], A. Guimier[1], C. Muller[1], M. Oufadem[1], D. Lehalle[2], C. Bole-Feysot[1], P. Nitschké[1], M. Singh[3], J. Epstein[3], F. Bajolle[4], D. Bonnet[4], P. Bouvagnet[5], S. Lyonnet[1,2], L. de Pontual[1], J. Amiel[1,2].* 1) Institut Imagine, INSERM U-1163, Paris, France; 2) Service de Génétique, Hôpital Necker-Enfants Malades, AP-HP, Paris, France; 3) Department of Cell and Developmental Biology, Cardiovascular Institute, University of Pennsylvania, Philadelphia, PA, USA; 4) Unité Médico-Chirurgicale de Cardiologie Congénitale et Pédiatrique, Centre de référence Malformations Cardiaques Congénitales Complexes - M3C, Hôpital Necker-Enfants Malades, APHP, Paris, France; 5) Laboratoire cardioGénétique - Hospices Civils de Lyon, Bron, France.

Truncus arteriosus (TA) is a rare congenital heart defect caused by defective septation of the outflow tract, associated with significant morbidity and mortality. We performed exome sequencing in three families in which the probands presented with either isolated TA or TA in association with other anomalies. Firstly, we identified a homozygous missense mutation in the GAP domain of *PLXND1* in three consanguineous sibs presenting TA and abnormal pulmonary venous return, associated with Hirschsprung disease (HSCR) in two out of three cases and with neuroblastoma (NB) in one. Investigation in the only surviving member of the sibship revealed generalised vascular anomalies. In the second family, we identified a frameshift mutation in *SMAD6* in siblings with isolated TA, inherited from their normal father, while in the third family, a splice site mutation in *TBX1* was identified in a patient with TA and athymia but without typical craniofacial features of DiGeorge syndrome. The cardiac phenotypes in these patients are consistent with the known roles of *Plxnd1*, *Smad6* and *Tbx1* in development of the conotruncus in mouse models, although heterozygous loss of *SMAD6* in humans appears incompletely penetrant. Our results confirm *PLXND1* mutations as a cause of cardiovascular defects in humans, and HSCR and NB may be associated phenotypes due to disrupted plexin-semaphorin signalling during neural crest development.

536T

Mutations in *TAX1BP3* cause Dilated Cardiomyopathy with Septo-Optic Dysplasia. *E. Reinstein.* Rabin Medical Center, Tel-Aviv, Israel.

We describe a Bedouin family with a novel autosomal recessive syndrome characterized by dilated cardiomyopathy and septo-optic dysplasia. Genetic analysis revealed a homozygous missense mutation in *TAX1BP3,* which encodes a small PDZ-containing protein implicated in regulation of the Wnt/β-catenin signaling pathway, as the causative mutation. The mutation affects a conserved residue located at the core of TAX1BP3 binding pocket and is predicted to impair the nature of a crucial hydrophobic patch, thereby interrupting the structure and stability of the protein, and its ability to interact with other proteins. *TAX1BP3* is highly expressed in heart and brain and consistent with the clinical findings observed in our patients, a knockdown of *TAX1BP3* causes elongation defects, enlarged pericard and enlarged head structures in zebrafish embryos. Thus, we describe a new genetic disorder that expands the monogenic cardiomyopathy disease spectrum and suggests that *TAX1BP3* is essential for heart and brain development.

537F

Impaired *PIEZO1* function in patients with a novel autosomal recessive congential lymphatic dysplasia. *B. Krock[1, 2], V. Lukacs[4], J. Mathur[6], R. Mao[1,2], P. Bayrak-Toydemir[1,2], M. Procter[2], S. Calahan[4], H. Kim[5], M. Bandell[6], N. Longo[8], R. Day[7], D. Stevenson[3], A. Patapoutian[4].* 1) Pathology, University of Utah, Salt Lake City, UT; 2) ARUP Institute for Clinical and Experimental Pathology, ARUP Laboratories, Salt Lake City, UT; 3) Department of Pediatrics, Division of Medical Genetics, Stanford University, Stanford, CA; 4) Howard Hughes Medical Institute, Molecular and Cellular Neuroscience, The Scripps Research Institute, La Jolla, CA; 5) Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA; 6) Genomics Institute of the Novartis Research Foundation, San Diego, CA; 7) Department of Pediatrics, Division of Pediatric Cardiology, University of Utah, Salt Lake City, UT; 8) Department of Pediatrics, Division of Medical Genetics, University of Utah, Salt Lake City, UT.

Defective development and function of lymphatic vessels can result in primary lymphedema, a debilitating accumulation of lymphatic fluid that typically affects the extremities. The known genetic etiologies of primary lymphedema delineate the developmental processes involved in lymphatic development, such as the VEGFR3 signaling pathway. However, the genetic causes of congenital lymphedema are incompletely understood, in particular for sporadic and autosomal recessively inherited forms. This observation suggests that unrecognized developmental pathways may underlie the genetic cause of undiagnosed congenital lymphedemas. Development of the lymphatic system is regulated by interstitial pressure and shear forces induced by fluid flow; however, the molecular mechanism of mechanosensation in these cells is largely unknown. Whole exome sequencing of a pair of siblings affected with an apparent autosomal recessive hereditary lymphedema identified *PIEZO1* as a candidate gene. Both siblings were compound heterozygotes for a splicing and missense variant in *PIEZO1*, a mechanically gated non-selective cation channel that is among the few ion channels reliably demonstrated to convey mechanical cues of the environment to mammalian cells. Functional analysis of the affected individual's cells revealed severely diminished PIEZO1 responses to both mechanical stimuli and a pharmacological PIEZO1-activator. Patch-clamp analysis of the missense PIEZO1 channel in a heterologous system demonstrated strongly diminished current amplitudes as compared to wild type, likely due to reduced cell surface expression. These findings suggest a role for PIEZO1 in human lymphatic development and implicate it as a novel genetic cause of an autosomal recessive congenital lymphatic dysplasia.

538W

Contribution of CRELD1 novel mutations and Polymorphisms in occurrence of AVSD among DS patients seen In North Indian population. *A. Asim[1], I. Panigrahi[2], A. Kapoor[3], S. Agarwal[1].* 1) Genetics, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, India, Lucknow, Uttar Pradesh, India; 2) Advanced Pediatrics Centre,Post Graduate Institute of Medical Science and Education Research ,Chandigarh,India; 3) Cardiology, Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, India, Lucknow, Uttar Pradesh, India.

Down syndrome (DS) is caused by trisomy of chromosome 21, is the commonest genetic disorder characterized by multiple congenital abnormalities of variable severity. Almost, 50% of DS individual have congenital heart defect (CHD) out of which 40-60% has atrioventricular defects (AVSD). CRELD1 (Cysteine-rich with EGF-like domains, OMIM 607170) is the gene identified as candidate gene for AVSD in DS. The aim of the study is to reveal the status ofCRELD1variants in DS with AVSD and without AVSD.A total of 60 patients of DS with AVSD (N = 24) and without AVSD (N=36), along with 10 samples having AVSD alone were recruited. Genomic DNA was extracted from Qiagen Mini kit and Sequencing of exon 9 and 10 were performed in all PCR amplified samples.Novel substitution mutation has been identified in exon 9 at R329H and E325K in DS with AVSD group**.** While, several other polymorphism have also been identified in exon 10 namely rs9878047 (c.1049-129T > C), rs3774207 (c.1119C > T), and rs73118372 (c.1136T > C) among DS with AVSD with the frequency of 4.1%, 50% and 29.1% respectively which has been reported earlier but for the first time in our north Indian population. However, Mutation taster analysis indicates the lethality of these mutations and polymorphisms.The results emphasized the need of screening of exon 9 and 10 of CRELD1 gene among DS with AVSD as the candidate gene marker in Indian patients.

539T

Cost analysis and strategy in molecular diagnosis of Familial Hypertrophic Cardiomyopathy: from Sanger to Next Generation Sequencing. *D.A. Coviello[1], M.I. Parodi[1], S. Davi[1], E. Canepa[1], M. Castagnetta[1], P. Spirito[2], C. Rapezzi[3], C. Autore[4], B. Musumeci[4], F. Formisano[2], S. Forleo[5], S. Favale[5], D. Degiorgio[1], M. Cecconi[1].* 1) Laboratory of Human Genetics, E.O. Ospedali Galliera, Genoa, Italy; 2) Department of Cardiology, E.O. Ospedali Galliera, Genoa Italy; 3) Department of Cardiology, Alma Mater Studiorum-University of Bologna, Italy; 4) Department of Cardiology, St. Andrea Hospital, Sapienza University, Rome, Italy; 5) Cardiology Unit, Department of Emergency and Organ Transplantation, University of Bari, Italy.

Routine clinical use of NGS technologies is helping to perform accurate diagnosis with a faster T.A.T. and with lower costs. In the last seven years we have routinely performed molecular diagnosis of Familial Hypertrophic Cardiomyopathy (FHC) using Sanger sequencing on 11 major genes, 8 sarcomeric and 3 metabolic genes. To move from Sanger to NGS we developed a custom panel of 19 selected pathogenic genes for FHC (*MYBPC3, MYH7, TNNT2, ACTC1, TPM1, TNNI3, MYL2, MYL3, MYH6, TNNC1, MYOM1, MYOZ2, VCL, ANKRD1, CALR3, CAV3, GLA, LAMP2, PRKAG2*) and we used Ion Torrent technology on PGM (Life Technologies).To validate the NGS methodology we analyzed 63 selected FHC patients; in these patients, previously tested by means Sanger, were been already identified 79 mutations in 11 genes out of 19. Among 79 mutations, 3 are located in uncovered regions of the panel and the NGS analysis has identified 75 out of remaining 76. We perform a cost analysis of NGS in comparison with Sanger sequence. We use NGS in the molecular diagnostic of FHC patients and, to date, this technology has allowed us to analyze additional 23 patients; 17 different mutations have been identified in 16 patients. We have identified some critical points: a) new version of the informed consent of the patient to perform the test; c) choice of an appropriate software for analysis of the data generated; d) relevance of collaboration with the clinicians for management of results, especially in large families; e) inevitable increasing identification of new variants with uncertain pathogenic value will affect the expectance of patients, especially in predictive or pre symptomatic testing f) incomplete representation and coverage of target regions lead to missing clinically relevant mutations regions g) necessity of Sanger sequencing to confirm mutant genotypes and to achieve a desirable coverage in GC-reach regions. The validation step of NGS technology permitted us to identify 75/76 (98,5%) mutations on covered target regions. This result has confirmed the validity of this technology and permitted us to introduce the NGS techniques within the diagnostic workflow for molecular analysis of HCM at lower costs. Sanger sequencing is still used to analyze well uncoveded genomic regions and to validate the presence of a causative genotype.

540F

Implementation of diagnostic Whole Exome Sequencing to improve diagnostic yield of genetic testing for patients with thoracic or abdominal aortic aneurysm. *H.T. Bruggenwirth[1], A.S. IJpma[1,2], J.M.A. Verhagen[1], G.M. Bolman[1], J.N.R. Kromosoeto[1], D. Heijsman[1], C.E.M. Kockx[3], F.J.G.T. Sleutels[3], H.J.F.M. Eussen[1], J.E.M.M. De Klein[1], M.A. Van Slegtenhorst[1], W.G. De Valk[3], R.W.W. Brouwer[3], W.F.J. Van IJcken[3], I.M.B.H. De Graaf-Van de Laar[1], M.W. Wessels[1], D.F. Majoor-Krakauer[1].* 1) Dept Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands; 2) Dept of Bioinformatics, Erasmus MC, Rotterdam, The Netherlands; 3) Center for Biomics, Erasmus MC, Rotterdam, The Netherlands.

Thoracic Aortic Aneurysm (TAA) and Abdominal Aortic Aneurysm (AAA) are genetically heterogeneous. A proportion of autosomal dominant inherited TAA syndromes is associated with mutations in genes interfering with the TGFB pathway. Non-syndromic TAA was associated with mutations in genes affecting smooth muscle cell functioning. Much less is known about genes involved in AAA. In our laboratory DNA analysis of several genes (8 genes were offered), involved in TAA and AAA (evidence-based) was previously performed by Sanger sequencing of the coding parts and exon-flanking intronic regions. In general, the diagnostic yield obtained was low. To improve the diagnostic yield we have implemented Whole Exome Sequencing (WES) using HaloPlex exome target enrichment (Agilent Technologies) with Illumina sequencing on a HiSeq platform. For TAA and AAA samples at least 6 gigabases of sequence is generated (75x average coverage). A set of core-genes (*ACTA2* [MIM 611788], *COL3A1* [MIM 130050], *FBN1* [MIM 154700], *SMAD3* [MIM 613795], *TGFB2* [MIM 614816], *TGFBR1* [MIM 609192], *TGFBR2* [MIM 610168]) was selected for which the coding regions should be covered by at least 20 reads, regions with lower coverage are supplemented by Sanger sequencing. For the primary diagnostic analysis a virtual gene panel, consisting of 23 genes known to be involved in aneurysms and a SNP in *MTHFR* known to be a risk factor, is analysed. Using Cartagenia software (Agilent) we select variants, present in those 24 genes. We filter for the presence of potentially damaging heterozygous variants in case of autosomal dominant inheritance and for the presence of compound heterozygous or homozygous variants in case of autosomal recessive disease. If during this first analysis no satisfactory explanation for the phenotype is found, as a second step the exome can be opened (on basis of research), provided that informed consent has been obtained. In this poster, we present our newly developed analysis pipeline and results of the primary diagnostic analysis.

541W

The genetics of inherited cardiac conditions. *B. Chong[1], S. Pantaleo[1], M. Chow[1], J.P. Plazzer[1,4], C. Love[2], P. James[3], D. du Sart[1].* 1) Victorian Clinical Genetics Services, Murdoch Childrens Research Institute, Melbourne, Victoria, Australia; 2) Translational Genomics Unit, Victorian Clinical Genetics Services, Murdoch Childrens Research Institute, Melbourne, Victoria, Australia; 3) Clinical Genetics, Royal Melbourne Hospital, Melbourne, Victoria, Australia; 4) InSIGHT Database Curator, Royal Melbourne Hospital, Melbourne, Victoria, Australia.

Inherited cardiac conditions, such as the cardiomyopathies or arrhythmias, have a complex genetic pathology. They are often heterogeneous, with overlapping clinical symptoms that may not always be penetrant. This adds to the complexity of determining whether gene variants are pathogenic, influence clinical severity or normal variation within the genes. Next generation sequencing (NGS) technology has enabled assessment of the large number and large size of these implicated genes. We use a custom capture NGS panel to screen for the arrhythmias, cardiomyopathies, aortopathies, congenital heart disease and unexplained sudden death conditions, with complete coverage of the entire coding region of all the genes analyzed. In 650 patients tested we identified, 54% pathogenic mutations and 19% variant of unknown significance in the arrhythmia panel; 59% pathogenic mutations and 31% variant of unknown significance in the cardiomyopathy panel; 34% pathogenic mutations and 22% variant of unknown significance in the aortopathy panel; 45% pathogenic mutations and 42% variant of unknown significance in the sudden death panel and the remaining patients in each panel were negative. The data provides detailed information about the relative contribution of the major genes to each phenotype and the distribution of variants within each gene. For some genes we compared the regions covered in our capture NGS panel with the whole exome NGS panel. By analyzing the entire gene panel in all samples, we have been able to gain an insight into the level of normal variation within these genes which in some cases questions the variant pathogenicity classification. We have also defined the key quality parameters to indicate whether a variant is in the patient's genome or an artefact or contaminant generated through the NGS process. With new population data from whole exome analyses in non-disease cohorts and our experience to date, we are now revising our NGS analysis process and improving our variant classification criteria to reflect a more evidence based classification.

542T

Clinical and molecular lessons from targeted next generation sequencing of 51 genes involved in primary electrical disease. *B. Loeys[1, 2], D. Proost[1], J. Saenen[3], G. Vandeweyer[1], A. Rotthier[4], M. Alaerts[1], E. Van Craenenbroeck[3], J. Van Crombruggen[1], G. Mortier[1], W. Wuyts[1], C. Vrints[3], J. Del-Favero[4], L. Van Laer[1].* 1) Center for Medical Genetics, Antwerp University Hospital/University of Antwerp, Antwerp, Belgium; 2) Radboud University Medical Center, Nijmegen, The Netherlands; 3) Department of Cardiology, Antwerp University Hospital/University of Antwerp, Antwerp, Belgium; 4) Multiplicom, Niel, Belgium.

Primary electrical disease (PED) is characterized by cardiac arrhythmias, which can lead to sudden cardiac death in the absence of detectable structural heart disease. PED encompasses a diversity of inherited syndromes, predominantly Brugada syndrome, early repolarization syndrome, long QT syndrome, short QT syndrome, arrhythmogenic right ventricular cardiomyopathy/dysplasia and catecholaminergic polymorphic ventricular tachycardia. To overcome the diagnostic challenges imposed by the clinical and genetic heterogeneity, we developed a targeted gene panel for next generation sequencing of 51 genes involved in PED. Twenty CEPH samples and 20 positive control samples were used to validate the panel. A technical sensitivity and specificity of 100% respectively 99.9% was obtained. After validation, we applied the assay to 114 PED patients. We identified 107 variants in 36 different genes, 18 of which were classified as pathogenic or likely pathogenic, 54 variants were of unknown significance and 35 were classified as likely benign. In the patient groups of BrS, ARVC/D and LQTS we reached causal mutation detection rates of 23% (18/80), 37.5% (3/8) and 36.4% (4/11) respectively. In total, 46.5% of patients (53/114) had a variant, either (likely) pathogenic or VUS, that required additional molecular and clinical follow-up. We hypothesize that several VUS will also be causal because we identified more variants per gene in our PED patients compared to what could be expected by chance, based on the number of variants present in the ExAC database for these genes (eg. 3.7 times more *SCN5A* variants (MAF $\leq$ 0.001) in our PED patients compared to the ExAC database). Our data also support a putative oligogenic nature as in 17 probands we identified more than one (possible) pathogenic variant or VUS. Even more intriguing is the observation that within one family, two different pathogenic mutations cause the identical phenotype in different individuals. The latter advocates for the application of the panel testing in different affected individuals to identify to full underlying genetic architecture. Finally, we identified the first *SCN5A* founder mutation in 17 Belgian families.In conclusion, the PED MASTR Plus assay is a proficient, highly reliable and reproducible technique to routinely screen patients suffering from primary arrhythmias.

543F

Target NGS extended panel for inherited cardiomyopathies: evaluating the increased diagnostic yield in hypertrophic cardiomyopathy. *J.D.C. Marsiglia[1], T.G.M. Oliveira[1], M.M. Neto[2], L.C. Teixeira[3], J.E. Krieger[1], E. Artega-Fernandez[4], A.C. Pereira[1].* 1) Laboratory of Genetics and Molecular Cardiology, Heart Institute, University of São Paulo, São Paulo, Brazil; 2) Fleury Group - Research and Cevelopment Department (R&D), São Paulo, Brazil; 3) CEFAP - Biomedical Science Institute (ICB/USP), São Paulo, Brazil; 4) Clinical Unit of Cardiomyopathies - Heart Institute, University of São Paulo, São Paulo, Brazil.

Introduction: Hypertrophic cardiomyopathy (HCM) is a primary cardiac disease, mainly characterized by unexplained left ventricle hypertrophy, in the absence of dilatation, usually asymmetric and predominantly septal. Up to now, more than 1400 mutations have been associated with HCM in over 20 genes related to sarcomeric myofibrils, Z-disc and calcium homeostasis. The main aim of this work was to compare the diagnostic yield of an expanded gene panel for the identification of mutations related with HCM that includes non-sarcomeric genes. Materials and Methods: 111 samples from a cohort of clinically diagnosed patients with HCM were processed using an Ion Torrent PGM platform for the assessment of the diagnostic yield of a genetic panel composed of 74 genes associated with inherited cardiomyopathies. The sample NA12878, a HapMap reference sample, was used in the assessment of intra- and inter-assay reproducibility in the validation process of a proposed pipeline for the detection of genetic alterations. Variants were considered potentially pathogenic if previously described as associated with HCM or if presenting a deleterious score in at least two of three impact prediction algorithms tested (PROVEAN, SIFT and PolyPhen-2) and MAF<0.01, if available. Results: Of all 111 samples, 44 had negative or inconclusive results, 42 had one or more mutations in sarcomeric genes, 20 had one or more mutations in non-sarcomeric genes, and 5 had two mutations, one in a sarcomeric and the other in a non-sarcomeric gene. The main non-sarcomeric genes harboring causal mutations were *FLNC* and *TRIM63*, both already associated with HCM. Conclusions: We observed an increase in diagnostic yield of 20% over a panel containing only sarcomeric genes. These results suggest that current molecular genetics diagnostic testing for hypertrophic cardiomyopathy should incorporate non-sarcomeric genes.

544W

WES as a first and second-tier test for arrhythmia and cardiomyopathy. *A.A. Singleton, B.E. Friedman, K. Retterer, G. Richard, D. Macaya.* GeneDx, Gaithersburg, MD.

NextGeneration sequencing panels for inherited arrhythmias and cardiomyopathies offer a sensitivity of 30-70% depending on the phenotype targeted. Our aim was to evaluate the positive rate of whole exome sequencing (WES) for these test indications, using WES either as first-tier test or after a negative or inconclusive targeted cardiology panel. We report results of 30 consecutive cases submitted for WES to GeneDx with a clinical indication of non-syndromic arrhythmia, cardiomyopathy, and/or sudden cardiac arrest/death. Record review indicated 13 individuals (43%) had prior testing including at least one targeted panel with negative or uninformative results. Of these 13, WES revealed 6 probands (46%) with pathogenic (P) or likely pathogenic (LP) variants in 6 distinct genes. One gene was newly described and therefore had not been available as part of a cardiology panel. The other genes were not included in the specific multi-gene panel initially chosen by the provider, likely due to phenotypic overlap. In the remaining 7 individuals, WES identified only variants of uncertain significance (VUS). These VUS were identified in genes known to be associated with a cardiogenetic presentation as well as in candidate genes. In contrast, 17 probands submitted for WES did not have prior targeted testing. Of these, 6 (35%) had a positive result (P or LP variants) in 8 distinct genes. Two probands were double mutation carriers. Since 3 out of the 8 genes were newly described, the molecular diagnosis in two of the probands (including one double mutation carrier) would have been missed by traditional panel testing. Three out of 17 probands (18%) had a negative results, and 8/17 (47%) had VUS in known or candidate genes. In summary, WES testing for 30 probands with arrhythmia, cardiomyopathy, sudden cardiac arrest/death, regardless of prior testing, yielded a positive rate of 40% (12/30).The molecular diagnosis would have been missed by traditional panel testing in 3/30 probands (10%). WES proves to be a good second-tier test especially for probands in whom the clinical diagnosis is not clear. A larger systematic comparison between WES and panel testing including diagnostic yield, cost, turn-around time would be necessary to determine the utility of WES as first-line test for these cardiogenetic disorders.

545T

Lack of specificity of ACMG classification rules decreases inter-curator concordance. ClinGen's adaptation of ACMG's framework to standardize interpretation of MYH7 related cardiomyopathy variants. *C. Caleshu[1,2,3], M. Kelly[1,4], A. Morales[1,5,6], E. Ashley[1,2,8], R. Hershberger[1,5,6,7], B. Funke[1,5,9].* 1) On behalf of the ClinGen Cardiovascular Domain Working Group; 2) Stanford Center for Inherited Cardiovascular Disease, Stanford, CA, USA; 3) Division of Medical Genetics, Stanford Medical Center, Stanford, CA, USA; 4) Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Boston, Massachusetts, USA; 5) Division of Human Genetics, The Ohio State University Wexner Medical Center, Columbus, OH, USA; 6) Davis Heart and Lung Research Institute, The Ohio State University Wexner Medical Center, Columbus, OH, USA; 7) Cardiovascular Division, The Ohio State University Wexner Medical Center, Columbus, OH, USA; 8) Division of Cardiovascular Medicine, Stanford Medical Center, Stanford, CA, USA; 9) Department of Pathology, MGH and Harvard Medical School, Boston, Massachusetts, USA.

Background: Recognition that variants in Mendelian disease genes are often classified differently in different labs has signaled the need for advances in variant classification systems. Recently, ACMG published updated guidelines, intended to serve as a framework for classifying Mendelian variants across all areas of genetics. We investigated the application of these new guidelines to *MYH7* and cardiomyopathy.**Methods:** Ten *MYH7* variants were independently reviewed by three members of the ClinGen Cardiovascular Working Group. Reviewers classified the variants using both the new ACMG criteria and criteria previously developed by each reviewer's cardiomyopathy genetics team. This analysis provided the foundation for an in-depth review of the ACMG criteria and the development of adapted rules with optimized specificity for *MYH7*. **Results:** Classification concordance was 8/10 for institutional criteria and 3/10 for ACMG guidelines. Disagreement between institutional classifications arose mainly from privately held data, while disagreement between ACMG classifications was largely due to ACMG criteria being applied differently by reviewers. Six of seven established pathogenic and likely pathogenic variants were classified as variants of uncertain significance with the ACMG criteria by at least one of the three reviewers. Subsequent in-depth review of the ACMG criteria resulted in 18 rule adaptations that are critical to provide optimal specificity for *MYH7*. Nine rules were modified, including changes informed by cardiomyopathy specific characteristics such as incomplete penetrance, prevalence of disease-causing alleles in the general population, and the existence of 1-5% of patients with >1 pathogenic variant. Eight rules were deemed inapplicable to *MYH7*, including loss of function as a known mechanism of disease and mutational hotspots. One rule was added to enable using the number of unrelated cases as a criterion supporting pathogenicity. Fifteen rule revisions were broadly applicable across cardiovascular genetics, reflecting a general lack of specificity.**Conclusions:** The use of ACMG guidelines in their current form can decrease the concordance of variant classifications. Expert review is therefore critical to adapt this new framework to disease domains and to reflect disease and gene specific characteristics and to enable consistent application. Our work serves as a stepping stone for additional genes and diseases.

546F

Whole Genome Sequencing is a clinically effective strategy for Inherited Cardiac Disorders. *P.A. James[1, 2], T. Thompson[1], D. Du Sart[3], R. Leach[4], D. Floerke[4], R. Tearle[4], D. Zentner[5], A. Trainer[5], J. Vohra[5], I. Winship[1,6].* 1) Genetic Medicine, Royal Melbourne Hospital, Melbourne, Victoria, Australia; 2) Department of Pathology, University of Melbourne, Melbourne, Victoria, Australia; 3) VCGS Pathology, Victorian Clinical Genetic Service, Murdoch Children's Research Institute, Melbourne, Victoria, Australia; 4) Complete Genomics Inc., Mountain View, California; 5) Cardiology Department, Royal Melbourne Hospital, Melbourne, Victoria, Australia; 6) Dept of Medicine RMH, Faculty of Medicine Dentistry and Health Sciences, University of Melbourne, Melbourne, Victoria, Australia.

Background: Identifying the underlying mutation(s) in families affected by inherited cardiac disorders allows optimal screening and risk management and is a recommended component of standard clinical care. In practice this remains unachievable in many families reflecting the genetic complexity of these disorders. We examined the potential of a whole genome sequencing (WGS) approach to genetic diagnosis and compared this to standard practice in the setting of a multidisciplinary cardiac-genetics clinic (CGC).**Methods**: Living index cases assessed in the CGC for inherited arrhythmia, cardiomyopathy or aortopathy (n=100) were consented for WGS alongside usual care. The whole genome was sequenced and annotated (Complete Genomics). Variants were filtered against public databases and whole genome data from a control cohort of healthy elderly. Clinical interpretation was applied to high quality calls for SNPs, InDels, structural variants and CNVs involving all genes associated with a cardiac phenotype (HPO database). Clinically significant results were validated by Sanger sequencing.**Results**: Analysis of 40 genomes has been completed (18 Cardiomyopathy, 19 LQTS/Brugada, 3 Aortopathy) revealing an average of 124 protein altering variants per genome in genes with a reported cardiac phenotype. For 11 individuals (27%) a potentially causative mutation was detected through targeted sequencing in standard care. This increased to 27 individuals (63%) following analysis of WGS. In 17% of cases more than one likely pathogenic change was detected. The addition of data on structural variation was an important contributor to the improved detection rate of WGS.**Conclusion**: Whole genome sequencing achieved a significant improvement in genetic diagnosis over standard investigation and gene testing. The results confirm the broad heterogeneity of inherited cardiac disorders and suggest a greater role of multiple mutations as determinants of severity across the range of inherited cardiac conditions.

547W

Molecular diagnosis of generalized arterial calcification of infancy (GACI). *K. Iravathy Goud[1], M. Kavitha[1], M. Adi Maha Lakshmi[1], R. Vempati[1], D. Sagarika[1], A. Panda[1], Y. Nitschke[2], F. Rutsch[2].* 1) Molecular Biology & Cytogenetics Laboratory, Apollo Hospitals, Hyderabad, India, India; 2) Department of generel pediatrics, Muenister University childrens Hospital, Munster, Germany.

Generalized arterial calcification of infancy (GACI) is a life-threatening disorder in young infants. Cardiovascular symptoms are usually apparent within the first month of life. The symptoms are caused by calcification of large and medium-sized arteries, including the aorta, coronary arteries, and renal arteries. Most of the patients die by 6 months of age because of heart failure. Recently, homozygous or compound heterozygous mutations for the ectonucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1) gene were reported as causative for the disorder. ENPP1 regulates extracellular inorganic pyrophosphate (PP(i)), a major inhibitor of extracellular matrix calcification. A newborn was diagnosed with GACI. The infant died at the age of 7 weeks of cardiac failure and the parents were referred to Molecular Biology and Cytogenetic lab for further workup. Cytogenetics analysis was performed on the parents, which showed normal karyotypes and mutational analysis for the ectonucleotide pyrophosphatase/phosphodiesterase 1 (ENPP1) gene was also performed. The mutational analysis showed that both father and mother of the deceased infant were heterozygous carriers of the mutation c.749C>T (p.P250L) in exon 7 of ENPP1 and it was likely, that the deceased child carried the same mutation homozygous on both alleles and died of GACI resulting from this ENPP1 mutation. The couple was counseled and monitored for the second pregnancy. Amniocentesis was performed at 15 weeks of gestation for mutational analysis of the same gene in the second pregnancy. The analysis was negative for the parental mutations. One month after the birth of a healthy infant, peripheral blood was collected from the baby and sent for reconfirmation. The results again were negative for the mutation and the baby was on 6 months follow up and no major symptoms were seen. The parents of the child benefited enormously by learning about the disease much in advance and also its risk of recurrence. The main aim of this study is to emphasize on two aspects: (i) the importance of modern molecular techniques in diagnosis such a syndrome and (2) the difficulties faced by the physician to provide appropriate diagnosis and the adequate genetic counseling to the family without molecular facilities. This study emphasis on importance of having molecular tools in tertiary care hospitals for diagnosis of rare genetic diseases.

548T

APOE genotype influences acute phase lipid changes in response to intracerebral hemorrhage. *C.D. Anderson[1-4], C.L. Phuah[1,4], M. Raffeld[1], A.M. Ayres[2,3], A. Biffi[1,3,4], A. Viswanathan[2,3], S.M. Greenberg[2,3], J. Rosand[1-4].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) The J. Philip Kistler Stroke Research Center, Massachusetts General Hospital, Boston, MA; 3) Hemorrhagic Stroke Research Group, Massachusetts General Hospital, Boston, MA; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA.

Background: In addition to their roles in modulating circulating lipid levels, APOE e2 and e4 alleles are genome-wide significant risk factors for intracerebral hemorrhage (ICH). ICH is accompanied by a decline in total cholesterol (TC) and low-density lipoprotein (LDL). Given the pleiotropic relationship between APOE epsilon alleles, ICH, and lipid levels, we investigated the influence of APOE status on changes in circulating lipids in the setting of ICH.*Methods:* ICH cases were drawn from an ongoing longitudinal cohort study. APOE epsilon variants were previously genotyped. Controls were drawn from a hospital clinical registry of non-cerebral acute illness. Measurements of serum lipids were standardized to four 6-month intervals before and after hospitalization for TC, LDL, triglycerides (TG) and high-density lipoprotein (HDL). Piecewise linear mixed-effects models were used to compare the impact of ICH on lipid trends before and after hospitalization. Linear regression analyses were utilised for comparing APOE allele-specific effects on temporal variation in serum lipid levels.*Results:* 212 ICH cases and 301 controls were analysed. Rates of decline in serum TC and LDL were greater at time of ICH compared to patients with non-cerebral acute illness (TC: -0.75 mg/dL/month for ICH, -0.13mg/dL/month for non-ICH subjects, p=0.0015; LDL: -0.31mg/dL/month for ICH, -0.13mg/dL/month for non-ICH subjects, p=0.31). This ICH-related decline in TC and LDL levels was associated with APOE e4 carrier status and not e2 or e3. APOE e4 carriers experienced a mean decrease of serum TC and LDL levels acutely in ICH of -20.8±33.0 mg/dL and -15.4±26.5 mg/dL, respectively compared with pre-ICH levels. In contrast, the difference in mean serum TC and LDL levels during the same time period were -3.3±36.2 mg/dL and -0.6±28.2 mg/dL, respectively in APOE e2 carriers. This decrease in TC and LDL at ICH in APOE e4 carriers remained significant ($p<0.05$ for TC, $p< 0.01$ for LDL) following adjustment for potential confounders including diabetes and statin use.*Conclusion:* ICH acutely impacts serum TC and LDL levels in ways that other non-cerebral acute illnesses do not, and APOE epsilon allele status is a strong predictor of these acute changes. APOE influence on temporal lipid trends in ICH may reflect APOE-specific differences in innate inflammatory or lipid-homeostatic mechanisms in reponse to acute cerebral injury.

549F

Knockout mice for the cardiac repolarization regulator *Nos1ap* display partial lethality and increased heart mass. *D.R. Auer, A. Kapoor, A. Chakravarti.* McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

We have previously identified the *NOS1AP* gene (MIM 605551) as a major gene for cardiac repolarization (QT interval: MIM 610141), a genetic risk factor for sudden cardiac death (MIM 115080), and a genetic modifier for Mendelian long QT syndrome (MIM 192500). Although *in vitro* gene targeting-based cellular and biochemical studies have shed light on *NOS1AP* function in the heart and other tissues, to enhance our understanding of its role in mammalian physiology and disease, we report the generation of *Nos1ap* knockout mice using traditional embryonic stem cell (ESC) targeting. ESCs carrying the knockout-first allele (with conditional potential; *tm1a*) were obtained from the Knockout Mouse Project. After germline transmission, *tm1a*/+ mice were bred with ubiquitously-expressing *cre* and *FLPe* recombinase lines to generate the *lacZ*-tagged, *neo*R- and exon4-deleted (*tm1b*) allele; the *lacZ*- and *neo*R-deleted, exon4-floxed (*fl*) allele; and the *lacZ*-, *neo*R-, and exon4-deleted (*exon4Δ*) allele. Among the mice obtained from heterozygous intercrosses (n=274), mice homozygous for the *tm1a*, *tm1b,* and *exon4Δ*alleles displayed reduced survival with two-thirds dying before reaching weaning age, whereas the expected numbers of homozygotes were obtained for the *fl* allele. At embryonic day 13.5 (E13.5), the expected numbers of *exon4Δ*homozygotes were observed out of 100 embryos screened, indicating that homozygous lethality occurs between E13.5 and weaning. Mice homozygous for the *tm1a*, *tm1b*, and *exon4Δ*alleles had (a) no detectable Nos1ap protein by western blotting, (b) a significant reduction in mRNA expression (0.4-fold of control for *tm1a*; P=1.1×10-5, 0.04-fold of control for *tm1b*, P=2.5×10-6; n=5 in each group), and (c) significantly increased heart weight as compared to controls (P=0.001; n=11 homozygous mutants, n=16 wildtype). Targeted RNA expression analyses in E13.5 hearts for genes that encode proteins known to interact with *Nos1ap* did not exhibit significant differences between homozygous mutants and wildtype controls. *Nos1ap* knockout mice are currently being characterized by cardiac histology, electrocardiography, echocardiography, and RNA-Seq to identify the physiological and global molecular consequences of disrupting *Nos1ap* dosage and to determine how the null phenotype is rescued in some animals.

550W

Response-eQTLs of human left ventricular tissue after ischemia. *T. Chang[1,2], V. Carey[2], S. Aranki[3], S. Body[1], C. Seidman[4], J. Seidman[5], B. Stranger[6], B. Raby[2], J. Muehlschlegel[1].* 1) Brigham and Women's Hospital, Department of Anesthesiology, Perioperative and Pain Medicine, Harvard Medical School, Boston, MA; 2) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 3) Brigham and Women's Hospital, Division of Cardiac Surgery, Harvard Medical School, Boston, MA; 4) Division of Cardiovascular Medicine, Brigham and Women's Hospital, Boston, MA; 5) Department of Genetics, Harvard Medical School, Boston, MA; 6) Department of Medicine, University of Chicago, Chicago, IL.

BBackground: Inter-individual variations in gene transcript levels are important in mediating disease susceptibility. To identify and functionally characterize genetic variation contributing to myocardial injury in humans, we performed an expression quantitative trait loci (eQTL) analysis in normal and ischemic human left ventricular myocardium to identify variants that alter expression in response to a stimulus.**Methods:** Apical punch biopsies from the left ventricle were taken from 135 patients undergoing cardiac surgery at two time points; immediately after aortic cross clamping (baseline), and immediately before aortic cross clamp removal (ischemia). We quantified genome-wide mRNA expression using an Illumina HiSeq and performed genotyping with the Illumina Omni2.5 array.The response-eQTL was defined as the delta between the ischemic and baseline sample. We examined the associations in *cis* between variants and gene expression levels with a distance of ±100kb, adjusting for demographics and clinical variables. A False Discovery Rate < 0.01 was applied to account for the large number of testing between variants and expressed genes, and only those SNPs with a MAF>5% were included. **Results:** In all, 1.5 million variants and 16000 genes were analyzed with a total of 4.3 million variant-gene pairs tested. Of those, 239 genes were shared between the baseline and ischemic state, 247 genes were unique to baseline, 84 were unique to ischemia. In addition, we identified 8 response-eQTLs.**Discussion:** We demonstrated the presence of genetic variants contributing to inter-individual variation for gene expression change in left ventricular myocardium, hereby providing evidence of a novel genetic basis for variable transcription in response to an ischemic insult in human tissue. Response-eQTL genes of particular interest were laminin alpha (*LAMA3, FDR=0, Chisq=16.8*), succinate dehydrogenase complex (SDHA, FDR=0.06, Chisq=16.3), and dihydrofolate reductase (DHFR, FDR=0, Chisq=19.9), linked to cell migration, mitochondrial respiratory function, or amino acid synthesis.

551T

Loss of Function Mutations in *NNT* Are Associated with Left Ventricular Noncompaction. *E.E. Davis[1], M.N. Bainbridge[2,3], W.Y. Choi[4], A. Dickson[4], H.R. Martinez[5], M. Wang[2], H. Dinh[2], D. Muzny[2], R. Pignatelli[5], N. Katsanis[1], E. Boerwinkle[2], R. Gibbs[2], J.L. Jefferies[6].* 1) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Codified Genomics, LLC, Houston, TX; 4) Department of Cell Biology, Duke University, Durham, NC; 5) Department Pediatrics-Cardiology, Baylor College of Medicine, Houston, TX; 6) The Heart Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

Left ventricular noncompaction (LVNC) is an autosomal dominant, genetically heterogeneous cardiomyopathy with variable severity, which may co-occur with cardiac hypertrophy. Despite recent advances towards understanding the molecular basis of the disorder, many of the causative LVNC genes remain to be identified. Here, we generated whole exome sequence (WES) data from multiple members from five families with LVNC. In four out of five families, the candidate causative mutation segregates with disease in known LVNC genes *MYH7* and *TPM1*, encoding myosin heavy chain beta and tropomyosin 1, respectively. Subsequent sequencing of *MYH7* in a larger LVNC cohort identified seven novel likely disease-causing variants. In the fifth family, we identified a frameshift mutation in *NNT*, a nuclear encoded mitochondrial protein, not implicated previously in human cardiomyopathies. Resequencing of *NNT* in additional LVNC families identified a second likely pathogenic missense allele. Suppression of *nnt* in zebrafish caused early ventricular malformation and contractility defects, likely driven by altered cardiomyocyte proliferation. *In vivo* complementation studies showed that mutant human *NNT* failed to rescue *nnt* morpholino-induced heart dysfunction, indicating a probable haploinsufficiency mechanism. Together, our data expand the genetic spectrum of LVNC and demonstrate how the intersection of WES with *in vivo* functional studies can accelerate the identification of genes that drive human genetic disorders.

552F

Genomic Subsets and Cell Types Contribute to the Polygenicity and Heritability of Coronary Artery Disease. *R. Do[1,6], H.-H. Won[2,3,4,5], P. Natarajan[2,3,4,5], A. Dobbyn[1,6], K. Lage[4,5,7,8], S. Raychaudhuri[5,9,10], E. Stahl[1,5,6,11].* 1) Genetics and Genomic Sciences, Icahn School of Medicine, New York, NY; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA; 3) Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA; 4) Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA; 5) Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA; 6) Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 7) Department of Surgery, Massachusetts General Hospital, Boston, Massachusetts, USA; 8) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts, USA; 9) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; 10) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK; 11) Division of Psychiatric Genomics, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

Coronary artery disease (CAD) and myocardial infarction (MI) is one of the leading causes of death and disability worldwide. Genome-wide association studies (GWAS) have identified up to 45 loci associated with CAD/MI. However, the molecular consequences of these loci remains largely unknown. Efforts by the National Institutes of Health (NIH) Roadmap Project, have allowed for systematic assessment of the molecular functions of the genome. By leveraging data from the NIH RoadMap Project, we investigated links between single nucleotide polymorphisms (SNPs) associated with MI/CAD and their biological impacts on gene function. First, we stratified MI/CAD according to genomic compartments and observed that SNPs that were proximal to the protein-coding region of genes exhibited significant polygenicity and >59% of the heritability for MI/CAD. Next, we showed that the polygenicity and heritability of MI/CAD are enriched in histone modification regions in specific cell types. By focusing on 45 MI/CAD-associated SNPs identified from GWAS, we showed that a higher number of these loci are enriched within specific regulatory elements, including active enhancer and promoter regions. Finally, we observed significant heterogeneity of this enrichment across cell types, with particularly strong signal observed in adipose nuclei, brain substantia nigra, brain angular gyrus, spleen cell types and smooth muscle tissues. These results suggest that the genetic etiology of MI/CAD is largely explained by tissue-specific regulatory perturbation within the human genome.

553W

Gene silencing and haploinsufficiency of *Jag1* in GWAS locus 20p12.2 increase blood pressure. *J. Kang[1], M. Kim[1], S. Ji[1], S. Kim[1,2], J. Lim[1], S. Hwang[3], B. Oh[1].* 1) Department of Biochemistry and Molecular Biology, School of Medicine, Kyung Hee University, Seoul, South Korea; 2) Department of Animal Biotechnology, The Graduate School of Future Convergence Technology, HanKyong National University, 327 Jungang-Ro, Ansung-Si, Gyeonggi-Do, South Korea; 3) Department of Chemical Engineering, College of Engineering, HanKyong National University, 327 Jungang-Ro, Ansung-Si, Gyeonggi-Do, South Korea.

Hypertension is a highly heritable trait, and around 40–60% of the inter-individual variation in hypertension is attributable to genetic factors. Genome-wide association studies (GWASs) have identified multiple genetic loci associated with blood pressure. The 20p12.2 locus was associated with hypertension, left ventricular wall thickness, stroke, and coronary artery disease in Europeans, East Asians, South Asians, and African ancestry individuals.We aimed to identify a causative gene in the 20p12.2 locus using mice models. Among genes on the locus, *JAG1* was selected as a candidate gene, based on the proximity to the lead SNPs (rs1327235, rs1887320), the hypertensive phenotype specified in Alagille syndrome individuals, and the role of Jagged-Notch signals in cardiovascular development. *In vivo* injection of *Jag1* siRNA in mice significantly increased blood pressure compared to control siRNA injection. Consistently, vascular endothelial specific, haploinsufficient *Jag1* mice (*Jag1*flox/+; Tie2-Cre) showed increase in blood pressure compared to wild-type mice (*Jag1*+/+; Tie2-Cre). We also found that the heart rate was increased in *Jag1*flox/+; Tie2-Cre mice compared to the wild-type. As presumed, injection of propranolol, a β-receptor blocker, reduced both elevated blood pressure and heart rate in mice, suggesting that the sympathetic surge increases the heart rate, resulting in elevation of blood pressure in *Jag1*flox/+; Tie2-Cre mice.Our study demonstrates that *JAG1* is a causative gene in the 20p12.2 locus and regulates blood pressure through sympathetic activation, and these findings may provide a novel hypertensive mechanism.

554T

The role of the anti-oxidant enzymes SOD2 and catalase in the pathogenesis of pulmonary arterial hypertension. *R.D. Machado[1], J. DeGeer[2], N. Lamarche-Vane[2].* 1) School of Life Sciences, University of Lincoln, Lincoln, United Kingdom; 2) Centre for Translational Biology, Glen site of the MUHC, 1001 Boul. Décarie, Montreal, QC.

Pulmonary arterial hypertension (PAH) is characterized by the progressive obstruction of pulmonary arterioles, amid metabolic abnormalities, leading to right-heart hypertrophy and failure. Patients with PAH display conspicuous metabolic imbalances in the pulmonary vasculature, in particular between vasoconstrictors and vasodilators associated with abnormally elevated levels of reactive oxygen species (ROS). Mutation of *BMPR2*, a receptor member of the TGF-beta signaling superfamily, is the primary genetic factor underlying disease. While this discovery broadly identifies the BMP pathway as being critical in maintaining the vascular bed, little is known of pathways downstream of BMPR-II relevant to vascular integrity.To interrogate the extent and nature of cellular dysfunction in PAH we conducted a yeast two-hybrid screen using the kinase domain of BMPR-II as bait in a lung cDNA library. This approach facilitates the elucidation of the wider BMP signaling network by the identification of novel interacting partners. Superoxide dismutase 2 (SOD2) and catalase (CAT) were independently isolated. Both enzymes are considered to be the front-line cellular defense against pathologically elevated levels of ROS and thereby of biological relevance to PAH pathogenesis. We demonstrate an association between BMPR-II and these molecules in yeast by GST pulldown and co-immunoprecipitation, in addition to identifying limited but consistent cellular co-localization in a proportion of co-transfected cells. Of considerable interest, by semi-quantitative Western blotting and immunofluorescence against endogenous protein we demonstrate that *BMPR2* haploinsufficient PAH patient cells display significant down-regulation and mis-localization of both enzymes, suggestive of a downstream effect of BMPR-II deficiency, a feature not observed in cancer cell lines used here as controls.Taken together, these data suggest that mis-handling of these enzymes by mutant BMPR-II receptors represents a significant and early defect in disease etiology. Current work now aims to determine whether restoration of SOD2/CAT expression by use of chemical mimetics has potential to reverse the cellular disease phenotype.

555F

Mechanical Stability of the Aorta in a *Col3a1* Mouse Model. *J. Meienberg[1], J. Muenger[1], J. Crabb[2], A. Mauri[2], C. Kaiser[3], G. Barmettler[3], C. Giunta[4], U. Ziegler[3], S. Zeisberger[5], E. Mazza[2], G. Matyas[1,6].* 1) Center for Cardiovascular Genetics and Gene Diagnostics, Foundation for People with Rare Diseases, Schlieren-Zurich, Switzerland; 2) Institute of Mechanical Systems, Swiss Federal Institute of Technology Zurich, Zurich, Switzerland; 3) Center for Microscopy and Image Analysis, University of Zurich, Zurich, Switzerland; 4) Division of Metabolism, University Children's Hospital, Zurich, Switzerland; 5) Swiss Centre for Regenerative Medicine, University of Zurich, Zurich, Switzerland; 6) Zurich Center for Integrative Human Physiology, University of Zurich, Zurich, Switzerland.

Ehlers-Danlos syndrome vascular type (EDS IV) is a rare autosomal dominant connective tissue disorder characterized by translucent skin, easy bruising, and arterial, intestinal and/or uterine fragility. The most severe complication is the increased risk for rupture of the aorta, leading to life-threatening internal bleeding. EDS IV is caused by mutations in the *COL3A1* gene, which encodes the alpha 1 chain of type III collagen, a fibrillar collagen expressed in walls of hollow organs. So far, only disease management and treatment of symptoms are available but no targeted therapy. Recently, a novel mouse model which has true haploinsufficiency of *Col3a1* due to a spontaneous deletion has been described. In heterozygous mice, this *Col3a1* deletion leads to reduced mechanical stability of the aorta and, in ~28% of cases, to spontaneous rupture of the aorta and thus to increased mortality similar to the phenotype of human patients. Our goal was to assess the mechanical stability of the aorta in this mouse model compared to wild-type mice.The thoracic aorta from wild-type and heterozygous *Col3a1* mice was dissected and cleaned of adherent connective tissue. 1.5-mm-long sections of aortic arch as well as ascending and descending aorta (Figure 1) were mounted on two 200-µm diameter stainless steel wires on a tissue puller (Danish Myo Technology). The mounted vessels were stretched radially until tissue damage, thereby recording the stretching force (in mM).Maximum force at tissue damage was significantly lower in heterozygous *Col3a1* mice compared to age- and gender-matched wild-type animals in both the ascending and descending parts of the aorta. For both genotypes, the mechanical stability of the aorta was decreasing with increasing distance from the heart.We developed a protocol for the assessment of the mechanical stability of mice aorta, which is suitable to detect significant differences between heterozygous and wild-type *Col3a1* mice. Our results open the way to test pharmacological substances for their potential to increase the mechanical stability of the aorta with the goal to find a therapy for patients with EDS IV and related aortic disorders.

556W

Cardiomyopathy in sialic acid deficient GNE myopathy. *L. Mian[1], N. Carrillo-Carrasco[2], S. Noguchi[3], M. Mori-Yoshimura[3], D. Maglic[1], V.M. Diaz[4], D. Despres[4], P. Zerfas[5], S.M. Wincovitch[6], S. Anderson[7], I. Nishino[3], W.A. Gahl[1, 8], M. Huizing[1], M.C. Malicdan[1, 8].* 1) Medical Genetics Branch, NHGRI, National Institutes of Health, Bethesda, MD; 2) Therapeutics for Rare and Neglected Diseases, National Center for Advancing Translational Sciences, NIH, MD, USA; 3) Department of Neuromuscular Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan; 4) National Institutes of Health, National Institute of Neurological Disorders and Stroke, NIH, USA; 5) Office of Research Services, Division of Veterinary Resources, Office of Research Services, NIH, MD, USA; 6) Cytogenetics and Microscopy Core, National Human Genome Research Institute, National Institutes of Health (NIH), MD, USA; 7) Cardiovascular and Pulmonary Branch, National Heart Lung and Blood Institute, National Institutes of Health (NIH), MD, USA; 8) NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, 20892 Maryland, USA.

Statement of Purpose: GNE myopathy (GM) is an adult-onset distal myopathy caused by mutations in the *GNE* gene which encodes a bifunctional, rate-limiting enzyme in sialic acid biosynthesis. GM is characterized by a progressive muscle weakness and atrophy, preferentially involving distal extremities. Muscle histology shows inclusion bodies and rimmed vacuoles in affected muscle fibers. Although GM is known to solely affect skeletal muscles, there are few reports of subjects with cardiac abnormalities. Through our Natural History Study (NCT01417533), we recognized that cardiac involvement may have a higher than expected occurrence in subjects with GM thus in this study we aimed to systematically characterize the cardiac phenotype in our GM mouse model. **Methods:** Patients and mouse models (Gne-/-hGNED176VTg) with GNE myopathy were subjected to cardiac physiology assessment, including electrocardiogram (ECG), Holter, and echocardiogram. In addition, GM mouse models and control littermates underwent cardiac MRI. Heart samples from these mice were collected at various ages for routine histology, immunohistochemistry, lectin histochemistry, electron microscopy, and protein analysis.**Results:** A number of patients in our cohort exhibited ECG, echocardiogram and Holter abnormalities. In mice, we found that a significant number of mice developed cardiac abnormalities progressively with age. GM mice had increased QRS interval, QT interval, and P amplitude in ECG. Echocardiogram findings show that GM mutant mice showed increased diastolic left ventricle volumes, suggesting of possible decreased cardiac functions, as compared with control. MRI findings reveal that GM mutant mice have decreased ejection fraction, decreased stroke volume, and decreased cardiac output. Histopathological analysis in GM mice revealed rimmed vacuoles in cardiomyocytes, fibrosis, decreased number of intercalated discs, and lower amount of cadherin as compared to control. **Conclusions:** Our findings provide evidence that cardiac muscles are involved in GM. It is important for clinicians to be aware of the occurrence of cardiac disease in GM for careful examination of the cardiac function and proper management. We propose that the cardiac abnormalities that we have seen are, at least in mice, compatible with restrictive cardiomyopathy. Future work will be needed to understand how aberrant sialylation in cardiomyocyes will lead to cardiomyopathy. .

557T

Blood-Pressure Associated Variants in Natriuretic Peptide Receptor C Affect Human Vascular Smooth Muscle Cells Proliferation and Angiotensin II-Stimulated Calcium Response. *M. Ren[1], F. Ng[1], K. Witkowska[1], M. Baron[2], A. Townsend-Nicholson[2], Q. Xiao[1], A. Hobbs[1], S. Ye[1], M. Caufield[1].* 1) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK; 2) Structural & Molecular Biology, University College London, London, UK.

*Background***:** Advances in genetic studies of blood pressure (BP) and hypertension have identified a number of trait-associated variants so far, and a recent genome-wide association study revealed a significant association between variation at natriuretic peptide receptor C (*NPR3)* gene locus and blood pressure (BP).*Objective*: To functionally investigate the role of BP-associated variants at NPR3 gene locus in the context of BP regulatory pathways.*Methods and Results:* Primary human umbilical smooth muscle (HUASMCs) and endothelial (HUVECs) cells were collected and genotyped for NPR3 gene variants rs1173743, rs1173747, rs1173756 and rs1173771. Endogenous mRNA and protein expression levels were assessed by qRT-PCR and western blotting. Open chromatin regions were examined using formaldehyde-assisted isolation of regulatory elements (FAIRE). We found the NPR3 risk allele (major allele) of intronic SNP rs1173747 was associated with diminished endogenous mRNA and protein levels in HUASMCs, which was consistent with its minor allele being located within an open chromatin state. Cell proliferation and migration were detected by cell counting and scratch assays. Intracellular calcium response induced by Angiotensin II (AngII) was detected using a fluorescent probe (Calcium 6). It is demonstrated cells carrying the major-allele of rs1173747 had increased proliferation and intracellular calcium level in response to AngII in HUASMCs. No difference in migration rates were detected. No such genotype-dependent characteristics were observed in HUVECs.*Conclusion:* This study has provided a mechanistic explanation for susceptibility variants at NPR3 gene locus to impact BP predominantly through modulating vascular smooth muscle cell behaviours.

558F

Apolipoprotein L1 genetic variants are associated with incident chronic kidney disease but not incident cardiovascular events in a population referred for cardiac catheterization. *H. Wang[1], L. Kwee[1], D. Craig[1], C. Haynes[1], M. Chryst-Ladd[1], L. Svetkey[1], U. Patel[1], P. Pun[1], E. Hauser[1], W. Kraus[1], M. Pollak[2], S. Shah[1].* 1) Duke University, Durham, NC; 2) Beth Israel Deaconess Medical Center, Boston, MA.

Genetic variants in the apoliprotein L1 (*APOL1*) gene have been implicated in the development of chronic kidney disease (CKD) in African Americans. However, the association between *APOL1* variants and cardiovascular disease is unclear. We assessed the association of *APOL1* variants with incident CKD, coronary artery disease (CAD) and incident cardiovascular events (CVD) in a large population referred for cardiac catheterization. Two single nucleotide polymorphisms in *APOL1* (G1 allele, rs60919145 and rs73885319) and a 6 base-pair insertion-deletion (G2 allele, rs71785313) were genotyped in African Americans (n = 1760) enrolled in the CATHGEN biorepository consisting of patients referred for cardiac catheterization at Duke University Medical Center (2001-2010). Incident CKD was defined as eGFR > 60 mL/min/1.73 m2 at index catheterization with at least two subsequent eGFR < 60 separated by at least 3 months; incident CVD included myocardial infarction and need for coronary interventions. Individuals were categorized as non-carriers (Ref/Ref), heterozygote (G1/Ref or G2/Ref) or homozygote carriers (G1/G1, G1/G2 or G2/G2) of *APOL1* risk alleles. The minor allele frequency was 0.22 (rs60919145), 0.14 (rs71785313) and 0.22 (rs73885319). Homozygote carriers had a lower median eGFR than heterozygotes and non-carriers (74.8 vs. 78.7 mL/min/1.73 m2, p = 0.004) and higher prevalence of baseline dialysis (12.1% vs. 3.4%, p < 0.001). Previously identified association between *APOL1* variants and incident CKD was confirmed (OR 2.0, 95% CI 1.0 – 4.1, p = 0.05). With regards to CAD, there was a trend towards less prevalent CAD and significantly lower burden of CAD in carriers vs. non-carriers (45.8% vs. 50.3%, p = 0.08; median CAD index 19 vs. 23, p = 0.02). However, no statistically significant associations were detected between *APOL*1 genetic variants and CAD in multivariable regression using additive, dominant and recessive models, adjusted for age, sex, BMI, diabetes status, hypertension, smoking status, hyperlipidemia and baseline eGFR. There were no significant differences between *APOL1* genetic variants and time-to-event for CVD events, even when stratifying by baseline kidney function and diabetes. Our results suggest that while *APOL1* variants are significantly associated with prevalent and incident kidney disease in African Americans, they are not associated with CAD or incident CVD events in a cohort of individuals with a high burden of cardiometabolic risk factors.

559W

Regulation of postprandial lipid homeostasis by *TM6SF2* . *N. Zaghloul, R. Yang, E. OHare, K. Ryan, S. Taylor, A. Shuldiner, L. Yerges-Armstrong, C. Sztalryd-Woodle.* Genetics & Genomic Med, Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD.

Postprandial hypertriglyceridemia is an important risk factor for cardiovascular diseases (CVD), but the mechanisms by which genetic factors modulate it remain unclear. It is primarily driven by very low density lipoprotein (VLDL) secretion from liver or chylomicron (CM) production from small intestine in response to dietary fat. We assessed the effect of a loss-of-function missense variant, E167K in the *TM6SF2* gene which was recently found to be associated with a favorable fasting blood lipid profile as well as other lipid traits such as hepatic steatosis. *TM6SF2* is predominately expressed in liver and small intestine consistent with a potential role in both tissues in regulation of postprandial lipemia. We genotyped the variant in 2059 subjects from the Old Order Amish (OOA) and performed association analyses to determine the effects of TM6SF2-deficiency on postprandial hypertriglyceridemia. Relative to the 7% minor allele frequency (MAF) in European-American populations, the E167K is enriched to 12% in the OOA. Analysis of carriers recapitulated associations with significantly lower fasting triglycerides (TG), total cholesterol (b=-10.4, p=1.2x10-5), and LDL-C (b=-10.8, p=6.2x10-7) and higher HDL-C (b=2.2, p=3.6x10-3). Importantly, following a high fat challenge, heterozygous and homozygous individuals exhibited significantly lower TG levels at all post-high-fat challenge time points ($p<0.05$ all), and lower total area (p=5x10-3) and incremental area under the curve (p=0.02). To gain further insight into the role of *TM6SF2* in small intestine lipid homeostasis, we developed a model of *tm6sf2*-deficiency in zebrafish. Consistent with the human clinical observations, we observed a significant decrease in fasting LDL-C and increased steatosis in liver. In addition, accumulation of lipid in enterocytes was also detected. Introduction of a 3-hour high fat diet (HFD) followed by clearance of dietary lipid for 24 hours produced significant differences in the intracellular lipid content of enterocytes consistent with a possible disruption of lipid clearance from enterocytes in response to dietary lipid. Finally, in human Caco-2 enterocytes treated with shRNA against *TM6SF2* exhibited significantly more lipid accumulation relative to control cells in response to an exogenous lipid load. Taken together, these data strongly support a role for *TM6SF2* in the regulation of post prandial lipemia and likely implicate its function in enterocyte lipid homeostasis.

560T

Role of Genetic Variants in *Cholesteryl Ester Transfer Protein* in Risk of Myocardial Infarction and Response to Therapy. *K. Hartmann[1], M. Seweryn[2], W. Sadee[1].* 1) Department of Pharmacology, The Ohio State University, Columbus, OH; 2) Faculty of Mathematics and Computer Science, University Lodz, Lodz, Poland.

Purpose Several pharmaceutical companies have designed inhibitors of cholesteryl ester transfer protein (CETP) as an adjunct or alternative to statins, which fail to prevent myocardial infarction (MI) in ~50% of cases despite lowering lipids. Initial results from clinical trials are conflicting, which may reflect the effect of genetic variants. SNPs in *CETP* have been consistently associated with HDL but less so with MI and treatment response. CETP is part of a gene family (PLTP, LBP, and BPI), which functions not only in cholesterol transport but also in immunity. A complex haplotype structure, containing frequent long haplotypes indicative of positive selection, may in part reflect activity in immunity. We propose that the lack of consistent association of CETP with MI may be due to failure to consider all relevant variants in combination, both in liver and immune related tissues.**Approach** We suggest two complementary hypotheses:1) relevant variants in *CETP* can be frequent, and potentially alter CETP expression in opposing directions; 2) the main contribution of *CETP* to MI is through lipid-immunity interactions. Our first aim was to identify SNPs associated with altered CETP mRNA expression using data from the Genotype and Tissue Expression Project (GTEx) and Framingham Heart Study (FHS), and analyze linkage patterns using 1000 genomes. We test predicted SNPs in FHS and Atherosclerosis Risk in Communities (ARIC). Our second aim was to assess the physiological role of CETP, by identifying co-expressed genes and functional networks.**Results** Our group has previously identified variants associated with decreased expression (rs247616) and increased splicing (rs5883). Analysis of expression data identified an additional SNP associated with increased expression on the haplotype opposite of rs247616, associated with decreased expression. This robust negative linkage pattern and a long, well-preserved haplotype including rs5883 suggest strong evolutionary pressure attributable to the lipid and/or potential immune functions. High expression of CETP in spleen and lymphocytes relative to liver supports a role in immunity. Co-expression analysis identified multiple genes, most with immune functions. Our initial analysis of candidate SNPs in GWAS yielded associations with both lipids and MI. Further work will test candidate SNPs in *CETP* and tightly co-expressed genes in combination to account for lipid levels and MI risk.Supported by U01GM092655 and TL1TR001069.

561F

Increased DNA damage and mutagen sensitivity: a new genetic modifier in Pulmonary Arterial Hypertension? *K.M. Drake, S.L. Meade, M.A. Aldred.* Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Cleveland, OH.

Pulmonary arterial hypertension (PAH) is a life-threatening condition characterized by obliterative lung vascular remodeling. It may be heritable (HPAH), idiopathic or associated with another underlying condition. HPAH shows autosomal dominant inheritance, but the penetrance averages only 27%, suggesting that other genetic and/or environmental modifiers are important in triggering disease. Our research focuses on the role of somatic mutations. We have shown that endothelial cells from PAH lungs frequently harbor chromosomal abnormalities. We therefore tested whether PAH patients are intrinsically more susceptible to DNA damage. Lung and blood cells showed significantly higher levels of micronuclei at baseline and in response to bleomycin or etoposide. Results were consistent across all three types of PAH. Surprisingly, we found similar results in unaffected first-degree relatives, demonstrating that this is not related to PAH treatments and suggesting a possible genetic basis. The goals of the present study are to identify the gene(s) underlying this increase in DNA damage, define the spectrum of mutagen sensitivity and determine if DNA repair is normal. We are initially focusing on PAH families, combining QTL extreme phenotype mapping with parametric linkage analysis and exome sequencing to identify candidate genes. The number of micronuclei, a marker of chromosomal breakage and loss, is analyzed in blood cells at baseline and in response to mutagens. Preliminary results in two families suggest dominant segregation independent of the underlying PAH mutation. Two unaffected relatives have baseline levels of micronuclei in the control range but an exaggerated response to mutagens in the “PH range”. The timecourse of repair of bleomycin-induced damage is largely normal, but not all patients return to baseline, suggesting a subtle difference in DNA repair that may lead to accumulation of damage over time. Early results with other potential mutagens show a highly significant increase in micronuclei in response to HDAC inhibitors. HDAC inhibitors have been proposed as a potential therapy for PAH, so delineating the differences in histone acetylation patterns is now an important priority. In conclusion, our results suggest that increased levels of DNA damage may be an independent genetic modifier of PAH and increased sensitivity to mutagenic agents has important implications for clinical management.

562W

SLC39A8 polymorphism influences cellular cadmium uptake and toxicity via signalling pathway activation in relation to hypertension. *R. Zhang[1], K. Witkowska[1], J. Guerra-Assuncao[2], F. Ng[1], M. Caulfield[1], S. Ye[1].* 1) Clinical Pharmacology , William Harvey Research Institute , London, United Kingdom; 2) Barts Cancer Institute, John Vane Science Centre, London, United Kingdom.

Objective. *hSLC39A8* (human solute carrier family 39 member 8) encodes a transmembrane protein that co-transpors divalent heavy metal cations, such as Cd2+,with elusive physiological role. Recent genome-wide association studies have identified a non-synonymous single nucleotide polymorphism rs13107325 to be associated with hypertension. The aim of this study is to investigate the functional impact of rs13107325 resulting in an amino acid substitution from Ala to Thr (A391T) in SLC39A8 on Cd2+ transport and the downstream signaling pathways.Methods. Intracellular Cd2+ uptake was measured in HEK293 cells overexpressing SLC39A8 (Measure-iTTM Pb and Cd assay kit), and in human umbilical vascular endothelial cells (HUVECs) of different genotypes. Cd2+- and genotype-dependence of ERK1/2 and NF-kB pathway activation were investigated by immunoblotting and dual-luciferase reporter assay. Cytotoxicity was measured by lactate dehydrogenase assay and MTS assay. Molecular dynamics simulations were performed to predict *in silico* the effect of A391T on the structure and dynamics of SLC39A8 by using Robetta, TMHMM and etc.Results. Overexpression of Ala variant in HEK293 resulted in higher Cd2+ uptake and higher cell death as compared with the Thr variant. This is associated with increased phosphorylation of ERK1 and NF-kB activation. Similar trends were observed in HUVECs with endogenous SLC39A8. Bioinformatics tools also suggested a conformational change of the a-helical structural transition (residual 390-392) in the Thr391 mutant, which potentially attenuates the protein function.Conclusion. Increased Cd2+ uptake by SLC39A8 Ala variant (blood pressure raising allele) is associated with higher cell death in human kidney and endothelial cells. Therefore its altered function due to rs13107325 may indicate a potentail therapeutic target in hypertension.

563T

Genetics of gene expression regulation in a case-control study for Acute Myocardial Infarction in a Pakistani population. *N.I Panousis[1,2,3], S. Tuna[4], L. Lataniotis[4], A. Rasheed[5], N. Shah[5], J. Danesh[6], E.T Dermitzakis[1,2,3], D. Saleheen[5,7], P. Deloukas[4].* 1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) Institute of Genetics and Genomics in Geneva (iG3), University of Geneva Medical School, Geneva, Switzerland; 3) Swiss Institute of Bioinformatics, Geneva, Switzerland; 4) William Harvey Research Institute, Queen Mary University of London, London, UK; 5) Center for Non-Communicable Diseases, Karachi Pakistan; 6) Department of Public Health and Primary Care, University of Cambridge, UK; 7) Department of Biostatistics and Epidemiology, University of Pennsylvania, USA.

The prevalence of coronary heart disease (CHD) is higher in populations of South Asia compared to Europeans; the risk of Myocardial infarction (MI) is 2-3 fold higher. By studying patterns of gene expression and genetic variation in monocytes, which is a relevant cell type we can discover biomarkers and develop a functional understanding of how genetic variation can alter risk of CHD. We report an RNA-seq analysis of monocytes from 71 cases of confirmed acute MI and 77 healthy individuals from the Pakistan Risk Of Myocardial infarction study (PROMIS). Principal component analysis (PCA) of expression differentiated MI from healthy individuals, demonstrating whole genome differences in expression. More specifically, we identified 5244 differentially expressed genes (5% FDR), including many that had previously been associated with cardiovascular disease (*IL1R2, PDGFD, LRP1*). PCA using only few differentially expressed genes cleanly differentiated MI and healthy individuals, suggesting a method for us to detect and/or validate novel biomarkers. We also explored genetic regulation of monocyte expression, in the context of GWAS variants associated with disease. As monocytes play a central role in CHD pathogenesis, we hypothesized that variants affecting gene expression are likely to be enriched in risk alleles for cardiovascular related diseases. We tested for genetic variants associated with gene expression levels within 1Mb of the transcription start site of a gene and discovered 4799 eQTLs as well as 179 variants associated with splicing changes (asQTLs) (FDR 5%). Analyzing cases and controls separately, we identified hundreds of MI specific eQTLs, which were not associated with expression in controls. Enrichment analysis of MI-specific eQTLs and GWAS loci, using the Regulatory Trait Concordance (RTC) method revealed several loci where the eQTL and GWAS signal were tagging the same functional variant. This included eQTL variants, which mediate GWAS hits for obesity (rs11247915), lipid levels (cholesterol: rs2277862, triglycerides: rs3198697) and type2 diabetes (rs3786897). We are currently combining the above findings with phenotypic risk factors (lipid profile, lifestyle, BMI) available in PROMIS to further investigate the molecular basis of CHD. In summary, we have shown that monocyte expression is an informative biomarker for MI and genetic variants active in the cell can be informative on genetic causes of CHD.

564F

A novel functional *VKORC1* promoter polymorphism (*rs397509427*) and Warfarin resistivity in Indian patients. *T. Shukla[1], K.P. Rao[1], S.K. Munpally[2], R. Tripathi[2].* 1) Genetics Dept, Osmania University, Hyderabad, India; 2) RAS Lifesciences, Hyderabad, India.

Warfarin is used as an oral anticoagulant for the prevention of thromboembolic diseases for subjects with deep vein thrombosis, atrial fibrillation, rheumatic heart disease or open heart surgeries. The dose requirement for warfarin therapy is highly variable, both inter-individually and inter-ethnically. Apart from the involvement of Genetic polymorphisms for the interindividual variability of humans to drugs, in the case of Warfarin, two important genes, viz., *VKORC1* and *CYP2C9* are known to produce interindividual variation in warfarin therapy. The present study was undertaken for their involvement in altered drug metabolism and action.*VKORC1* is the target enzyme for warfarin, which plays an essential role in carboxylation of vitamin K-dependent coagulation factors, by recycling Vitamin K. Mutations in the *VKORC1* gene and its haplotypes were known for variable dosages during warfarin therapy. Functional studies revealed that the dominant *VKORC1* -1639 G>A promoter SNP (*rs9923231*) correlates with lower protein expression levels. In the present study, A novel mutation *rs397509427* SNP (an insertion of G, which was present at -1586 position with respect to start codon) was observed in the promoter region of the *VKORC1* gene along with the *rs9923231* SNP in 8 patients, all of whom were on a higher dosage of warfarin (>7mg/day). The in-silico studies for Transcription factor binding site prediction, Secondary structure analysis and prediction of conserved short sequences/enhancers showed that the mutation could possibly act as an enhancer binding site or affecting the binding of transcription inhibitors.Further, studies were conducted for the functional study in HepG2 cell line to assess the role of mutant promoters. The relative quantitation of luciferase mRNA expression levels in control of *VKORC1* mutant promoter, showed approximately 6 fold higher expression levels in the cells, while the luciferase assays showed a 3 fold increase in protein expression. The results correlated well with the proposed hypothesis of increased transcription and translation due to *rs397509427* mutation. It may be concluded that the mutation might enhance the efficiency of the regeneration of the reduced vitamin K which ultimately would produce more gamma-carboxylation of the vitamin K dependent clotting factors.. It is an important finding and should play a significant role in the treatment regime as a diagnostic marker, when studied on a wider range of ethnic groups.

565W

The Gly364Ser variant in the Catestatin domain of Chromogranin A enhances the risk for hypertension in Indian populations. *M. Kiranmayi[1], V.R. Chirasani[1], P.K.R. Allu[1], S. Lakshmi[1], B.S. Sahu[1], D. Vishnuprabu[2], S. Sharma[3], R. Kumaragurubaran[1], D. Bodhini[4], M. Dixit[1], A.K. Munirajan[2], M. Khullar[3], V. Radha[4], V. Mohan[4], A.S. Mullasari[5], S. Senapati[1], N.R. Mahapatra[1].* 1) Department of Biotechnology, Indian Institute of Technology Madras, Chennai, India; 2) Department of Genetics, University of Madras, Chennai, India; 3) Department of Experimental Medicine and Biotechnology, Postgraduate Institute of Medical Education and Research, Chandigarh, India; 4) Department of Molecular Genetics, Madras Diabetic Research Foundation, Chennai, India; 5) Institute of Cardiovascular Diseases, Madras Medical Mission, Chennai, India.

Chromogranin A (CHGA) is a 48-kDa protein involved in the sorting/packaging of catecholamines and neuropeptides into the secretory granules of endocrine and neuro-endocrine tissues. CHGA is also a pro-hormone that gives rise to several bioactive peptides including catestatin (CST: human CHGA352-372), a catecholamine-release inhibitor. To identify and score naturally occurring polymorphisms in the CST region of CHGA in Indian populations, we re-sequenced and genotyped the region in a South Indian (Chennai) population (n≈2680). We discovered one novel (Gly367Val) and one previously reported polymorphism (Gly364Ser). We further replicated the genotyping of Gly364Ser in a geographically-distinct North Indian (Chandigarh) population (n≈760) using Taqman assay. Of note, there was a drastic difference in the frequency of 364Ser allele between the Chennai (6.28%) and Chandigarh (3.48%) populations. In both the populations, 364Ser was found to occur in a higher frequency in hypertensive cases than controls (unadjusted odds ratio =1.440 (p=0.015) for Chennai and =2.662 (p=0.002) for Chandigarh). In the Chennai population, the 364Ser carrying individuals displayed higher blood pressure levels: SBP by ~3 mm Hg (p=0.047), DBP by ~2 mm Hg (p=0.045) and MAP by ~2.5 mm Hg (p=0.030) and higher plasma glucose levels (by ~8 mg/dl; p=0.018). In the Chandigarh population as well, 364Ser variant was found to be associated with elevated blood pressure levels: SBP by ~8 mm Hg (p=0.004), DBP by ~6 mm Hg (p=0.001) and MAP by ~7 mm Hg (p=0.001). The CST-Ser peptide generated lower Nitric Oxide (NO) levels in HUVECs (assayed by DAF-2DA) as compared to CST-WT peptide, providing a plausible explanation to these observations. Molecular modelling and molecular dynamics simulation studies of the peptides showed that the presence of 364Ser caused dramatic changes in the CST peptide secondary structure, thus, revealing a structural basis for their differential biological activities. In light of the recent reports showing direct links between NO generation and beta2-adrenergic receptor (ADRB2), molecular dockings of CST peptides to ADRB2 were carried out. While CST-WT could effectively block the binding of the antagonist cyanopindolol to the ADRB2 active site CST-Ser bound at an entirely different location, keeping the active site free for interactions. In conclusion, the Gly364Ser variant increases the risk for hypertension via modulation of NO pathway in vascular endothelial cells.

566T

Functional Interrogation of an Intronic SNP in *SMAD3* Linked to CAD (Coronary Artery Disease). *A. Turner, A. Martinuk, M. Nikpay, P. Lau, S. Soubeyrand, R. McPherson.* University of Ottawa Heart Institute, Ottawa, Ontario, Canada.

As part of recent large meta analyses of genome-wide association studies (GWAS), we have identified several new loci associated with CAD, including *SMAD3*. SMAD3 is an important mediator of TGFb signaling, but its specific role in CAD pathogenesis is not currently clear. The novel CAD-associated SNP in *SMAD3*, rs56062135, has an effect allele frequency of 0.79 (odds ratio 1.07). Many SNPs identified via GWAS are not causal in disease pathogenesis themselves, and may instead act as proxies for the actual functional SNP(s). Our goal was to identify candidate functional CAD-associated SNP(s) at the *SMAD3* locus and characterize the roles these SNPs play in CAD pathogenesis. Bioinformatics revealed rs56062135 is perfectly linked with several other SNPs, with rs17293632 representing the best candidate to be a causal variant. rs17293632 is located in an intron in *SMAD3* with chromatin marks indicative of an enhancer and is predicted to disrupt a conserved, consensus AP-1 binding site. Luciferase assays in primary human aortic smooth muscle cells, HeLa cells, and HepG2 cells reveal rs17293632 is located in a very strong enhancer sequence. The minor allele disrupts the AP-1 site and drastically lowers enhancer activity ($p<0.001$). To determine if rs17293632 is associated with SMAD3 mRNA levels in humans *in vivo*, we selected age and gender matched patients across the three different genotypes (approximately 50 patients per genotype). RNA from whole blood was extracted from PAXgene Blood RNA tubes (PreAnalytiX/QIAGEN) and SMAD3 mRNA levels were quantified by qRT-PCR. Subsequent expression quantitative trait locus analysis revealed rs17293632 genotype is significantly associated with SMAD3 mRNA levels in whole blood ($p<0.05$). rs17293632 was also significantly associated with SMAD3 mRNA levels in 96 genotyped carotid plaque samples ($p<0.05$). Chromatin immunoprecipitation experiments in genotyped human aortic smooth muscle cells where the *SMAD3* AP-1 site is intact reveal binding of several AP-1 components to the region encompassing rs17293632. Allele-specific chromatin immunoprecipitation in aortic smooth muscle cells heterozygous at rs17293632 show AP-1 components bind less preferably to the minor allele. Our *SMAD3* data highlights that non-coding DNA variants are important in the pathogenesis of common diseases such as CAD. Ongoing functional analysis of the *SMAD3* locus is expected to identify novel mechanisms and pathways leading to atherosclerosis.

567F

Integrity of induced pluripotent stem cell (iPSC) derived megakaryocytes as assessed by genetic and transcriptomic analysis. *K. Kammers[1], J.T. Leek[1], I. Ruczinski[1], J. Martin[2], M.A. Taub[1], L.R. Yanek[2], A. Frazee[1], D. Hoyle[3], N. Faraday[2], D. Becker[2], L. Cheng[3], Z.Z. Wang[3], L. Becker[2], R.A. Mathias[2].* 1) Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 2) The GeneSTAR Program, Johns Hopkins School of Medicine, Baltimore, MD; 3) Johns Hopkins School of Medicine, Baltimore, MD.

Aggregation of platelets in the blood on ruptured or eroded atherosclerotic plaques may initiate arterial occlusions causing heart attacks, strokes, and limb ischemia. Understanding the biology of platelet aggregation is important to prevent inappropriate vascular thrombosis. GWAS studies have identified common variants associated with platelet aggregation, but because they are intronic or intergenic, it is not clear how they are linked biologically to platelet function. To examine this, we are funded to produce pluripotent stem cells (iPSCs) from people with informative genotypes, and then derive megakaryocytes (MKs), the precursor cells for anucleate platelets, from the iPSCs to determine patterns of gene transcript expression in the MKs related to specific genetic variants.To this end it is essential that the iPSC-derived MKs retain their genomic integrity during production or expansion. This was examined using three alternative measures of integrity of the MK cell lines: (1) mutation rates comparing parent cell DNA to iPSC cell DNA and onward to the differentiated MK DNA; (2) structural integrity using copy number variation (CNV) on the same; and (3) transcriptomic signatures of the derived MK cells. For the genotype and CNV data, we used the HumanOmniExpressExome-8v1 array on 14 paired donor DNA – iPSC and paired iPSC - MK lines, and for the RNASeq data we extracted non-ribosomal RNA from 14 paired iPS and MK cell lines.A comparison of genotypes between matched pairs of cell lines indicated a very low rate of discordance; estimates ranged from 0.0001%-0.01%, well below the genotyping error rate (0.37% estimated from controls). No CNVs were generated in the iPSCs that got passed on to the MKs. Finally, looking specifically for genes 'turned on' in MKs following differentiation from the iPSCs, we observed the following highly biologically relevant gene sets in the list of top 12 identified: platelet activation, immune response, inflammatory response, platelet formation, and regulation of cell proliferation. The most highly expressed genes in iPSCs were ribosomal proteins, while in MKs they were platelet related proteins, such as GP9 and PF4.In conclusion, the genetic and transcriptomic data strongly support high integrity of the iPSC-derived MKs. We are currently performing extensive eQTL analysis to categorize 'functional' relevance of the GWAS-identified determinants of platelet aggregation leveraging the genotype and RNASeq data.

568W

Elevated levels of oxidative DNA damage and Folate gene polymorphisms in children with congenital septal defects and their mothers. *s.B. Sunayana[1], k. Srujana[1], K. Nageswara rao[2], G. Sandhya[1], D. Sujatha[3], M. HemaPrasad[1].* 1) Human Genetics, Osmania University, Hyderabad, India; 2) Care Hospital, Banjara Hills, Hyderabad-500034; 3) National Institute of Nutrition, Hyderabad.

Congenital septal defects (CSD) are the most common type of birth defect and may result in death in uterus, in infancy, or in childhood. Maternal oxidative stress and folic acid levels are the major factors that adversely affect the fetal heart development. The present study is aimed to understand the possible etiological factors that contribute to an increase in the severity of disease. Cytogenetic analysis using fluorescence in situ hybridization (FISH) and comet assay were performed to identify the microdeletions in chromosome 22 and to measure the oxidative DNA damage in the lymphocytes, respectively. Single nucleotide polymorphisms (SNPs) of Methylenetetrahydrofolate reductase (MTHFR) (677C >T), reduced folate carrier (RFC) (A80G) and methionine synthase reductase (MTRR) were analyzed to identify the role of folic acid.Methods: One hundred and sixty two children with non-syndromic CSD and their mothers were studied with age-matched controls. Blood samples were collected from the Department of Pediatric Cardiology, Care Hospital, Hyderabad, with prior clearance from the Institutional Ethics Committee and written consent from the parents. The lymphocytes were isolated from fresh blood for comet assay as well as for the identification of 22 q11.2 microdeletions. The plasma and red blood cells (RBCs) were used for the estimation of folic acid and vitamin B12. The C-reactive protein (CRP) was analyzed using the enzyme-linked immunosorbent assay. Genomic DNA was isolated for PCR and restriction fragment length polymorphism analysis of SNPs 677C> T, 1298A>G, A80G and 66A>G genes.Results:The mean values of RBC folate in mothers with affected prepositus having 3 months infants to 15 years children with CSD showed one and a half fold reduction in RBC folate (349.6 ± 162.9ng/ml) when compared to controls (527.6±227.4ng/ml). Further, vitamin B12 levels in mothers (73.7 ± 46.2ng/mL) with CSD offspring were significantly low when compared to the controls (264.2 ± 159.1ng/mL). The mean comet tail length and CRP levels were significantly increased ($p<0.05$) in CSD children and mothers compared to the controls. The SNPs of MTHFR 677C> T, 1298A>G, RFC 80A> G and MTRR 66A>G was observed in 79.65% of CSD children and in 85.8% mothers, respectively. Conclusions: Low folate levels and genetic polymorphism have shown increased oxidative DNA damage in children with congenital septal defects and in their mothers are the important risk factors for CSD .

569T

A Gene-by-Environment Interaction Informs Aortic Segment-Specific Vulnerability for Aneurysm Formation in Mendelian Aortopathies. *N. Wilson, J. Doyle, C. Antonescu, S. Wheelan, D. Bedja, H. Dietz, MIB-AVA Leducq consortium.* Institute of Genetic Medicine, Johns Hopkins Medical Institutions, Baltimore, MD.

Most heritable aortopathies including Marfan (MFS) and Loeys-Dietz (LDS) syndrome show aortic root aneurysm (AoRA). Bicuspid aortic valve with more distal ascending aortic aneurysm (BAV/AscAA) is the most common inherited aneurysm condition, occurring in up to 1% of people, yet its etiology remains mysterious with heterozygous loss-of-function mutations in *NOTCH1* accounting for less than 1% of cases. It remains unclear whether anatomic differences between these conditions foreshadow intrinsic differences in mechanistic underpinnings and hence therapeutic vulnerabilities. We showed that TGFb-linked ERK activation drives aortic disease in a MFS mouse model (*FBN1*C1039G/+) and that inhibition of ERK phosphorylation with either angiotensin receptor blockers (ARBs) or the ERK kinase inhibitor RDEA119 rescues AoRA in MFS mice. In contrast, calcium channel blockers (CCBs) specifically induce more distal ascending aortic enlargement and tear in MFS mice that closely mimics the pathology typically seen in BAV/AscAA. Concomitant administration of either ARBs or RDEA119 prevents both AoRA and AscAA in CCB-treated MFS mice. In an unbiased approach to mechanistically characterize this gene-by-environment interaction, we applied strict a priori filters to RNA sequencing data from aortic specimens, imposing a change in CCB-treated animals that was abrogated upon use of RDEA-119. Given that expression-profiling analysis of end-stage tissues makes it difficult to distinguish between pathogenic, compensatory and irrelevant events, we focused on alterations that occurred in tight temporal sequence with conditional provocation. Enrichment algorithms identified Notch signaling as the lead candidate pathway. Informatively CCB-treated MFS mice exposed to dibenzazepine (DBZ), a potent Notch inhibitor, showed marked acceleration of AscAA progression and rupture; DBZ had no effect on wild-type animals but also accentuated AoRA in MFS mice without CCB treatment. We conclude that: 1) there is intersection between the pathogenesis of MFS and BAV/AscAA; 2) in the face of an underlying disease predisposition, Notch signaling is protective; and 3) multiple therapeutic strategies including Notch agonism have shared potential in phenotypically discordant aneurysm conditions. Studies are underway to interrogate the relevance of the 2nd-most enriched pathway in CCB-treated MFS mice, regulation of androgen receptor activity, to the overt male preponderance among patients with BAV/AscAA.

570F

***Notch1* haploinsufficiency under the influence of maternal hyperglycemic environment increases risk of congenital heart defects by an epigenetic mechanism.** *M. Basu[1], K. Bosse[1], V. Garg[1,2].* 1) Center for Cardiovascular and Pulmonary Research and The Heart Center, Nationwide Children's Hospital , 700 Children's Drive, Columbus, OH 43205; 2) Departments of Molecular Genetics & Pediatrics, The Ohio State University.

Rationale: Congenital heart disease (CHD) is the most common birth defect with high incidence of infant morbidity and mortality. Majorities of CHDs are proposed to have a multifactorial etiology and are due to a combination of genetic and environmental factors. Pre-gestational maternal diabetes is a well-recognized non-genetic risk factor which disturbs embryonic development and increases susceptibility towards congenital malformations. Hyperglycemia results in increased production of reactive oxygen (ROS) and nitrogen species leading to endothelial-cell dysfunction and we recently demonstrated a genetic interaction between endothelial nitric oxide (NO) synthase and *Notch1* important for cardiac development. Therefore, we hypothesized that maternal diabetes in combination with *Notch1* heterozygosity in developing embryo will predispose to the development of CHD.**Methods:** We tested our hypothesis using streptozotocin-induced mouse model of type-I diabetes in mice heterozygous for a null allele for *Notch1* and performed morphological analyses on E13.5 embryos. Protein and mRNA expression were measured using immunohistochemical staining and qRT-PCR. *In vivo* results were also validated using an atrioventricular mesenchymal (AVM) cell line and chick embryo model system in presence of high glucose to dissect the underlying mechanism of gene-environment interaction.**Results:** Histologic analyses indicated that wildtype (WT) embryos [2/9(22%)] exposed to maternal diabetes have increased incidence of ventricular septal defect (VSD) compared to [0/9 (0%)] non-diabetic embryos. The incidence of VSD was further increased for diabetic *Notch1*+/- mice [12/14(85.7%)]. Analysis of non-diabetic and diabetic WT embryonic hearts revealed that hyperglycemia is associated with a gradual decrease in *Notch1* mRNA and its downstream target genes (*Hey1, Hey2, Nrg1*) with concomitant upregulation in *Jarid2* mRNA expression, an epigenetic histone mark known to regulate *Notch1*. We also detected increased production of ROS and decreased NO bioavailability in AVM cells treated with 25mM of glucose compared to normoglycemia (5.5mM). *Jarid2* expression also increased with hyperglycemia both at RNA and protein levels. *In vitro* chromatin immunoprecipitation data suggested relative enrichment of Jarid2 on *Notch1* locus with hyperglycemia. Together our data demonstrate that maternal hyperglycemia disrupts cardiac development by affecting the Notch1 signaling pathway through Jarid2-mediated repression.

571W

Common Genotype and Environmental Risk Factors of Myocardial Infarction and High Blood Cholesterol in a Retrospective Population Study. *Y. Liang[1], A. Kelemen[2].* 1) 1Department of Family and Community Health, University of Maryland, Baltimore, Baltimore, MD; 2) 2Department of Organizational Systems and Adult Health, University of Maryland, Baltimore.

This article assesses whether risk alleles related to MI may also relate to high cholesterol; and to evaluate prediction power of the joint genetic and environmental factors influencing the risk of MI and cholesterol. The retrospective study consists of 1837 subjects, 818 having history of MI and 1019 with no history of MI aged 35 to 69 years. The data includes 29 variables such as smoking status, alcohol use, etc. 81 SNPs linked to a metabolic pathway were genotyped and preselected from the SeattleSNPs variation discovery panel. SNPs *car_CEPT_3* [OR=1.431, (1.047, 1.956)] and *car_NOS_2* [OR=1.413, (1.063, 1.877)] were found to be significantly associated with risk of MI. High cholesterol has strongest effect on MI [OR=5.397, (4.069, 7.160)] compared with other factors. HBP [OR=2.420, (1.819, 3.219)], male gender [OR= 4.252, (3.087, 5.857)], smoking status [OR=3.399, (2.376, 4.862) for current smoker; [OR=.447, (.321, .622)] for former smoker], drinking status in past 1-2 yrs [OR=.672, (.502, .901)], awere found significantly associated with MI. The interaction between drinking and smoking status was significant [OR=.459, (.220, .956)] for current smokers who drank in the past 1-2 yrs. The prediction power was increased from 72% to 80% when including both genetic and environmental factors and their interactions compared with environmental factors only. The presence of the two SNPs, *il1_2_12* and *alox_2_22* increased the odds of the likelihood of high cholesterol by 27% [OR=1.272, (1.045, 1.549)] and 42% [OR=1.424, (1.027, 1.974)] respectively. Adjusting for the demographic and clinical factors, the odds of the likelihood of high cholesterol were 33% lower [OR=0.671, (0.511, 0.881)] for people who had the *cept_2_12* allele than for people who did not have it; the odds of the likelihood of high cholesterol were 99% lower [OR=0.011,(.000, .249)] for people having no history of MI than for people having it. Gene-gene and gene-environmental interaction effects were found significantly associated with high cholesterol (gender and *cept_2_12*, family history of heart diseases and *car_esr1_2*, no MI history and age, *Alox_2_22* and *cept_2_12*). Having history of either MI or high cholesterol conditions has significant effect on each other in both ways. The common alleles gene variants (*il1* and car_*cept*.) associated with both MI and high cholesterol has been discovered, which may be potential therapeutic target for personalized treatment and precision medicine.

572T

Utilizing gene expression to uncover genotype-dependent effects of BMI in multiple tissues. *C.A. Glastonbury[1], A. Viñuela[1,2,3,4], A. Buil[2,3,4], G. Thorleifsson[5], U. Thorsteinsdottir[5,6], E. Dermitzakis[2,3,4], T. Spector[1], K. Small[1].* 1) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 2) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 3) Institute for Genetics and Genomics in Geneva (iGE3), University of Geneva, Geneva, Switzerland; 4) Swiss Institute of Bioinformatics, Geneva, Switzerland; 5) deCODE Genetics, Sturlugata 8, IS-101 Reykjavik, Iceland; 6) Faculty of Medicine, University of Iceland, 101 Reykjavik, Iceland.

On the population level obesity is robustly associated to a range of co-morbidities. However, despite a demonstrated causal role of obesity in co-morbidities such as Type 2 Diabetes, co-morbidity incidence amongst obese individuals is heterogeneous. This heterogeneity may reflect genotype-dependent responses to obesity, here termed gene x environmentBMI (GxEBMI) interactions. We leveraged the power of transcriptomics to identify GxEBMI regulatory effects in RNA-seq data from three primary tissues and one cell line (Adipose, Skin, Whole Blood & LCLs) from 856 healthy female twins from the TwinsUK cohort. A genome-wide *cis* GxEBMI scan (±1MB from TSS, MAF >5%) identified four genome-wide significant (GWS) GxEBMI effects in Adipose (*ALG9, HACL1, GAA, SMG6*) and one in Skin (*RHPN2*). Two of three GWS effects testable in the DeCODE adipose expression dataset (N = 673) replicated ($p < 0.05$, consistent direction of effect) demonstrating our findings are robust. We tested the GWS GxEBMI variants for *trans* effects (±5MB from TSS) and found rs3851570 (*cis* GxEBMI of *ALG9)*, is a *trans*- GxEBMI regulator of 151 genes in adipose (Top hit: *ZNF423* p=8.2x10-13, bonferroni corrected cutoff = 4.3x10-7**)**. The 151 *trans*- GxEBMI genes are enriched for oxidative phosphorylation (B-H, p = 1.4x10-3) and respiratory chain processes (B-H, p=8.3 x 10-3) indicating rs3851570 interacts with BMI to regulate cellular metabolism. We characterized properties of adipose GxEBMI using 423 effects identified at a relaxed threshold of p < 1x10-5 These GxEBMI are highly tissue-specific (ϖ1=0), not directly associated to BMI ($p > 0.05$; GIANT dataset, N = 250K) nor enriched for main effect *cis*-eQTLs (ϖ1=0.03). GxEBMI genes are enriched for GO metabolic process ($p = 0.01$) and GxEBMI SNPs are enriched for low p-values in large GWAS's of Insulin ($p = 0.001$) and HOMA-IR ($p = 0.006$). These findings suggest utilizing gene expression to detect GxE is a powerful approach to understand the development of obesity and adipose-relevant complex traits. As adipose tissue is a heterogeneous mix of cells, we computationally deconvolved each sample to estimate the relative proportion of 19 cell-types. The most abundant cell type estimated are adipocytes (median=0.72) followed by Macrophages (median=0.07). Overall, the immune cell fraction is strongly correlated to adiposity (% trunk fat, $r2 = 0.39$, $p < 2.2x10-16$). Ongoing work aims to identify cell-specific *cis*-eQTLs and their relationship with complex traits.

573F

What next after GxE GWAS: A bioinformatic functional characterization of the *EBF1* gene and stress interaction signal. *A. Singh, M.A. Babyak, B.H. Brummett, R. Jiang, I.C. Siegler, W.E. Kraus, S.H. Shah, R.B. Williams, E.R. Hauser.* Duke University Medical Center, Durham, NC.

Identifying the functional relationship between risk genotypes and phenotypes and understanding the influence of the environment are important steps in characterizing the genetic architecture of GxE interactions. In our previous study (Singh et al. 2014) a GxE GWAS showed that in the presence of chronic psychological stress in two population-based cohorts (MESA and Framingham), common variants in the *EBF1* gene (SNPs rs4704963, rs17056278, R2=1) contribute to inter-individual differences in human obesity, diabetes, and cardiovascular disease risk factors. The gene-by-stress interaction was replicated in three additional studies (Family Heart Study, Duke Caregiver Study, and CATHGEN Cohort) (Singh et al. 2015). In this work, we performed functional characterization using bioinformatic approaches, an existing curated knowledgebase (i.e., Ingenuity® Pathway Analysis (IPA)), and gene expression data from the CATHGEN cardiovascular cohort study (Burns et al., 62nd ASHG, 2012). In the IPA knowledgebase the T2D pathway showed the role of *GLUT4* gene in the development of insulin resistance and the *EBF1* gene interaction network showed an interacting-link between *EBF1* and a known diabetes gene *SLC2A4* (i.e., *GLUT4*). We compared the genomic sequence of the GWAS signal in multiple vertebrate species (Human, Rhesus, Mouse, Dog, Elephant, Opossum, etc.); a SNP from the GWAS signal (rs17056278) is located on a highly conserved site that includes a candidate for conserved transcription factor binding site *CCAT* (tfbscons) suggesting a potential functional role for the associated SNP. Using linear regression under the default additive genetic model adjusting for age and sex, we tested the *EBF1* SNP in an eQTL analysis with and without a stress interaction in CATHGEN Whites where chronic psychosocial stress was a computed variable (Singh et al 2015). There was a statistically significant differential expression of *EBF1* (P=0.034), expression levels being lower for the minor allele group. We did not detect a significant SNP x chronic psychosocial stress interaction on expression of *EBF1;* however, the effects were consistent with the original GWAS association in terms of the direction and trend, i.e. *EBF1* gene expression increased with increasing chronic psychosocial stress for the carriers of the minor allele but not for major allele homozygotes. To examine the relationship between chronic psychosocial stress, *EBF1* variants, and gene expression, a larger study is required.

574W

Comprehensive analysis of established dyslipidemia-associated loci in the Diabetes Prevention Program. *T.V. Varga[1], A.H. Winters[2], K.A. Jablonski[3], E.S. Horton[4,5], P. Khare-Ranade[6], W.C. Knowler[7], S.M. Marcovina[8], F. Renström[1,9], K.E. Watson[10], R. Goldberg[11,12], J.C. Florez[4,5,13,14], T.I. Pollin[2], P.W. Franks[1,15,16], Diabetes Prevention Program Research Group*. 1) Department of Clinical Sciences, Lund University, Malmö, Skåne, Sweden; 2) Division of Endocrinology, Diabetes, and Nutrition, Department of Medicine, and Program in Genetics and Genomic Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA; 3) The Biostatistics Center, George Washington University, Rockville, Maryland, USA; 4) Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA; 5) Diabetes Research Center, Diabetes Unit, Massachusetts General Hospital, Boston, Massachusetts, USA; 6) Washington University, St. Louis, Montana, USA; 7) Diabetes Epidemiology and Clinical Research Section, NIDDK, Phoenix, Arizona, USA; 8) Northwest Lipid Metabolism and Diabetes Research Laboratories, University of Washington, Seattle, Washington, USA; 9) Department of Biobank Research, Umeå University, Umeå, Sweden; 10) Department of Medicine, UCLA School of Medicine, California, USA; 11) Lipid Disorders Clinic, Division of Endocrinology, Diabetes, and Metabolism, Leonard M. Miller School of Medicine, University of Miami, Miami, Florida, USA; 12) The Diabetes Research Institute, Leonard M. Miller School of Medicine, University of Miami, Miami, Florida, USA; 13) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 14) Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA; 15) Department of Nutrition, Harvard Chan School of Public Health, Boston, Massachusetts, USA; 16) Department of Public Health & Clinical Medicine, Umeå University, Umeå, Sweden.

Background: We assessed whether 234 established dyslipidemia-associated loci modify the effects of metformin treatment and lifestyle intervention (vs. placebo control) on lipid and lipid sub-fraction levels in the Diabetes Prevention Program (DPP) randomized controlled trial.*Methods*: We tested gene-treatment interactions on 1-year changes in blood lipid concentrations (high and low density lipoprotein cholesterol [HDL-C, LDL-C], total cholesterol, triglycerides) and lipoprotein sub-fraction particle concentrations and size in 2,993 non-diabetic DPP participants. *Results*: Of the previously reported SNP associations, 32.5% replicated at $P<0.05$ at baseline. Trait-specific genetic risk scores (GRS), defined as sums of the risk alleles of all trait-associated variants, were robustly associated with their respective baseline traits ($3\times10\text{-}4>P>1.1\times10\text{-}16$) for all but two traits (medium HDL and intermediate density lipoprotein, $P>0.05$). Metformin and lifestyle modified the effect of the GRS for large HDL particle numbers, such that from baseline to 1-year, each risk allele of the GRSHDL-large associated with lower large HDL levels in the metformin (b=-2.7% per GRS allele; 95%CI -4.5, -1.0; $P=2\times10\text{-}3$; *Pinteraction*=0.02 for metformin vs. placebo) and the lifestyle arm (b=-2.8% per GRS allele; 95%CI -4.6, -1.1; $P=2\times10\text{-}3$; *Pinteraction*=$5\times10\text{-}3$ for lifestyle vs. placebo), but not in the placebo arm ($P>0.05$). Moreover, lifestyle modified the effect of GRSHDL-size, such that a higher GRS associated with lower HDL particle size from baseline to 1-year in the lifestyle arm (b=-0.02nm per GRS allele; SE=0.007; $P=1\times10\text{-}3$; *Pinteraction*=$8\times10\text{-}3$ for lifestyle vs. placebo), but not in the metformin or placebo arms ($P>0.05$).*Conclusions*: Improvements in HDL particle size conferred by metformin and lifestyle may be diminished by high genetic loading for smaller particle size.

575T

Understanding the Biological Basis of G×E Interactions in Cardio-metabolic Disease: Methylation Profiles, Genetic Variation and Caregiver Stress. *R. Jiang[1], L. Kwee[2], A. Singh[1], S.H. Shah[2], B.H. Brummett[1], M.A. Babyak[1], I.C. Siegler[1], R.B. Williams[1], E.R. Hauser[2].* 1) Department of Psychiatry and Behavioral Sciences,Duke University Medical Center, Durham, NC; 2) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC.

Previous studies have demonstrated and replicated an interaction between genetic variants and psychosocial stress. Minor alleles of two SNP variants (rs439401, rs157580) were associated with an adverse metabolic profile (*e.g.*, higher triglyceride levels) in stressed caregivers, but a favorable or neutral metabolic profile in non-stressed controls. We hypothesized that this interaction between psychosocial stress and genetic variation may be due to differential DNA methylation at three levels: average methylation in the candidate region (AMCR), average global methylation (AGM) and genome-wide, probe-specific methylation. We selected age- and gender-matched caregiver/control pairs and assayed the Cpg methylation level using the Infinium HumanMethylation450 BeadChip. Sample and methylation probe quality control was performed using the RnBeads package in R, resulting in 112 samples with 462,936 CpG sites for association analysis of methylation levels and SNP genotypes. The CpGAssoc package was used for statistical analysis of AMCR, AGM and individual CpG sites. We did not observe genome-wide significant p-values ($p<10\text{-}7$) with any of the individual CpG sites or a significant p-value for AGM ($p>0.05$) for association with caregiver stress or genotype. AMCR was calculated for 73 CpG sites in the 104 kb PVRL2-TOMM40-APOE-APOC region, controlling for age, gender and six PCs. Our results showed that rs439401 C/C carriers had lower AMCR than C/T and T/T carriers (p=0.045). Further analysis of rs439401 by caregiver/control status showed that this pattern was seen only in caregivers, but not in the controls. C/C caregivers had lower AMCR compared with C/T caregivers (P=0.014), and T/T caregivers (P=0.076). A similar pattern was observed for rs157580 in that G/G caregivers had higher AMCR compared to A/A (P=0.004) and A/G in caregivers (P=0.008), but not in controls. The effect of SNP variants on methylation level was seen only in caregivers, and the direction of the pattern was consistent with the relationship between SNP variants and metabolic profile in our previous studies. These findings indicate that the stressed-caregiver/non-stressed control status may moderate the association of SNP variants through methylation levels in the candidate gene region, and thus, have an impact on the metabolic phenotypes. Further replication in a larger sample is required as well as additional studies of the role of stress, SNP variants, and DNA methylation.

576F

Genome-wide association study identifies variants that predict survivorship in individuals with coronary artery disease. *J.R. Dungan[1], X. Qin[2], M. Hurdle[2], C. Haynes[2], E. Grass[2], D. Craig[2], S.G. Gregory[2], E.R. Hauser[1,2], S.H. Shah[2], W.E. Kraus[1,2].* 1) School of Nursing, Duke University, Durham, NC; 2) Department of Medicine, Duke University, Durham, NC.

Background: We have previously reported a novel phenotype, 'survivorship in coronary artery disease (CAD)'; using a candidate gene (*LSAMP,* identified from linkage mapping in early-onset CAD) we demonstrated proof-of-principle for genetic effects unique to CAD cases for both mortality risk and improved survival. To evaluate such effects at a genome-wide level we now conduct a genome-wide association (GWA) analysis of this phenotype and hypothesize that novel candidate genes could predict survival in clinically-appreciable CAD.**Methods:** A GWA analysis was conducted for survivorship in individuals with CAD using discovery (*N*=1,099) and replication cohorts (*N*=414) of Caucasians selected from a cardiovascular biorepository (CATHGEN) of sequential patients referred for cardiac catheterization. CAD cases were defined as having a CAD index >32 (at least one-vessel CAD) and indication for catheterization was concern for ischemic heart disease. Illumina Omni Quad v1.0 GWAS chip was used with stringent QC metrics. Cox proportional hazards regression models for each SNP (additive) adjusted for age, sex, four Eigenstrat-defined principal components and seven major cardiovascular risk covariates were constructed using the R statistical package. Models were censored on time from cardiac catheterization to death (all-cause) or last follow-up time (*med* 5.2 years). Genome-wide significance was defined as $p<10\text{-}5$ for discovery and $p<0.05$ for replication. Meta-analyses of discovery and validation cohorts were conducted with a combined Z-score approach using METAL.**Results:** Of the 93 SNPs meeting genome-wide significance in the discovery analysis three SNPs replicated with nominal significance, having the same direction of effect in both cohorts: rs13007553 (*MYT1L/TSSC1*; *HR* 1.31, *95% CI* 1.01-1.7, *p*discovery=5.93x10-5, *p*replication=0.04 , *p*meta=1.05x10-5; rs587936 (*DAB2IP; HR* 0.80, *95%CI* 0.61-1.05, *p*d=3.75x10-5 *p*r=0.02, *p*m=2.13x10-5); and rs1993024, (*FSTL5/NAF1*; *HR* 0.75, *95%CI* 0.55-1.03, *p*d=4.04x10-5*, p*r=0.02, *p*m=0.03). Of the discovery SNPs, 30 were significant in meta-analysis ($p<0.001$) and had the same direction of effect in both cohorts.**Discussion:** We have identified genome-wide variants that appear to predict survivorship in CAD cases and which were not identified from prior GWAS of CAD. Similar to previously reported candidate gene results for *LSAMP*, both *TSSC1* and *DAB2IP* are related to tumor suppression suggesting a novel shared mechanism mediating survivorship in CAD.

577W

Comprehensive characterization of the genetic architecture of sudden cardiac death. *F.N. Ashar[1], C. Albert[2], S.S. Chugh[3], A. Cupples[4], M. Eigelsheim[5], P. Goyette[6], A. Huertas-Vasquez[3], H. Huikuri[7], J. Juntilla[7], X. Jouven[8], S. Kaab[9], M. Kortelainen[10], P. Kwok[11], T. Lehtimaki[12], L. Lyytikainen[12], M. Muller-Nurasyid[13], C. Newton-Cheh[14], B. Psaty[15], S. Pulit[14], D. Siscovick[15], B. Stricker[5], N. Sotoodehnia[15], D.E. Arking[1], CHARGE-SCD.* 1) Johns Hopkins School of Medicine, Baltimore, MD; 2) Division of Preventive Medicine, Cardiovascular Division, Brigham and Women's Hospital, Boston, Massachusetts, USA; 3) The Heart Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA; 4) Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA; 5) Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; 6) Università de Montréal, Montreal, Quebec, Canada;; 7) Dept. of Internal Medicine, University of Oulu and University Central Hospital, Oulu, Finland; 8) Università Paris Descartes, Assistance Publique-Hopitaux de Paris, Hôpital Européen Georges Pompidou, Paris, France; 9) Department of Medicine I, University Hospital Grosshadern, Ludwig-Maximilians-Universitàt, Munich, Germany; 10) Medical Research Center Oulu, University Hospital of Oulu, and University of Oulu, Finland; 11) Department of Dermatology, Cardiovascular Research Institute, and Institute for Human Genetics, University of California, San Francisco, California, USA; 12) Department of Clinical Chemistry, Fimlab Laboratories, School of Medicine, University of Tampere, Finland; 13) Institute of Genetic Epidemiology, Helmholtz Zentrum München-German Research Center for Environmental Health, Neuherberg, Germany; 14) Center for Human Genetic Research, Cardiovacular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA; 15) Cardiovascular Health Research Unit, University of Washington, Seattle, Washington, USA.

As part of the CHARGE-Sudden Cardiac Death (SCD) working group, we have conducted the largest GWAS (4,496 cases, >25,000 controls) to date for sudden cardiac death, a phenotype defined operationally as a sudden unexpected pulseless condition due to a ventricular arrhythmia. Although the primary cause of SCD is considered to be electrical instability, the majority of SCD occurs in the setting of coronary artery disease (CAD).While no loci reached genome-wide significance following replication in our GWAS, we used this dataset to interrogate the effects of the group of SNPs associated with SCD risk factors, on sudden death risk. We employed a multi-SNP genetic risk score (GRS) approach that uses weighted linear regression to establish the relationship between SNP effect estimates for SCD and a risk factor of interest. We chose to focus on SNPs associated with two major etiological classes of SCD risk factors—ECG traits (QT, QRS, and PR intervals), and CAD associated traits (CAD, type2 diabetes [T2D], BMI, HDL, LDL, total cholesterol and triglyceride levels) at five different alpha cutoffs (α) (5e-8, 1e-5, 1e-3, 0.05 and 0.99).Of the ECG parameters, we see a statistically significant enrichment of SNPs associated with QT interval at all alpha cutoffs (at α=5e-8, p=0.007; when α=0.99, p=7.41e-7). In contrast, we find no evidence of association between the set of SNPs associated with QRS duration or PR interval and SCD at any alpha cutoff. When we focus on CAD associated traits, GRS models constructed with T2D or CAD SNPs showed statistically significant enrichment in SCD at all alpha cutoffs examined (at α=5e-8, p for CAD=2.48e-6; p for T2D=0.012). Furthermore, while the risk scores constructed from SNPs associated with lipid levels, did not predict SCD risk, GRS with BMI SNPs at 5e-8 cutoff was strongly predictive of SCD risk (p=1.29e-4).Taken together, this analysis is the first to establish the link between common variants associated with risk factors related to SCD pathophysiology and SCD risk. Furthermore, this study not only offers further support for existing epidemiological observations establishing prolonged QT interval, CAD, and T2D as SCD risk factors but also presents a compelling case to investigate the biological role for loci associated with BMI in sudden death.

578T

DNA methylation profiling identifies a locus near a HMG-CoA synthase gene associated with triglyceride levels and modified by BMI. *F. Gagnon[1], V. Truong[1], N. Zwingerman[1], I. Kassam[1], D. Aïssi[2], J. Dennis[1], M. Wilson[3], P. Wells[4], P.-E. Morange[5], D.-A. Trégouët[2].* 1) Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada; 2) INSERM, UMR_S 1166, Team Genomics & Pathophysiology of Cardiovascular Diseases, Paris, France; 3) Genetics and Genome Biology Program, SickKids Research Institute, Toronto, Canada; 4) Department of Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Canada; 5) INSERM, UMR_S 1062, Nutrition Obesity and Risk of Thrombosis, Marseille, France.

In a candidate gene study using biobanked whole blood DNA, we reported an association of DNA methylation marks in the *CTP1A* gene with triglyceride (TG) levels in families and cases of venous thromboembolism (VTE). The *CTP1A* gene encodes for a protein expressed in liver that is essential for fatty acid oxidation, which supports the role of DNA methylation in TG regulation.To further explore the role of epigenetic marks in inter-individual variation of fasting TG levels, we conducted a genome-wide association study between TG and DNA methylation levels measured in whole blood from 214 individuals from the *French-Canadian Family study on F5L Thrombophilia* (F5L). The F5L study is composed of 5 large families ascertained on single probands with VTE. Replication was tested in 350 unrelated VTE cases from the *MARseilleTHrombosis Association Study* (MARTHA). We used linear mixed regression models adjusted for age, sex, cell type proportions, and familial correlation when analyzing the *F5L study*. Sensitivity analyses assessing potential confounders and effect modifiers were conducted.We replicated the recently reported association between TG and a DNA methylation mark in *ABCG1*, a lipid-related gene, and discovered a novel genome-wide significant association in the *PHGDH* gene. The effect size at the *PHGDH* locus was stronger in the F5L study (b=-0.21, p=2.3×10-7) than in MARTHA (b=-0.08, p=0.048). The prevalence of overweight and obese individuals (BMI≥25) was 56% and 36%, respectively, in the F5L and MARTHA studies. A stratified analysis in MARTHA revealed BMI as an effect modifier, with an effect only detected in overweight and obese individuals (b=-0.17, p=9.8×10-3) compared to individuals with normal BMI (b=-0.001, p=0.99).Follow-up analyses showed that no SNPs within 1Mb of the identified *PHGDH* locus confounded the association between the methylation marker and TG levels; but a SNP located in *HMGCS2*, a gene with similar functions as *CTP1A* and only 34 kb from *PHGDH, was* associated with the TG-methylation marker. *HMGCS2* belongs to the HMG-CoA synthase family known for its pivotal role in cholesterol synthesis and ketogenesis. A preliminary Mendelian Randomization analysis suggests an important but indirect role of the *PHGDH* methylation mark on TG level variation.These causal inference results shed light on the molecular mechanisms underlying TG level variation, and provide putative insights on the VTE-obesity association.

579F

Coronary Collateralization Shows Sex and Ethnic Differences in the Presence of Obstructive Artery Disease. *Z. Liu[1], L. Wang[1,2], N. Vasudeva[2], P.J. Goldschmidt-Clermont[3], M.A. Pericak-Vance[1,2], D.M. Seo[3], G.W. Beecham[1,2].* 1) Human Genetics, John T. Macdonald Foundation, University of Miami, Miami, FL; 2) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 3) Division of Cardiology, Miller School of Medicine, University of Miami, Miami, FL.

BACKGROUND: Coronary collateralization is a natural process of revascularization in the cardiovascular system. Collateral circulation protects cardiac tissues from myocardial infarction (MI) by allowing oxygenated blood to bypass the blockage in main coronary arteries. As such, collateral circulation may partly affect MI outcomes such as sudden cardiac death.**METHOD:** To better understand the genetic mechanisms of coronary collateralization, we assessed 879 patients with obstructive coronary artery disease (over 50% stenosis presented in any of the four major coronary branches: left main, LAD, LCX, RCA) from our cardiac catheterization patient cohort. Patients were assessed for presence/absence of collaterals, as well as age, sex, diabetes, anti-cholesterol medicine usage, etc. DNA samples were genotyped using the Affymetrix 6.0 genotyping platform. Due to diverse genetic backgrounds of our samples, we calculated local and global ancestry, and subsequently performed admixture mapping using logistic regression, with the presence of collaterals as an endpoint.**RESULTS:** We found sex differences in collateralization: men had statistically significantly higher rates of collaterals than women (p-value = 1.38E-04). Different race/ethnicity groups showed different rates of collateralization: white Hispanics/Latino showed higher rates of collaterals (59%) than African Americans (50%) and non-Hispanic Whites (48%) (chi-square p-value = 1.7E-02). The admixture mapping analysis showed Native American ancestry was associated with the presence of collaterals at a region on chromosome 17 (chr17:36120051–40782083, beta = 0.55, min p-value = 1.27E-04). African ancestry also showed association with collaterals at a different region on chromosome 17 (chr17:32796666–33440166, beta = 0.38, min p-value = 7.2E-04).**CONCLUSIONS:** In our study, men have higher rate of collateralization than women, which may in part explain why women have poor outcomes after MI than men. Different race ethnicity groups have different rates of collateralization, which may partially explain racial disparities in MI outcomes. Finally, we identified two regions on chromosome 17 that are likely to harbor genetic variations that influence collateralization. A following up in-depth study of these two regions is ongoing.

580W

Exome chip meta-analysis identifies novel loci contributing to lipid levels in Asian population. *X. Lu[1], H. Zhang[1], W. Zheng[2], S. Ganesh[1], P. Sham[3], K. Mohlke[4], W. Sheu[5], J. Rotter[6], Z. Mo[7], R. Dorajoo[8], E. Tai[8], X. Lin[9], T. Wu[10], D. Gu[11], C. Willer[1], the Asian exome chip lipids Consortium.* 1) Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, Michigan, USA; 2) Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN, USA; 3) Department of Psychiatry, the University of Hong Kong, Pokfulam, Hong Kong; 4) Department of Genetics, University of North Carolina, Chapel Hill, NC, USA; 5) Department of medicine, Taichung veterans general hospital, Taichung, Taiwan; 6) Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, California, USA; 7) Center for Genomic and Personalized Medicine, Guangxi Medical University, Nanning, Guangxi, China; 8) Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore; 9) Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences and Graduate School of the Chinese Academy of Sciences, Shanghai, China; 10) MOE Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China; 11) Fuwai Hospital, National Center for Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China.

Genome-wide association studies (GWAS) have identified ~160 loci contributing to lipid levels, which are heritable risk factors for cardiovascular diseases. However, most of the published variants are non-coding and explain less than half of the heritability of lipid levels. Thus, we attempted to evaluate the effect of the coding variants on lipid levels, particularly in the Asian population who have not been studied as extensively as Europeans. To identify coding and low-frequency variants associated with plasma lipids levels, we collected Exome chip genotyping data from more than 39,000 Asian samples from 20 cohorts. Lipid levels were transformed using the inverse normal distribution after adjusting for age, age2, sex, principal components. Single variant analyses in each cohort were carried out either using rvtests or RareMetalWorker. The meta-analysis of single variant and gene-level association tests were performed using RAREMETALS for HDL, LDL, triglyceride and total cholesterol levels. We identified a novel non-synonymous variant in *CD163* (HDL, $P = 3.98 \times 10\text{-}9$). Furthermore, we identified a common non-coding variant in *ACVR1C* with TC ($P = 4.83 \times 10\text{-}8$) that were not previously reported in large-scale GWAS. We also replicated 28 loci reported in prior GWAS, involving 38 coding variants. In gene-based tests, a burden of rare/lower frequency variation in *PCSK9*, *HMGCR*, *CETP*, *APOA1*, and *CD36* were associated with lipids levels. Results of the gene-level burden tests were driven primarily by single variant associations. Our study identifies novel coding variant associations and extends the allelic spectrum of variation underlying lipid levels.

581T

Molecular basis of regulatory variation at coronary heart disease associated loci. *C.L. Miller[1], M. Pjanic[1], T. Wang[1], T. Nguyen[1], J. Lee[2], B. Liu[2,3], T. Assimes[1], S. Montgomery[2,3], T. Quertermous[1].* 1) Cardiovascular Medicine, Stanford University School of Medicine, Stanford, CA; 2) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 3) Department of Pathology, Stanford University School of Medicine, Stanford, CA.

Genome-wide association studies have identified 46 replicated genetic loci for coronary heart disease (CHD), and 104 loci associated at a 5% false discovery rate. However, the regulatory mechanisms of these associations largely remain undefined. Given that the majority of these CHD-associated loci reside in non-coding regions, they are predicted to function via context-specific gene regulation. Recent high-throughput assays of regulatory function include the assay for transposase-accessible chromatin using sequencing (ATAC-seq) and chromatin immunoprecipitation-sequencing (ChIP-seq). ATAC-seq utilizes a Tn5 transposase to fragment and tag accessible DNA sequences, which are often coupled to transcription factor occupancy identified by ChIP-seq. Importantly, this assay may reveal the spatio-temporal regulatory profiles in limited numbers of primary cells. Using ATAC-seq in primary human coronary artery smooth muscle cells (HCASMC) we identified 147,173 accessible chromatin peaks in control versus 198,976 peaks in TGF-beta-stimulated cells (136,446 shared peaks). Using *de novo* motif enrichment analysis we identified significant enrichment of specific AP-1 family members (29.2% vs. 5.1% background), chromatin remodeling, and SMC differentiation transcription factors. Using functional enrichment analysis of overlapping regions in open chromatin, ChIP-seq and CHD loci we observed enrichment of the hypoxia inducible factor 1 (HIF-1) and TGF-beta signaling pathways (1.5x10-22 and 5.6x10-18, respectively) and relevant phenotypes, including cell migration and blood vessel morphology. Finally, we utilized these regulatory maps to explore the causal mechanisms underlying CHD-associated variants at four loci using haplotype-specific chromatin immunoprecipitation (haploChIP), luciferase reporter assays, and allelic expression imbalance in HCASMC. Taken together, these results suggest that genome-wide approaches such as ATAC-seq can be leveraged to map context-specific regulatory mechanisms of non-coding variants associated with complex diseases such as CHD, and reveal new biological and molecular insights into targeting heritable disease risk.

582F

A genome-wide association study of ischemic stroke and its subtypes identifies a novel locus near *TSPAN2*. *S.L. Pulit, National Institute of Neurological Disorders and Stroke - Stroke Genetics Network (SiGN).* University Medical Center Utrecht, Utrecht, Netherlands.

Stroke is the 2nd leading cause of death worldwide and the annual number of stroke deaths is rising. Hypertension, atrial fibrillation, and smoking are known risk factors, but a large proportion of ischemic stroke (IS) risk remains unexplained. Given the established heritability of IS, genetic variation likely accounts for some of the residual risk. IS is highly heterogeneous; the 3 most common subtypes are cardioembolic, large artery atherosclerosis (LAA), and small artery occlusion (SAO). Genome-wide association studies (GWAS) have identified only 4 confirmed risk loci; 3 are subtype specific. With the aim of elucidating disease etiology, the Stroke Genetics Network sought to perform a comprehensive GWAS by assembling the largest sample to date and employing Causative Classification of Stroke (CCS), a system that considers neuroimaging results and ancillary examinations to obtain a comprehensive evaluation of subtype. As it is centrally-adjudicated and web-based, CCS also reduces inter- and intra-observer variability.After extensive quality control to merge cases collected worldwide and publicly-available controls, we ran discovery analyses in 16,851 cases and 32,473 controls. We ran GWAS for all IS and for each subtype in CCS and TOAST, the most commonly employed subtyping system used in all previous GWAS. Since only TOAST cases were available for replication, we calculated pairwise correlations of SNP effects between CCS and TOAST GWAS. Genetic correlation between CCS and TOAST primary subtypes was moderate to strong (r = 0.61–0.75), indicating that TOAST cases were suitable for replication. SNPs with p<10-6 were selected for *in silico* lookup in an independent set of 21,842 TOAST-subtyped IS cases and 367,842 controls.Joint analysis of all data revealed a novel association at rs12122341, a common SNP near *TSPAN2* conferring risk to LAA (OR[G] = 1.2, p = 1.3 x 10-9). The 12q24 locus, previously implicated in all IS, was associated predominantly with SAO (rs10744777, OR[T] = 1.17, p = 2.9 x 10-9).*TSPAN2* knockout mice exhibit neuroinflammation and TSPAN is highly expressed in arterial endothelial, smooth muscle cells, and white blood cells. Further studies will determine if the rich repository of phenotypic data created by CCS will allow for genetic analysis of novel phenotypes, or if leveraging the genetic overlap of CCS and TOAST through broader inclusion of cases reveals more risk loci, a vital first step in improving IS diagnosis and treatment.

583W

Chromosome 22q11 microdeletion syndrome: association of congenital heart disease with copy number variants. *G. Repetto, L. Leon, C. Vial, K. Espinoza, F. Benavides, M.L. Guzman.* Center for Genetics and Genomics, Clin Alemana-Univ Desarrollo, Santiago, Chile.

Chromosome 22q11.2 microdeletion syndrome (22q11DS) is one of the most common pathogenic genomic rearrangements in humans. Although the majority of patients share a deletion of similar size and location, 22q11DS is notorious for its highly variable clinical presentation. One of the relevant features is congenital heart disease (CHD), present in 50-60% of patients. It usually affects the cardiac outflow tract and includes defects such as tetralogy of Fallot, ventricular septal defects and truncus arteriosus, among others. The reason for the incomplete penetrance of CHD remains unknown. We hypothesize that structural genetic variation outside of the deletion region may have a modifier effect. We evaluated the association of copy number variants (CNVs) and the presence of CHD in a large cohort of Chilean patients with 22q11DS. The study included 133 patients with CHD (cases) and 120 with normal cardiac anatomy (controls). Frequency, size and location of constitutional CNVs were characterized using Affymetrix SNP 6.0 arrays. CNV calls were made with the GC model wave adjustment procedure using PennCNV software. Genome-wide CNV association was evaluated using ParseCNV. Association was assessed by Fisher´s exact testand odds ratio. The large 3Mb LCR A-D deletion was found in 92% of patients; the remainder had 2 or 1.5 Mb deletions (LCR A-C or A-B, respectively) . Statistically significant difference in frequencies were found for a duplication on chromosome region 1p36.13. The duplication was present in 8.5% of cases and 1.7% of controls (p value 0.02, OR 5.35, 95% confidence interval 3.83-6.88). The region contains noncoding RNAs, and we propose that they may exert a modifying role over the primary effect of the deletion. Funded by Fondecyt-Chile grant # 1130392.

584T

Causal effect of blood plasminogen activator inhibitor type 1 level on increased risk of coronary heart disease. *C. Song[1,2], A.D. Johnson[1,2], C.J. O'Donnell[1,2].* 1) National Heart, Lung and Blood Institute/ Framingham Heart Study, Framingham, MA; 2) NHLBI Division of Intramural Research.

IntroductionPlasminogen activator inhibitor type 1 (PAI-1) plays an essential role in the fibrinolysis system and thrombosis. Previous studies have reported elevated blood PAI-1 levels associated with increased risk of coronary heart disease (CHD). However, it is unclear whether the association reflects a causal influence of PAI-1 on CHD risk versus an effect of shared regulators. **Methods**To evaluate for a causal role in the association between PAI-1 and CHD, we performed a systematic review and meta-analysis. We conducted an electronic literature search in PubMed of PAI-1 and CHD/Coronary artery disease/Myocardial infarction published in English between 1992 and 2015. To investigate the causal association between PAI-1 and CHD, we used genetic risk scores (GRSs) composed of PAI-1-related SNP variants, reported by the largest available genome-wide association study (GWAS) of PAI-1, as instrumental variables (IVs). We then tested whether elevated PAI-1 levels were associated with a higher CHD risk by the Mendelian Randomization (MR) approach using summarized statistics from GWASs for PAI-1 and for CHD. We integrated results by using PAI-1-related GRS composed of SNPs with different P-value thresholds in GWAS for PAI-1 (SNP association thresholds P<5e-8, 0.001, 0.01, and 0.05) as IVs to address potential bias.**Result**In the systematic meta-analysis, the highest quantile of blood PAI-1 level was associated with higher CHD risk comparing with the lowest quantile [Odds ratio (OR) =1.68; 95% confidence interval (CI): 1.41, 2.01] in an age-and sex-adjusted model. The effect size was reduced but remained significant in studies using a multivariable-adjusted model. The MR analyses consistently showed a similar causal effect size of increased PAI-1 level on risk of CHD using IVs at different P-value thresholds, and the significance level increased when using a GRS with more SNPs (OR=1.02 per unit increased of PAI-1; 95% CI: 1.02, 1.03 when using PAI-1-SNPs with P<0.05).**Conclusion**Our study indicated a small to moderate causal effect of elevated PAI-1 level on CHD risk. Further study is warranted to understand the specific molecular determinants of circulating PAI-1 on CHD risk.

585F

Genetic determinants associated with BP phenotypes. *N. Vasudeva[1], L. Wang[1,2], Z. Liu[2], P. Goldschmidt[3], M. Pericak-Vance[1,2], D. Seo[3], G. Beecham[1,2].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Dr. John T. Macdonald Foundation, Department of Human Genetics, University of Miami, Miami, FL; 3) Division of Cardiology, Miller School of Medicine, University of Miami, Miami, FL.

Hypertension is a leading cause of mortility and morbidity in the developed countries. These traits are well established as a heritable traits. Genetic variants may lead to changes in BP. However the underlying environmental factors as well as their complex interactions makes discovering the genetic determinants affecting these trait a challenging task. Multiple variants are believed to cumulatively contribute to this polygenic trait. Genome-wide genotyping was performed on 1975 subjects using Omni 2.5 Million SNPs chip. Our dataset consists of multi-ethnic populations with ~12% Blacks, ~45% Hispanics and ~37% Whites. Sample and SNP quality control were followed by Eigenstrat. To account of the effect of blood pressure lowering medications, we added 7% and 6% to systolic blood pressure (SBP) and diastolic blood pressure (DBP), mean arterial pressure (MAP), pulse pressure (PulseP) measures, respectively, for each antihypertensive medication taken. Linear regression was used to analyze single variant association adjusting for covariates such as sex, age, age^2, principal component one, anti-cholesterol medicines and diabetes. Various SNPs approached the genome-wide significant level. SNPs at chr19:2,242,829 in gene SF3A2 (p-val: 9.31E-08) and chr3:88,668,147 (p-val: 2.68E-07) for SBP, chr11: 112,260,754 (p-val: 1.27E-08), chr19: 2,242,829 (1.78E-08) for MAP approached genome-wide significant association with BP phenotypes. We will further perform a gene-based association test using the SKAT R package. A web based tool WebGestalt will be used in order to find pathways connected to these large numbers of genes. To further analyze these genotypes and pathways, we will use alternate gene based methods such as VEGAS to better understand the genetic etiology of hypertension in our population.

586W

Genome-wide association study of CVD-related loci with lipid traits in the Heart Strategies Concentrating on Risk Evaluation (Heart SCORE) study. *X.B. Wang[1], S. REIS[2], I. Halder[2], M. Barmada[1], F.Y. Demirci[1], M.I. Kamboh[1].* 1) HUMAN GENETICS, University of Pittsburgh, PITTSBURGH, PA; 2) Medicine, University of PITTSBURGH, PA.

Cardiovascular disease (CVD) is the leading cause of death worldwide. Plasma levels of total cholesterol (TC), high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C) and triglycerides (TG) are significant risk factors for CVD and about 50% of the variation in these four quantitative traits is under genetic control. Two meta-analyses of published lipid genome-wide association studies have identified 157 loci that explain approximately 25-30% of the genetic variance for each of the four traits, indicating that additional genetic factors remain to be identified. In this study, we used a Illumina CVD BeadChip that contains approximately 49,000 SNPs in approximately 2,100 candidate genes selected based on their potential role in CVD to test for association with four lipid traits in 464 African American (AA) and 770 European American (EA) subjects as part of the Heart SCORE study. Seventy of the reported 157 genome-wide significant lipid loci/genes were represented on this chip. We observed significant associations (P<0.05) with SNPs in 55 of the 70 loci with strongest associations between HDL-C and *CETP*/ rs7499892 (P=3.92E-09), LDL-C and *APOE*/rs7412 (P=1.80E-06), and TG and *APOB*/rs2678379 (P=2.53E-06). In addition to the replication of known loci, we also observed multiple novel suggestive associations at P<1E-04 with the four major lipid traits. One of these novel loci, *SGK* (P=8.18E-05), was replicated in an independent study. In conclusion, we have replicated many of the known lipid loci and identified potentially novel loci for the four lipid traits using a novel CVD chip.

587T

CETP genotype is associated with phenotypic variability of HDL among other loci identified in a lipid vQTL study. *Y. Ma[1], Y. Cao[1], P. Wei[1], C.M. Ballantyne[2], J.M. Cheverud[3], C.S. Guild[4], C.E. Ndumele[5], E. Boerwinkle[1], T.J. Maxwell[6].* 1) Human Genetics Center, University of Texas Health Science Center, Houston, TX; 2) Section of Atherosclerosis and Vascular Medicine, Baylor College of Medicine, and Center for Cardiovascular Disease Prevention, Methodist DeBakey Heart and Vascular Center, Houston, TX; 3) Department of Biology, Loyola University, Chicago, IL; 4) Department of Medicine, University of Mississippi Medical Center,Jackson, MS; 5) Division of Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD; 6) Computational Biology Institute, The George Washington University, Ashburn, VA.

A variance-heterogeneity quantitative trait locus (vQTL) is a genetic locus associated with phenotypic variability. vQTL can be the result of linkage disequilibrium (LD) with a functional QTL. It can also be induced by gene-by-gene (GxG) or gene-by-environment (GxE) interactions. We performed vQTL screening for triglycerides, cholesterol, high-density lipoprotein (HDL), and low-density lipoprotein (LDL) in European Americans (EA) and African Americans (AA) separately using the whole-exome SNP genotype data of the Atherosclerosis Risk in Communities (ARIC) study. A parametric bootstrap-based mean and variance joint test followed by a variance only test was performed for vQTL screening. We identified eight significant vQTLs in or near CETP and one vQTL in LIPG for HDL, one vQTL for triglycerides, and one vQTL for cholesterol in EA. We also identified five HDL vQTLs in or near CETP in AA, four of which correspond to the HDL vQTLs identified in EA. Given that vQTL can be the result of GxG interactions, subsequent GxG screening involving identified vQTLs was performed. Significant GxG interactions were identified between the cholesterol vQTL and twelve single nucleotide polymorphisms (SNPs) from eleven unique locations in EA. In addition, we also identified significant interactions between two of the HDL vQTLs and four SNPs from three different regions in EA, and four pairs of interactions involving HDL vQTLs in AA. We then used the Framingham Heart Study (FHS) data to replicate the vQTLs and significant GxG interactions identified in EA using ARIC data. Mixed effect models incorporating familial correlation were used to analyze the FHS data. Three HDL vQTLs in CETP, rs1800775, rs7499892, and rs1532624, were confirmed using FHS data. One interaction between HDL vQTL rs9989419 and rs11721627 was replicated. Our analyses have shown that CETP genotype is associated with phenotypic variability of HDL, in addition to its well-known association with mean heterogeneity of HDL. We have also demonstrated the candidate role of vQTLs in GxG study.

588F

Genome-wide Association Studies Meta-analysis on Long Term Average Blood Pressure among Asian Population. *C. Li[1], Y.K. Kim[3], R. Dorajoo[4], H. Li[5], I-T. Lee[10], C-Y. Cheng[12,13,14], M. He[9], W.H Sheu[15], X. Guo[15], J. He[1,2], J.E. Hixson[7], D.C. Rao[6], J. Chen[2], J. Zhao[1], D-F. Gu[8], E-S. Tai[4], X. Lin[5], J-Y. Wu[11], T.Y. Wong[12,13,14], T. Wu[9], Y.I. Chen[15], J.I. Rotter[15,16], B-J. Kim[3], S.K. Ganesh[17], A. Chakravarti[18], T.N. Kelly[1], AGEN consortium.* 1) Department of Epidemiology, Tulane University School of Public Health and Tropical Medicine, New Orleans, LA; 2) Department of Medicine, Tulane University School of Medicine, New Orleans, LA; 3) Division of Structural and Functional Genomics, Center for Genome Science, Korea National Institute of Health, KCDC, Korea; 4) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 5) Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences and Graduate University of the Chinese Academy of Sciences; 6) Division of Biostatistics, Washington University School of Medicine, St. Louis, MO; 7) Department of Epidemiology, Human Genetics and Environmental Sciences, University of Texas School of Public Health, Houston, TX; 8) State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center of Cardiovascular Diseases, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100037, China; 9) MOE Key Lab of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science & Technology, Wuhan 430030, Hubei, China; 10) Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan; National Yang-Ming University Hospital, Taipei, Taiwan; 11) National Center for Genome Medicine, Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; School of Chinese Medicine, China Medical University Hospital, Taichung, Taiwan; 12) Singapore Eye Research Institute, Singapore National Eye Center, Singapore 168751, Singapore; 13) Duke-NUS Graduate Medical School, Singapore; 14) Department of Ophthalmology, National University of Singapore, Singapore; 15) Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center; 16) Division of Genomic Outcomes, Departments of Pediatrics and Medicine, Harbor-UCLA Medical Center; 17) Cardiovascular Medicine, Department of Internal Medicine Department of Human Genetics University of Michigan; 18) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

Objectives: We aimed to identify novel genomic mechanisms underlying blood pressure (BP) regulation by conducting genome-wide single marker and gene-based meta-analyses of the novel long term average (LTA) BP phenotype.**Methods and Results:** A total of 18,422 East Asian participants were included in the discovery stage meta-analyses (stage-1). Promising findings were tested for replication among 46,629 participants of European ancestry (stage-2). Significant SNPs and genes were determined by a P<5.0×10-8 and 5.0×10-6, respectively, in joint analysis of stage-1 and stage-2 data. For 4 loci reported by the previous LTA BP GWAS meta-analysis in participants of European ancestry, evidence of trans-ethnic replication was determined by consistency in effect direction and a Bonferroni-corrected P<1.67×10-2 (0.05/4). The current study identified one novel *ARL3* variant, rs4919669 at 10q24.32, influencing LTA systolic BP (stage-1 P=5.03×10-8, stage-2 P=8.64×10-3, joint P=2.63×10-8) and mean arterial pressure (stage-1 P=3.59×10-9, stage-2 P=2.35×10-2, joint P=2.64×10-8). Nine associations for 3 previously reported BP loci at *WBP1L*, *NT5C2*, and *ATP2B1* were also identified. Gene-based analysis revealed one novel gene, *KCNJ11,* that was associated with LTA systolic BP (stage-1 P=8.55×10-6, stage-2 P=1.62×10-5, joint P=3.28×10-9) and mean arterial pressure (stage-1 P=9.19×10-7, stage-2 P=9.69×10-5, joint P=2.15×10-9). Fourteen genes (*TMEM180*, *ACTR1A*, *SUFU*, *ARL3*, *SFXN2*, *WBP1L*, *CYP17A1*, *C10orf32*, *C10orf32-ASMT*, *AS3MT*, *CNNM2*, and *NT5C2* at 10q24.32; *ATP2B1* at 12q21.33; and *NCR3LG1* at 11p15.1) implicated by previous GWAS meta-analyses were also identified. Among the loci identified by the previous GWAS meta-analysis of LTA BP, we trans-ethnically replicated associations of the *KCNK3* marker rs1275988 at 2p23.3 with LTA systolic BP and mean arterial pressure (P=1.27×10-4 and 3.30×10-4, respectively).**Conclusion:** The current study identified 1 novel variant and 1 novel gene for LTA BP. Furthermore, we present the first evidence of relevance of the *KCNK3* locus for LTA BP in an East Asian sample.

589W

***PIK3CG* rQTL for blood pressure (SBP & DBP) modifies CHD/LDL and CHD/DBP relationships among other rQTL and GxG found in a blood pressure rQTL screen.** *T.J. Maxwell[1], Y. Ma[2], Y. Cao[2], C.M. Ballantyne[3], J.M. Cheverud[4], C.S. Guild[5], C.E. Ndumele[6], E. Boerwinkle[2], P. Wei[2].* 1) Computational Biology Institute, The George Washington University, Ashburn, VA; 2) Human Genetics Center, University of Texas Health Science Center, Houston, TX; 3) Section of Atherosclerosis and Vascular Medicine, Baylor College of Medicine, and Center for Cardiovascular Disease Prevention, Methodist DeBakey Heart and Vascular Center, Houston, TX; 4) Department of Biology, Loyola University, Chicago, IL; 5) Department of Medicine, University of Mississippi Medical Center,Jackson, MS; 6) Division of Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD.

A screen for relationship loci (rQTL) was performed between systolic (SBP) and diastolic blood pressure (DBP) separately for European (EA) and African Americans (AA) with genome-wide genotype data in the Atherosclerosis Risk in Communities (ARIC) study. An rQTL exists when the correlation between multiple traits varies by genotype. Differential gene-by-gene (GxG) and/or gene-by-environment (GxE) interactions are often responsible for rQTL. Screening for this single-locus pattern is a powerful approach to identifying loci with context-dependent effects. Parametric bootstrapping was used to identify 37 genome-wide significant rQTLs from 10 unique locations in EA. Subsequent screens for GxG using whole-exome SNP genotype data identified significant loci for DBP from 15 unique locations interacting with 5 of the rQTL in EA. We used the Framingham Heart Study (FHS) for replication. Twenty EA rQTL from 4 locations replicated in FHS using an analogous mixed effect model incorporating familial relationships. One EA GxG interaction replicated in FHS while another replicated in ARIC AA. One EA rQTL is from a cluster of correlated SNPs intronic to and 3' of *PIK3CG*. It was identified because it affects the relationship between SBP and DBP but it also is nominally significant as an rQTL for coronary heart disease (CHD) risk and DBP, and for CHD risk with both LDL and total cholesterol. *PIK3CG* has been extensively studied and is related to cardiac contractility, modified LDL uptake by macrophages, foam cell formation and blood pressure related measures. Our *PIK3CG* SNPs are separate from other loci 5' of *PIK3CG* reported in other GWAS studies. One EA rQTL (rs10520840) was involved in eleven significant GxG interactions from nine unique locations, one of which (rs7188697) was previously reported as a large hit for QT interval. We found 17 significant rQTLs from 14 locations in AA with 8 of the rQTLs interacting with SNPs from 13 locations with all but one for SBP. Currently we do not have an appropriate replication sample for the AA results. However, we did have 3 AA rQTL and one GxG replicated in FHS and one rQTL replicated in ARIC EA. One of these rQTL is a nominal rQTL for CHD and SBP and another is directly associated with CHD. Overall, there are numerous loci that affect the relationship between SBP and DBP, some due to interactions with other loci and some subsequently modifying the risk relationship between blood pressure and CHD.

590T

A smoking and diabetes status stratified analysis of peripheral arterial disease (PAD) identifies new loci and variants that my interact with these risk factors to modify the risk of PAD. *N.R. van Zuydam[1,2], M. de Andrade[3], E. Vlachopoulou[4], E. Ahlqvist[5], E. Dahlström[6], V. Salomaa[7], C.N.A. Palmer[2], J.C. Hopewell[8], H.M. Colhoun[2], I.J. Kullo[3], GoLEAD and SUMMIT.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Medical Research Institute, University of Dundee, UK; 3) College of Medicine, Mayo Clinic, Rochester MN, USA; 4) University of Helsinki, Finland; 5) Lund University, Malmö, Sweden; 6) Folkhälsan Institute of Genetics, Folkhälsan Research Center, Biomedicum Helsinki, Helsinki University Central Hospital, Helsinki, Finland; 7) National Institute for Health and Welfare, Helsinki, Finland; 8) CTSU, Nuffield Department of Population Health, University of Oxford, Oxford, UK.

Peripheral arterial disease (PAD) affects approximately 200 million people worldwide. Smoking and diabetes are the largest risk factors for PAD. There are three established loci for PAD: genome wide significant (p<5x10-8) signals near *CHRNA3* and *LPA,* and another highly significant signal (p<5x10-7) in the 9p21.3 region. This study sought to expand the number of PAD associated regardless of risk factor profile and to identify variants that are specifically associated with PAD dependent on smoking and diabetes status.PAD was defined as an ankle brachial pressure index <0.9, mid-thigh to mid-foot amputations, corrective procedures related to PAD, dispensed medication to treat claudication or claudication with evidence of vascular disease. We were able to analyse a total of 6,685 PAD cases and 39,324 PAD free controls. We stratified these subjects by diabetes status (2,621 cases/14,579 controls with diabetes and 2,827 cases /21,988 controls without diabetes) and by smoking status (3,964 cases/22,597 controls from ever smokers and 1,083 cases /11,922 controls from never smokers). We performed a fixed-effects variance weighted meta-analysis and variants were tested for heterogeneity between strata. We detected genome-wide significant associations (p<5x10-8) with PAD in all individuals for known variants: rs10757272 (OR[95%CI]=1.17[1.12-1.22], p=5x10-9), in the 9p21.3 region, and rs1051730 (OR[95%CI]=1.13[1.09-1.18], p=1x10-9), near *CHRNA3.* Rs2036527 (OR[95%CI]=1.28[1.21-1.35], p=1x10-11), R2=0.9 with rs1051730), near *CHRNA3,* was associated with PAD in non-diabetic subjects and showed evidence of interaction with diabetes status (P het=5x10-7, in subjects with diabetes: OR[95%CI]=1.00[0.94-1.08], p=NS). We identified a novel association for rs10894808 (OR[95%CI]=1.55[1.41-1.69], p=9x10-9), near *B3GAT1*, in never smokers (P het=5x10-8, in ever smokers: OR[95%CI]=0.95[0.87-1.04], p=NS).In this study, presenting the largest GWAS of PAD so far, we describe one novel locus for PAD based on stratification for known risk factors. Two loci, detected at genome wide significance, showed strong evidence of interaction with either diabetes or smoking status. These results suggest that there may be different pathways contributing to risk of PAD dependent on smoking and diabetes status. .

591F

GDF-15 gene variants influence GDF-15 levels, an independent prognostic marker for cardiovascular events. *N. Eriksson[1], D. Ardissino[3], A. Budaj[4], E. Hagström[1,2], C. Held[1,2], W. Koenig[5,6,7], A. Siegbahn[1,8], P.G. Steg[9,10,11,12], R.A.H. Stewart[13], D. Waterworth[14], H.D. White[13], L. Wallentin[1, 2] on behalf of the STABILITY Investigators.* 1) Uppsala Clinical Research Center, UUppsala Clinical Research Center, Uppsala University, Uppsala, Sweden. ppsala, Uppsala, Sweden; 2) Department of Medical Sciences, Cardiology, Uppsala University, Uppsala, Sweden; 3) Azienda Ospedaliero-Universitaria di Parma, Parma, Italy; 4) Postgraduate Medical School, Grochowski Hospital, Warsaw, Poland; 5) Department of Internal Medicine II-Cardiology, University of Ulm Medical Center, Ulm, Germany; 6) Deutsches Herzzentrum München, Technische Universitàt München, Munich, Germany; 7) DZHK (German Centre for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany; 8) Department of Medical Sciences, Clinical Chemistry, Uppsala University, Uppsala, Sweden; 9) INSERM-Unité 1148, Paris, France; 10) Assistance Publique-Hôpitaux de Paris, Département Hospitalo-Universitaire FIRE, Hôpital Bichat, Paris, France; 11) Università Paris-Diderot, Sorbonne-Paris Cité, Paris, France; 12) NHLI Imperial College, ICMS, Royal Brompton Hospital, London, UK; 13) Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland, New Zealand; 14) Genetics, GlaxoSmithKline, King of Prussia, PA, USA.

Background: Growth differentiation factor 15 (GDF-15) is an emerging biomarker in the field of cardiovascular diseases. It is an independent prognostic biomarker for cardiovascular events and major bleeding during antithrombotic treatment. Whether there is a causal association between GDF-15 and outcomes is still unknown. Genetic markers located in the *GDF15* gene are associated with altered circulating GDF-15 concentrations. In this genome-wide association study (GWAS) we assess the genetic effects on GDF-15 in a large, global, clinical trial on patients with stable coronary heart disease (CHD).**Methods:** We performed a GWAS on 10,788 stable CHD patients enrolled in the Stabilization of Atherosclerotic Plaque by Initiation of Darapladib Therapy trial (STABILITY, ClinicalTrials.govnumber, NCT00799903). Genotype data was generated on the Illumina OmniExpressExome array. In total 10,786 patients and 733,565 SNPs passed genotype quality control of which 10,462 patients had GDF-15 measurements at baseline. GWAS analyses of GDF-15 were performed using PLINK, adjusted for established cardiovascular risk factors and genetic principal components. Analyses were performed sequentially adjusting for each genome-wide significant variant until no signals were left.**Results:** Overall the median (interquartile range) level of GDF-15 was 1248 (909 – 1814) ng/L. We identified three independent genome-wide significant SNPs associated with the GDF-15 level: rs17725099, rs1227731 and rs12459782 (lowest P-value = 3.06E-158) with minor allele frequencies of 0.25, 0.14 and 0.33, respectively. The SNPs were located close to or within the *GDF15* gene on chromosome 19 and were in predicted regulatory regions. The top hit SNP, rs17725099, explained 4.7% of the variance in GDF-15 level with an increase in median GDF-15 level between major and minor genotype of 44% (from 1128 (G/G) to 1626 (A/A) ng/L), SNPs combined explained 6.9% of the variance. Further, rs17725099 showed similar effect (beta estimate) across all self-reported ethnicity groups except black (n=118) and the effect remained when stratifying by country (n=28).**Conclusion:** We identified three independent SNPs affecting GDF-15 concentrations. The SNPs explained 6.6% of the variance in GDF-15 and the top hit SNP had similar effect on GDF-15 across all countries. The results indicate that these SNPs may be suitable for Mendelian randomization analyses in the search for a possible causal effect of GDF-15 on clinical outcomes.

592W

Population-specific genomics identifies predictor of venous thromboembolism in African Americans. *W. Hernandez[1], E.R. Gamazon[2], E. Smithberger[1], T.J. O'Brien[3], A.F. Harralson[3], M. Tuck[4], A. Barbour[3], R.A. Kittles[5], L.H. Cavallari[6], M.A. Perera[1].* 1) Dept Med, Section of Genetic Medicine, The University of Chicago, Chicago, IL; 2) Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN; 3) The George Washington University, Department of Pharmacology and Physiology Washington, DC; 4) Uniformed Services University of the Health Sciences, Department of Veterans Affairs, Washington, DC; 5) Division of Urology, Department of Surgery, University of Arizona College of Medicine, Tucson, AZ; 6) University of Florida, Department of Pharmacotherapy and Translational Research, Gainesville, FL.

Background: Venous thromboembolism (VTE) is the third most common life-threatening cardiovascular condition in the U.S. with African Americans (AAs) having a 30 - 60% higher incidence compared to other ethnicities. The mechanisms underlying population differences in the risk VTE are poorly understood. Genome-wide association studies (GWAS) in Caucasians have confirmed the well-established risk variants, Factor V Leiden (rs6025) and prothrombin G20210A (rs1799963). However, these variants are nearly absent in AAs.Methods: To more comprehensively account for genetic variation observed in VTE risk among AAs, we conducted the firstGWAS on AAs comprised of 578 subjects (cases=146, controls=432) followed by replication of highly significant findings in an independent cohort of 159 AA subjects (cases=94, controls=65).GWAS genotypes were imputed to approximately 10 million SNPs using reference files from the 1,000 genomes haplotypes - Phase I integrated variant set. Logistic regression was used to estimate associations between genetic variants and risk of VTE.We utilized ioinformatics analysis of the top signalsto identifiy expression quantitative trait loci (eQTLs) in whole blood and we investigate the mRNA expression differences between VTE cases and controls.Results:Genome-wide significant associations were identifiedon chromosome 7 (rs73692310, p=1.73x10-9) and chromosome 18 (rs28496996 and rs58952918, p<1.1x10-8) and replicated SNP associations on chromosome 20 (rs2144940 and rs1998081, meta-analysis p<5x10-8) which increased VTE risk by ~2.20 fold. These risk variants were found at higher frequency among AAs (>20%) compared to other ethnic groups (<8%). We demonstrate that chromosome 20 SNPs are *cis*-eQTLs for thrombomodulin (THBD) – a key player in the regulation of coagulation, and also show the expression of THBD is lower among VTE cases compared to controls (p=9.87x10-6). The direction of these associations supports the role of decreased TMBD in VTE risk.Conclusion: We have identified and replicated common novel polymorphisms associated with increased risk of VTE in two independent AA cohorts. In addition, our study provides further evidence that the well-established risk variants among Caucasians (rs6025, prothrombin G20210A, and ABO variants) are not the major cause of VTE in AAs. Our findings provide new molecular insight into a mechanism regulating VTE susceptibility in AAs; a population disproportionately affected by this disease.

593T

Genome-wide Association Study with 1000 Genomes Imputation Identifies 7 new loci Associated with Blood-Pressure Traits in African-American Population. *J. Liang[1], D. Edwards[2], B. Tayo[3], E. Fox[4], Y. Lu[5], R. Jensen[6], J. Smith[7], X. Guo[8], H. Tang[9], G. Chen[10], L. Yanek[11], A. Morrison[12], S. Tajuddin[13], A. Johnson[14], K. Liu[15], R. Charles[16], M. Fornage[12], S. Kardia[7], K. Rice[6], S. Redline[17], J. Rotter[18], C. Richard[3], D. Levy[14], B. Keating[19], A. Chakravarti[11], T. Edwards[2], X. Zhu[1], N. Franceschini[20], COGENT-BP investigators.* 1) Case Western Reserve University, Cleveland, OH; 2) Vanderbilt University, Nashville, Tennessee; 3) Loyola University Chicago, Chicago, IL; 4) University of Mississippi Medical Center, Jackson, MS; 5) Mount Sinai Hospital, New York, NY; 6) University of Washington, Seattle, WA; 7) University of Michigan, Ann Arbor, MI; 8) Cedar-Sinai Medical Center, Los Angeles, CA; 9) Stanford University, Stanford, CA; 10) National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland; 11) John Hopkins University, Baltimore, Maryland; 12) The University of Texas Health Center, Houston, TX; 13) National Institutes of Health, National Institutes of Aging, Bethesda, Maryland; 14) National Institutes of Health, National Heart, Lung and Blood Institute; 15) Northwestern University Feinberg School of Medicine, Evanston, IL; 16) National Human Genome Research Institute, Bethesda, Maryland; 17) Harvard University, Harvard Medical School, Cambridge, MA; 18) University of California, Los Angeles, David Geffen School of Medicine, Los Angeles, CA; 19) University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA; 20) University of North Carolina, Chapel Hill, NC.

African-Americans are more susceptible to high blood pressure (BP) and complications associated with high BP than any other ethnic group in the United States. Many large BP genome-wide association studies (GWASs) have been reported with common variants associated with BP in different ethnic populations, although the BP variation accounted for by the associated variants is still modest. The published GWASs are less successful in African ancestry populations than other populations, possibly due to relatively small sample size and different linkage disequilibrium patterns. We reported a GWAS using genotypes imputed from the 1000G Phase I Integrated Release Version 3 or recent including 30,086 individuals for 17 African American and one Nigerian cohorts. We observed nine independent loci (*BC041459, HBG2, RUFY4 (CXCR2P1), NDST4/TRAM1L1, GPR20, PTH2R/MAP2, ZNF621/CTNNB1, EVX1/HOXA, CDH17*) significantly associated with one or multiple BP traits (systolic BP, diastolic BP, pulse pressure and hypertension status) ($p < 5.0 \times 10\text{-}8$). Two of them have been reported previously (*EVX1/HOXA, CDH17)*. Among the nine index SNPs, two were also studied in our previous report (Franceschini et al. AJHG, 2013) but with less significance. Fine mapping, functional follow-up analysis and replication analysis are ongoing to further confirm the findings. Our study strongly suggests that searching for genetic variants underlying complex traits in African ancestry populations can be fruitful when variants are imputed to the 1000 Genomes reference panel.

594F

Identification of genetic loci associated with heart rate variability by the V$_g$HRV consortium. *I.M. Nolte[1], L.M. Muñoz[1], V. Tragante[2], R. Jansen[3], A. Vaez[1], . MOPMAP Investigators[4], K.F. Kerr[5], . LifeLines Cohort Study[6], S. Kääb[7], N. Verweij[8], T. Lehtimäki[9], B.P. Krijthe[10], D. Levy[11], M.R. Jarvelin[12], A. Minassian[13], A. Britton[14], A.J. Oldehinkel[15], A. Kluttig[16], D.L. Kuh[17], E. Ingelsson[18], P.K. Stein[19], H. Huikuri[20], H. Tiemeier[21], C.M. Nievergelt[13], K. Stolarz-Skrzypek[22], J.F Thayer[23], H. Riese[15], B.J.J.M. Brundel[24], H. Snieder[1], E.J.C. de Geus[25].* 1) Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, Netherlands; 2) Department of Cardiology, University Medical Center Utrecht, Utrecht, The Netherlands; 3) Department of Psychiatry, VU University Amsterdam, Amsterdam, The Netherlands; 4) Departments of Epidemiology and Medicine, University of North Carolina, Chapel Hill, NC, USA; 5) Department of Biostatistics, University of Washington, Seattle, WA, USA; 6) LifeLines Cohort Study, University Medical Center Groningen, Groningen, The Netherlands; 7) Department of Medicine I, University Hospital Munich, Ludwig-Maximilians-University Munich, Munich, Germany; 8) Department of Cardiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; 9) Department of Clinical Chemistry, Fimlab Laboratories and University of Tampere School of Medicine, Tampere, Finland; 10) Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 11) National Heart, Lung, and Blood Institute's Framingham Study, Framingham, MA, USA; 12) Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment & Health, School of Public Health, Imperial College London, UK; 13) Department of Psychiatry, University of California San Diego, La Jolla, California, USA; 14) Department of Epidemiology and Public Health, University College London, London, UK; 15) Interdisciplinary Center Psychopathology and Emotion regulation, Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands; 16) Institute of Medical Epidemiology, Biostatistics and Informatics; Martin-Luther-University Halle-Wittenberg, Halle (Saale), Germany; 17) MRC Unit for Lifelong Health and Ageing at University College London, London, UK; 18) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 19) Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA; 20) Medical Research Center of Oulu, University of Oulu, University Hospital of Oulu, Oulu, Finland; 21) Departments of Epidemiology & Child and Adolescent Psychiatry, Erasmus Medical Center, Rotterdam, The Netherlands; 22) First Department of Cardiology, Interventional Electrocardiology and Hypertension, Jagiellonian University Medical College, Krakow, Poland; 23) Department of Psychology, The Ohio State University, Columbus, OH, USA; 24) Departments of Clinical Pharmacology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 25) Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands.

Background: Reduced cardiac vagal control reflected in low heart rate variability (HRV) is associated with greater risk of cardiac morbidity and mortality. Heritability of HRV ranges from 34-53% for the standard deviation of normal-to-normal intervals (SDNN), from 36-71% for the root mean square of successive differences (RMSSD), and from 25-63% for peak-valley respiratory sinus arrhythmia and high frequency power (pvRSA/HF) in the spectral decomposition of heart rate (HR).**Objective:** to identify loci associated with HRV at rest by performing meta-analysis of genome-wide association studies (GWAS).**Design:** The VgHRV consortium performed three two-stage meta-analyses of GWAS in 53,542 individuals of European ancestry for the HRV measures SDNN, RMSSD, and pvRSA/HF. These traits were determined from 10s to12h heart beat recordings, preferably under sitting/supine resting conditions. Independent SNPs that reached genome-wide significance after replication were followed-up in 11,234 Hispanic and 6,899 African American individuals. Post-GWAS analyses included *in silico* sequencing and annotation, eQTL analysis in whole blood, genetic risk score analysis, genetic correlation analysis between HRV and HR, and functional network analysis.**Results:** We identified 11 independent SNPs in eight loci that were significantly associated with at least one of the three HRV measures. These hits together accounted for 0.8-0.9% of the variance in HRV in the general European population. Replication in Hispanic and African American samples confirmed association for seven and six hits, respectively, in six of the loci. Five of the identified SNPs were previously associated with resting HR, but six were unique to HRV including

our top finding (p=4.6x10-46). Two of our hits were either themselves or in high LD with non-synonymous SNPs and three were eQTLs. LD Score regression showed that genetic correlation is almost 1 between the HRV traits and is around -0.6 between HRV and HR. Genes near our hits were broadly related to two categories of significantly enriched biological processes: 1) cellular signaling and response (e.g. to glucagon) and 2) metabolic processes in the cell (e.g. glycosylation).**Conclusion:** The VgHRV consortium identified 11 independent SNPs in eight loci that were genome-wide significantly associated with HRV explaining 0.9% of its variance. Post-GWAS analyses yielded interesting functional clues providing insight into the biology underlying cardiac vagal control.

595W

Genome-wide association analysis of self-reported blood clots in 6,135 research participants identifies 8 loci associated with thrombosis. *M. Sabater-Lleal[1], D.A. Hinds[2], A. Buil[3], A. Martinez-Perez[4], R. Malik[5], L. Folkersen[6], D. Ziemek[1,7], A. Malarstig[1,7], N. Bing[7], J.M. Soria[4], J.C. Souto[4], A. Franco-Cereceda[8], A. Hamsten[1], E.T. Dermitzakis[3], J.Y. Tung[2], METASTROKE.* 1) Cardiovascular Genetics and Genomics, Department of Medicine, Karolinska Institutet, Stockholm, Sweden; 2) 23andMe, Inc; 3) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 4) Hospital de la Santa Creu i Sant Pau, IIB-Sant Pau, Barcelona, Spain; 5) Institute for Stroke and Dementia Research, Klinikum der Universität München, Ludwig-Maximilians Universität, Munich, Germany; 6) Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Denmark; 7) Pfizer Worldwide R&D; 8) Cardiothoracic Surgery Unit, Department of Molecular Medicine and Surgery, Karolinska Institutet, Karolinska University Hospital Solna, Stockholm, Sweden.

Venous thromboembolism (VTE), which includes deep-vein thrombosis (DVT) and pulmonary embolism (PE), is a complex disease determined by well-established environmental factors and genetic risk factors. To add insights into the genetic regulation of thrombotic disease, we conducted a genome-wide association study (GWAS) of self-reported blood clots in 6,135 VTE cases and 252,827 controls of European ancestry belonging to the 23andMe cohort of research participants.Eight loci that exceeded genome-wide significance were revealed. Our study replicated previously known VTE loci near *F5, FGA-FGG, F11, F2, PROCR, FVIII* and *ABO* genes, and the more recently discovered loci near *SLC44A2*, and *TSPAN15*.In addition, we present data showing common genetic risks factors for stroke and coronary artery disease.Analyses of expression profiles and expression quantitative trait loci (eQTL) across different tissues suggested three plausible candidate genes for the chromosome 19 locus, the first thrombosis-related locus that does not harbour seemingly coagulation-related genes. Finally, our resultsprovide evidencethat self-reported data for VTE GWAS yield results comparable with those using clinically diagnosed VTE. This observation opens up the potential for larger meta-analyses, which will enable elucidation of the genetics of thrombotic diseases, and serves as an example for the genetic study of other diseases.

596T

Meta-analysis of exome chip for platelet count and mean platelet volume identifies novel common and rare loci. *J.D. Eicher[1], N. Chami[2], T. Kacprowski[3], A. Nomura[4], L. Yanek[5], N. Faraday[5], P. Auer[6], A. Reiner[7,8], G. Lettre[2], A.D. Johnson[1], Blood Cell Consortium.* 1) Population Sciences Branch, Intramural Research, Framingham Heart Study, NHLBI, NIH, Framingham, MA; 2) Montreal Heart Institute and Università de Montréal, Montréal, Québec, Canada, H1T 1C8; 3) Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics University Medicine and Ernst-Moritz-Arndt University Greifswald Friedrich-Ludwig-Jahn-Str. 15a, 17475 Greifswald, Germany; 4) Center for Human Genetics Research, Massachusetts General Hospital, 185 Cambridge St., Boston, MA 02114; 5) Department of Medicine, Johns Hopkins University 1830 E Monument St, Baltimore, MD 21287; 6) School of Public Health, University of Wisconsin-Milwaukee 1240 N 10th Street Milwaukee, WI 53201; 7) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109; 8) Department of Epidemiology, University of Washington School of Public Health, Seattle, WA 98195.

Platelet production is a tightly controlled developmental process indicative of the important role platelets play in hemostasis and thrombosis. Identifying genetic determinants of platelet count (PLT) and mean platelet volume (MPV), two common clinical blood count measurements, may elucidate mechanistic insights into platelet production, release, clearance, and their possible roles in chronic disease. Previous genome-wide association and exome chip studies have identified both common and rare genetic variation that influences PLT and MPV. Here, using platelet phenotypes and exome chip genotype data collected in 21 different studies, we performed a meta-analysis on PLT and MPV association results in 131,857 and 41,529 individuals, respectively. Individual study exome chip association results, completed using RareMetalWorker or RVTests, were combined with RareMetals for single variant tests and burden testing using Variable Threshold (VT) and SKAT.For PLT, we confirm association of many known rare (e.g., *JAK2*, *TUBB1*, *MPL*, and *SH2B3*) and common (e.g., *CABLES4*, *TMCC2*, *ARHGEF3*, and *SH2B3*) loci (p<2.07x10-7). Similarly, we replicate several known associations between rare (e.g., *TUBB1*, *KARLN*, and *IQGAP2*) and common (e.g., *GCSAML*, *WDR66*, and *JMJD1C*) loci with MPV (p<2.20x10-7). In addition to recapitulating the association of many known loci, we also identify novel loci currently undergoing conditional analyses and replication efforts. These include novel platelet associations of several genes associated with red and white blood cell (e.g. *TMPRSS6*, *ST5*, and *HFE*) traits previously in the literature and in concurrent analyses in the Blood Cell Consortium, which may reflect influences in common hematopoietic or clearance pathways. We also identify novel associations of genes (*PEAR1* and *MRVI1*) previously associated with platelet reactivity as well as several lipids associated genes including *APOH* and *LDLRAP1*. In addition to these novel associations of previously identified hematological or cardiovascular related genes, we find novel association with *ZMIZ2* (p=3.09x10-17), a known factor in androgen receptor mediated transcription and mesodermal development. Our results demonstrate the utility of exome chip data and increased sample size in genetic discovery for quantitative blood cell traits. Future work will seek to confirm the novel loci and examine the importance of novel coding variation and genes on platelet biology using animal and cellular models.

597F

Simvastatin and RORα Ligands Increase AMP-Activated Protein Kinase (*AMPK*) Level *In vitro*. *N. Coban, C. Gulec, O. Yildirim, N. Erginel-Unaltuna.* Genetics, Institute for Experimental Medicine, Istanbul University, Istanbul, Turkey.

Aim: AMP-activated protein kinase (*AMPK*) is a trimeric enzyme comprising a catalytic α-subunit and regulatory β, γ subunits. *AMPK* was first identified as an upstream kinase that phosphorylates and hence inactivates 3-hydroxy-3-methylglutaryl coenzyme A reductase, the key enzyme controlling cholesterol biosynthesis. Cholesterol is known to be one of the natural ligands of nuclear receptor *RORα*, and *RORα*-mediated control of *AMPK* expression level was shown to play role in metabolism in the mice. Since cholesterol looks like to act as a linker in *RORα-AMPK* pathway, we aimed to test whether this cholesterol-based pathway is effective in human macrophages which play crucial role in atherosclerosis.**Methods:** To test the dependence of *AMPK* on cholesterol level and *RORα* aktivity in the macrophages, we measured the expression levels of *AMPK* and it's downstream gene *SREBF2* in THP-1 macrophages treated with simvastatin and/or synthetic *RORα* ligands. **Results:** We observed that simvastatin increased the expression of *AMPK*, and that this increment was partially enhanced by *RORα* ligands. Simvastatin was also found to increase the expression of *AMPK* target gene *SREBF2*. The latter increment, however, was partially reduced by *RORα* ligands. **Conclusion:** *AMPK* has emerged as a new target for the treatment of cardiometabolic diseases. *RORα* has also potential to be used as a target in cardiometabolic diseases. Statins, on the other hand, have been used in the treatment of cardiovascular diseases for decades. In recent study, we aimed to test the interrelationship between *AMPK* and *RORα* in macrophages. Our results suggest that cholesterol might play role in regulation of *RORα-AMPK* pathway in human macrophages, and this regulation can be modulated by administration of simvastatin and synthetic *RORα* ligands. These results offer *RORα-AMPK* pathway as a potential therapeutic target not only for metabolic disorders, but also for atherosclerosis.

598W

Better Characterization of Coronary Artery Disease Across Multiple Tissues Achieved through Improvement of Bayesian Networks Using Causal Inference Testing. *A. Cohain[1], O. Franzen[1], R. Ermel[2], Y. Mo[1], H. Talukdar[3], H. Foroughi[3], J. Skogsberg[3], K. Hao[1], C. Betsholtz[5], J. Kovacic[6], T. Michoel[4], A. Ruusalepp[2,7], E.E. Schadt[1,7], J.L.M. Björkegren[1,2,6,7].* 1) Department of Genetics & Genomic Sciences, Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Cardiac Surgery, Tartu University Hospital, Tartu, Estonia; 3) Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Stockholm, Sweden; 4) Division of Genetics and Genomics, The Roslin Institute, University of Edinburgh, Edinburgh, UK; 5) Deptartment of. Immunology, Genetics and Pathology, Uppsala University, Sweden; 6) Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York, NY; 7) Clincial Gene Networks AB, Stockholm, Sweden.

Given the huge societal impact of Coronary Artery Disease (CAD) around the world, extensive genome-wide association studies (GWAS) have been pursued to elucidate the complexity of this highly heritable disease, resulting in the identification of >150 loci that collectively explain <10% of the expected genetic variance. The genetic architecture and downstream mechanisms of CAD required to translate genetic loci into opportunities for diagnosis, therapy and prevention, have been resolved for only a few of these disease-associated loci. To improve our understanding of CAD in light of these genetic findings, the Stockholm-Tartu Atherosclerosis Reverse Network Engineering Team (STARNET) sequenced RNA from 9 disease-relevant tissues, genotyped DNA, and collected extensive phenotypic information from a total of 700 patients who either underwent coronary artery bypass graft surgery ("cases") or other forms of open heart surgery ("controls").To move beyond viewing the complexity of the human system and disease as a combination of isolated genetic loci and genes, we carried out a multi-tissue systems approach to integrating the extensive STARNET dataset to better define the molecular underpinnings of CAD. First, we detected causal interactions among gene expression traits and between these traits and clinical phenotypes by employing causal inference techniques that leverage DNA variation common to these traits as a perturbation source. Systematic application of this causal inference procedure among all gene expression traits and clinical features across all CAD loci reveals the complex architecture of both upstream and downstream effects of these disease associated SNPs. Using the results of the causal relationships predicted across all tissues in STARNET as structure priors, we constructed Bayesian networks that provide a highly integrated, directed, multi-tissue view of disease that enables data-driven predictions relating to the pathophysiology of CAD. We believe these disease networks are the most comprehensive and informative representations of the regulatory and molecular landscape of CAD to date and as such should provide significant insights into novel therapeutic points of intervention and novel therapies for CAD.

599T

From Death to Life/Back to the Future: Detailed premorbid clinical and family history can save the lifes and adress the final diagnosis in Sudden Unexplained Deaths with negative autopsy. *S.G. Temel[1,2], C.L Baydar[3], B. Turkgenc[4,5], M. Ozen[6], S. Usar Incirli[7], C. Conkbayir[8], M.C Yakicier[4,9].* 1) Near East University, Faculty of Medicine, Department of Histology & Embryology, Nicosia, North Cyprus; 2) Uludag University, Faculty of Medicine, Department of Medical Genetics, Bursa, Turkey; 3) Near East University, Faculty of Medicine, Departmen of Forensic Medicine, Nicosia, North Cyprus; 4) Acibadem Genetic Diagnostic Center, Istanbul, Turkey; 5) 2. Marmara University, Faculty of Medicine, Department of Medical Biology and Genetic, Istanbul, Turkey; 6) Dr. Nalbantoglu State Hospital, Department of Pathology, Nicosia, North Cyprus; 7) Dr. Nalbantoglu State Hospital, Department of Neurology, Nicosia, North Cyprus; 8) Near East University, Faculty of Medicine, Department of Cardiology, Nicosia, North Cyprus; 9) Acibadem University, Faculty of Science, Department of Molecular Biology and Genetic, Istanbul, Turkey.

Sudden cardiac death is responsible for a large proportion of sudden deaths in young individuals. In forensic medicine, many cases remain unexplained after routine postmortem autopsy and conventional investigations. These cases are called sudden unexplained deaths (SUD). İnherited cardiac disorders such as Long Qt, Brugada syndrome and cardiomyopaties comprise a substantial proportion of SUD cases. In this study, two forensic SUD case and their family members are investigated for inherited cardiac disorders. We report a 35 and 31 years old unrelated asymptomatic forensic SUD males died suddenly during asleep. There was another SUD individual in the first case's family and second case had a syncope history. Postmortem examinations excluded the extracardiac causes of sudden death as well as the absence of structural abnormalities. Because we had no available material for the first case, detailed cardiologic examination was performed to the sibling of him. She was positive for ajmaline test and diagnosed as Brugada syndrome; intracardiac defibrillator was planned and implanted. Targeted sequencing was performed to the sibling of the first and the second forensic SUD case on ion Personel Genome Machine using Ion Ampliseq Panel comprising 68 genes known to be associated with cardiological arrhythmias. Torrent Suite Software was used for data analysis. Three heterozygote missense variations were identified in the sibling of the first SUD case; c.3611C>T (p.Pro1204Leu) in TRPM4 gene, c.109G>A (p.Val37Ile) in RANGRF gene, paternally inherited and c.11229G>C (p.Met3743Ile) in AKAP9 gene, maternally inherited. Molecular autopsy of the second SUD case was revealed 3 missense variations; c.1889G>A (p.Arg630Gln) in KCND3 gene, maternally inherited and c.5369T>G (p.Val1790Gly) in AKAP9 gene, c.325G>A (p.Val109Ile) in KCNE1 gene, paternally inherited. RANGRF variation was novel, KCNE1 and TRPM4 variations were reported as pathogenic allele and the remaining 2 were reported as "unknown" allele. Carriers were identified in two families, also. For future directions functional analysis of found variations are under estimated. Irrespective of genetic testing; all SUD family required detailed clinical testing to identify relatives who may be at risk. Molecular autopsy and detailed premorbid clinical and family history can survive family members of SUD cases.Acknowledgements: İnherited arrhythmia panel is supported by grant from SANTEZ Project (0253.STZ.2013-2), Turkey.

600F

Multivariate genome-wide association analysis of circulating cellular adhesion protein levels in the Multi-Ethnic Study of Atherosclerosis. *N.B. Larson[1], P.A. Decker[1], C.L. Wassel[2], J.S. Pankow[3], W. Tang[3], N.Q. Hanson[4], M.Y. Tsai[4], S.J. Bielinski[1].* 1) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 2) Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA; 3) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN; 4) Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN.

The cellular adhesion pathway is critical in the pathophysiology of atherosclerosis, and genetic factors contributing to regulation of circulating levels of related proteins may be relevant to risk prediction of cardiovascular disease. Moreover, the regulatory genetics and corresponding effects of the cellular adhesion pathway may be race/ethnicity-specific and could contribute to disparities in disease incidence. In contrast to conducting separate protein quantitative trait loci (pQTL) analyses of each individual protein, joint genetic association analyses of multiple quantitative traits can leverage cross-trait co-variation and identify simultaneous regulatory effects on protein levels across the pathway. In this study, we conducted a multi-pQTL analysis of 15 proteins related to cellular adhesion assayed on n = 2313 from the Multi-Ethnic Study of Atherosclerosis (MESA). Genetic data consisted of the Affymetrix Genome-Wide Human SNP Array 6.0 supplemented by the Illumina Exome BeadChip, Cardio-MetaboChip, and the iSelect ITMAT/Broad/CARe (IBC) Chip. We applied the MQFAM multivariate association analysis method in PLINK on normalized residuals derived from univariate linear regression, adjusting for age, sex, genetic ancestry, BMI, and smoking status. Participants were African (n = 519), Chinese (n = 594), non-Hispanic white (n = 613), and Hispanic American (n = 587), and analyses were conducted stratified by race/ethnicity. Analysis results were also combined across race/ethnicity by meta-analysis using Stouffer's combined Z-score approach. Stratified analyses identified multiple significant (P<5e-08) loci associated with co-variation of protein levels, many of which were in proximity to coding genes of the assayed proteins and varied in statistical significance by strata. However, we additionally identified consistent SNP associations across race/ethnicity in genes *ABO* [MIM 110300] and *FUT2* [MIM 182100], which respectively encode the protein defining the ABO blood group system and a Golgi stack membrane protein involved in the creation of H-antigen precursor.Supplementary analyses suggested the *ABO* multi-pQTL SNP associations to be driven by ABO blood type. These results indicate the biological relevance of blood group antigens on regulation of circulating cellular adhesion pathway proteins while also demonstrating race/ethnicity-specific genetic regulatory effects.

601W

Differential mRNA and miRNA gene expression in hypertensive and non-hypertensive women is influenced by race. *D.F. Dluzen, N.N. Hooten, Y. Zhang, K.D. Jacob, A.B. Zonderman, M.K. Evans.* National Institute on Aging, National Institutes of Health , Baltimore, MD.

Essential hypertension occurs more frequently among African Americans (AA) than any other population sub-group, has earlier onset, and more frequently results in end-organ complications. In addition, AA women have the highest incidence and hospitalization rates for hypertension in the United States. Previous data suggests that gene expression patterns may increase individual susceptibility to selected chronic diseases. Therefore, we hypothesized that differential gene expression may influence the disproportionate incidence and prevalence of hypertension among AAs. Transcriptional profiling of peripheral blood mononuclear cells (PBMCs) from AA or White, normotensive or hypertensive females identified thousands of mRNAs differentially expressed by race and/or presence of hypertension. Since microRNAs (miRNAs) are well-known post-transcriptional regulators of mRNA expression levels, and recent data indicates that miRNAs play an important role in the etiology of cardiovascular disease, we wanted to examine whether changes in miRNA abundance influence mRNA changes in hypertension. To test this, we profiled global miRNA expression in the same cohort. Analysis of microarray expression changes using Ingenuity Pathway Analysis identified mRNA-miRNA regulatory networks in hypertension-related pathways, i.e. the renin-angiotensin system, nitric oxide signaling, and actin-cytoskeletal signaling pathways, which differ by race and hypertension status. mRNA and miRNA gene expression changes were validated using RT-qPCR in an expanded cohort and gene functionality and miRNA target validation was investigated *in vitro* using primary human umbilical vein cells. We have identified several miRNAs, including miRs *20a-5p, 30c-5p, 4763-5p, 4709-3p* and *4717-3p*, which are significantly ($P<0.05$) and differentially expressed by race and/or presence of hypertension. Together, these findings identify several pathways and miRNA gene candidates whose expression differ with race and may contribute to hypertension-related health disparities.

602T

Discovery of gene networks in monocytes with evidence of mediating the protective effects of HDL-cholesterol on monocytes function. *R.A. Verdugo[1], F.A. Medina[1], K. Orostica[1], T. Zeller[2], P.S. Wild[3,4], T. Munzel[3], K. Lackner[5], E. Ninio[6], D.A. Tregouet[6], F. Cambien[6], S. Blankenberg[2], L. Tiret[6].* 1) Programa de Genética Humana, Facultad de Medicina, Universidad de Chile, Chile; 2) Department of General and Interventional Cardiology, University Heart Center Hamburg; 3) Department of Medicine II, University Medical Center Mainz; 4) Center for Thrombosis and Haemostasis, Clinical Epidemiology, University Medical Center Mainz; 5) Institute for Clinical Chemistry and Laboratory Medicine, University Medical Center Mainz; 6) INSERM UMR_S 1166, Pierre and Marie Curie, Paris, France.

Although high-density lipoprotein (HDL) cholesterol is a known protective factor against atherosclerosis, the molecular mechanisms mediating its effects are not well understood. The objective of this study is to identify gene networks whose expression may explain such protective effect in humans. We tested graphical models of causality on 17,186 genes expressed in monocytes from 1,466 participants of the Gutenberg Health Study. There were 3,596 genes whose expression was associated with either HDL, 258 with number of atherosclerotic plaques, and 129 were associated to both (FDR < 0.1). Independent component analysis (ICA) identified 163 components in 500 runs. Of these, 44 were selected because high stability across runs, were not explained by outliers, and were not correlated with technical variables. Nine independent patterns of expression were significantly associated to HDL cholesterol level in blood. The rows of matrix A from ICA were used as independent expression patterns called metagenes. The structure of causal associations among the expression of nine metagenes, HDL, sex, smoking, and number of plaques were investigated by Bayesian Networks. Graphical models were scored by the BIC score and the space of models was explored by the hill-climbing algorithm. The process was repeated in 1000 bootstraps and the causal structure was represented as the average networks across boot replicates using the bnlearn R-package. The average causal model identified Meta31 (ADRB2 gene in the center) as a parent of HDL in the network, and therefore it's a causal candidate to modify cholesterol uptake or efflux from monocytes. By contrast, Meta29 and Meta119, with genes PACSIN1 and DEFA4 in their respective centers, were the only children of HDL, suggesting that HDL directly regulates the expression of these modules in the monocyte. Meta31 was the most correlated metagene to HDL cholesterol (r2=0.045) and it had the highest proportion of genes with expression correlated to this lipoprotein (50%) and to number of atherosclerotic plaques (10%). The gene module associated to this pattern contained 340 genes that were enriched for the "response to wounding" and "immune response" GO terms. Meta29 contained 258 genes that were enriched for the "response to endoplasmic reticulum stress" GO term and code mostly "integral to membrane" proteins. We investigate gene-level models of causality that can be tested experimentally.

603F

Extensive Transcriptional Changes in Hypertrophied Cardiomyocytes Derived from Human iPSCs. *W. Li[1], P. Aggarwal[1], A. Turner[1], A. Matter[1], E. Storvick[1], D. Arnett[2], S. Hunt[3], B. Lewis[2], U. Broeckel[1].* 1) Dept. of Pediatrics, Medical College of Wisconsin, Milwaukee, Wisconsin, US; 2) Dept. of Epidemiology, University of Alabama at Birmingham, Birmingham, Alabama, US; 3) Cardiovascular Genetics Division, School of Medicine, University of Utah, Salt Lake City, Utah, US.

Rationale. Left ventricular hypertrophy (LVH) is a significant risk factor for cardiovascular disease. Cardiac hypertrophy is accompanied by significant gene expression changes in cardiomyocytes (CMs). However, our knowledge on gene expression patterns is largely focused on the net transcript abundance level. So far, we have limited knowledge about the level of alternative splicing (AS) and associated isoform expression changes. To gain insight into the transcriptional changes associated with cardiac hypertrophy, we perform RNA-sequencing on CMs with or without hypertrophy stimulation. We identify splicing patterns and differential isoform expression unique to CM hypertrophy. **Methods.** We developed a model of human CM hypertrophy using induced, pluripotent stem cell derived CMs (hiPSC-CMs). Four lines of CMs were included in our study. Stimulating the hiPSC-CMs with Endothelin1 induced characteristic changes consistent with the development of LVH, *e.g.*, an increase in cell surface area and elevated expression of well-established LVH markers. Transcripts were assembled using STAR and Bowtie2. Transcripts abundance was quantified using Cufflinks2. Splice sites with at least 10 uniquely mapped reads were pursued for further analysis. **Results.** Upon hypertrophic induction, 446 to 982 genes show differential expression at isoform level. A significant portion of the differentially expressed isoforms (DES) is specific to each cell line. Distinct functional motifs are found in these unique DESs. For each CM line, between 160 and 1131 genes exhibit altered splicing sites (SSs) specific to hypertrophic CMs. 11% to 25% of these SSs produce novel exons that are concentrated in highly expressed genes. Compared to annotated exons, a dominant portion (42%-72%) of the novel exons are created through alternative 5' splicing. Interestingly, AS and differential isoform expression are enriched in genes involved in hypertrophic cardiomyopathy and cardiac contraction. **Conclusions.** We reveal extensive transcriptional changes upon hypertrophy induction in our cellular model of LVH. Compared to normal CMs, hypertrophied CMs have a distinct profile of AS and isoform expression patterns. Our study suggest that alterations at both splicing and isoform-expression levels are an important regulatory component of cardiac hypertrophy. Further, unique SSs and isoform expression may serve as biomarker candidates for future investigations.

604W

Increased mutational burden is associated with more severe presentation in left ventricular noncompaction. *M. Bainbridge[1], K. Miszalski-Jamka[2], W. Mazur[3], J. Jefferies[4], S. Jhangiani[1], D. Muzny[1], R. Gibbs[1].* 1) Human Genome Sequencing Center, Baylor College Med, Houston, TX; 2) Department of Cardiology, Congenital Heart Disease and Electrotherapy, Silesian Center for Heart Disease, Zabrze, Poland; 3) Ohio Heart and Vascular Center Cincinnati, OH; 4) The Heart Institute, Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

Left ventricular non-compaction (LVNC) is traditionally understood to be a dominant, monogenic disease that is typified by spongy, non-compacted myocardium, and hypertrabeculation. LVNC has a prevalence of ~1/5000 and is believed to be a compensatory mechanism for impaired myocardial function. We extensively phenotyped and whole exome sequenced 190 subjects with LVNC or hypertrabeculation and identified possible contributory variants (CVs) in 113 subjects (59.4%), in 55 genes that are known to cause LVNC, muscular dystrophy or non-LVNC cardiomyopathy. The most commonly mutated gene was TTN with 17 truncating variants followed by MYH7 with 10. Multiple variants were found in genes not previously associated with LVNC including long-QT genes KCNH2, KCNE1, KCNQ1, KCNJ2, myopathy gene FLNC and coronary artery disease gene MEF2A. Of the 113 subjects with at least one CV, 37 (32.7%) had 2 or more, including a pair of siblings with 4 CVs. We found increased mutational burden in these genes between our LVNC cohort and an unselected control population. We also found strong correlation between mutational burden in individual patients and measured phenotypes of cardiac dysfunction, including left ventricular ejection fraction ($p<0.001$) and non-compacted to compacted myocardium ratio ($p<0.001$). These findings remained even when considering variants with minor allele fractions that would normally be considered too high to cause a rare, dominant, monogenic disease (0.1-1%). Our findings indicate that LVNC may be caused by, and its phenotype exacerbated by, multiple, uncommon variants that act epistatically. These findings have a profound impact in LVNC genetic diagnosis and help to explain the variable penetrance of LVNC seen in families.

605T

Dilated cardiomyopathy-associated long intergenic non-coding RNAs in zebrafish cardiac transcriptome. *L. Wang[1], Y. Shih[2], X. Xu[2], Y. Zhang[1].* 1) Division of Biostatistics and Bioinformatics, University of Maryland Greenebaum Cancer Center, and Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD; 2) Division of Cardiovascular Diseases, Department of Biochemistry & Molecular Biology, Mayo Clinic, Rochester, MN.

The long intergenic non-coding RNAs (lincRNAs) have been found to serve as key regulators of many cellular processes, such as genomic imprinting, chromatin regulation, mRNA regulation and nuclear organization. Because of temporospatial-specific expression property, it is important to investigate lincRNAs in the context of various tissue types. In this study, we aimed to identify heart-specific lincRNAs using a genome-wide transcriptome analysis. Specifically, we focused on lincRNAs that express in zebrafish heart at embryonic and adult stages, an emerging vertebrate model to study cardiomyopathy and heart failure. We identified 120 known lincRNAs from zebrafish heart, including 68 at embryonic stage, 94 at adult stage, and 42 overlapping lincRNAs at both stages. Among these lincRNAs, 55 lincRNAs are cardiac specific without expression in zebrafish muscle. Furthermore, we identified 592 putative novel lincRNAs using our computational bioinformatics pipeline. The co-expression analysis indicated that certain lincRNAs were potentially functionally associated with dilated cardiomyopathy genes in zebrafish. Our studies suggest that lincRNAs play crucial biological roles during heart development and may serve as potential key regulators in biological processes that are closely related to heart disease.

606F

Heterozygous loss-of-function mutations in *DLL4* cause Adams-Oliver Syndrome. *J.A.N. Meester[1], L. Southgate[2], A. Stittrich[3], H. Venselaar[4], S.J.A. Beekmans[5], N. den Hollander[6], E.K. Bijlsma[6], A. Helderman-van den Enden[6], J.B.G.M. Verheij[7], G. Glusman[3], J.C. Roach[3], A. Lehman[8], M.S. Patel[8], B.B.A. de Vries[9], C. Ruivenkamp[6], P. Itin[10], K. Prescott[11], S. Clarke[12], R. Trembath[2], M. Zenker[13], M. Sukalo[13], L. Van Laer[1], B. Loeys[1], W. Wuyts[1].* 1) Centre of Medical Genetics, University of Antwerp, Antwerp University Hospital, Belgium, Antwerp, Belgium; 2) Barts and The London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London, EC1M 6BQ, United Kingdom; 3) Institute for Systems Biology, Seattle, Washington, SA 98109, United States of America; 4) Centre of Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboudumc, Nijmegen, 6525 GA, The Netherlands; 5) Department of plastic and reconstructive surgery, VU Medical Center Amsterdam, Amsterdam, 1081 HZ, The Netherlands; 6) Department of Clinical Genetics, Leiden University Medical Center, Leiden, 2333 ZA, The Netherlands; 7) Department of Medical Genetics, University Medical Centre Groningen, University of Groningen, Groningen, 9700 RB, The Netherlands; 8) Department of Medical Genetics and Child and Family Research Institute, University of British Columbia, Vancouver, BC, V6H 3N1, Canada; 9) Department of Human Genetics and Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, 6500 HB, The Netherlands; 10) University Hospital, Dermatology, Basel, 4031, Switzerland; 11) Department of Clinical Genetics, Chapel Allerton Hospital, Leeds, LS7 4SA, United Kingdom; 12) Department of Dermatology, Chapel Allerton Hospital, Leeds, LS7 4SA, United Kingdom; 13) Institute of Human Genetics, Otto-von-Guericke-Universitàt Magdeburg, University Hospital Magdeburg, Magdeburg, 39106, Germany.

Adams-Oliver syndrome (AOS) is a rare developmental disorder characterized by the presence of both aplasia cutis congenita (ACC) of the scalp vertex and terminal limb reduction defects, such as brachydactyly, oligodactyly, syndactyly, hypoplastic nails or transverse amputations. Cardiovascular anomalies, comprising pulmonary hypertension, ventricular septum defects, tetralogy of Fallot and anomalies of the great arteries and their valves are also frequently observed. Mutations in five genes have been identified as a cause for AOS prior to this report. Mutations in *EOGT* and *DOCK6* cause autosomal recessive AOS, whereas mutations in *ARHGAP31*, *RBPJ* and *NOTCH1* lead to the autosomal dominant form of AOS. As *RBPJ*, *NOTCH1* and *EOGT* are all involved in the Notch signaling cascade, we hypothesized that mutations in other genes involved in this pathway may also be implicated in AOS pathogenesis. Using a candidate gene based approach, we prioritized *DLL4*, a critical Notch ligand, due to its essential role in vascular development and angiogenesis in the context of cardiovascular features in AOS patients. Targeted resequencing of the *DLL4* gene using a custom enrichment panel was performed in 89 independent families, in which we found seven mutations. In addition, a defect in *DLL4* was also detected in two pedigrees with whole exome/genome sequencing. In total, nine heterozygous mutations in *DLL4* were identified, including two nonsense variants, which are predicted to lead to nonsense mediated decay, and seven missense variants. These missense variants encompass four mutations that replace or create cysteine residues, which are likely critical for maintaining the structural integrity of the protein and three mutations that affect conserved amino acids in two functional domains of the protein. These functional domains include the DSL domain and the MNNL domain, which are both involved in binding of the ligand to the Notch receptor. Affected individuals with *DLL4* mutations present with variable clinical expression and incomplete penetrance with no emerging genotype-phenotype correlations. Our findings demonstrate *DLL4* mutations as an additional cause of autosomal dominant AOS or isolated ACC and provide yet further evidence for a key role of Notch signaling in the etiology of this disorder.

607W

Recessive *MYH6* mutations in hypoplastic left heart with reduced ejection fraction. *J. Theis, M. Zimmermann, J. Evans, B. Eckloff, E. Wieben, M. Qureshi, P. O'Leary, T. Olson.* Mayo Clinic, Rochester, MN.

The molecular underpinnings of hypoplastic left heart (HLH) are poorly understood. Staged surgical palliation has dramatically improved survival, yet eventual failure of the systemic right ventricle necessitates cardiac transplantation in a subset of patients. We sought to identify genetic determinants of HLH with latent right ventricular dysfunction in individuals with a Fontan circulation. Evaluation of cardiac structure and function by echocardiography in patients with HLH and their first degree relatives identified five individuals with right ventricular ejection fraction ≤40% following Fontan operation. Whole genome sequencing (WGS) was performed on DNA from 21 family members, filtering for genetic variants with allele frequency <1% predicted to alter protein structure or expression. Secondary family-based filtering for *de novo* and recessive variants revealed rare inherited missense mutations on both paternal and maternal alleles of *MYH6*, encoding myosin heavy chain 6, in two patients who developed right ventricular dysfunction >10 years postoperatively. Parents and siblings who were heterozygous carriers had normal echocardiograms. Protein modeling of the four highly conserved amino acid substitutions, residing in both head and tail domains, predicted perturbation of protein structure and function. In contrast to dominant *MYH6* mutations with variable penetrance identified in other congenital heart defects and dilated cardiomyopathy, this study reveals compound heterozygosity for recessive *MYH6* mutations in patients with HLH and reduced systemic right ventricular ejection fraction. These findings implicate a shared molecular basis for developmental arrest and latent myopathy of left and right ventricles, respectively.

608T

Novel *FLNC* truncation variants found in a large cohort of dilated cardiomyopathy cases. *R.L. Begay[1], S.L. Graw[1], D.B. Slavov[1], F. Brun[2], K.L. Jones[1], K. Gowan[1], D. Miani[2], G. Sinagra[2], L. Mestroni[1], M.R.G. Taylor[1].* 1) Department of Medicine, University of Colorado - CU Cardiovascular Institute, Aurora, CO; 2) University of Trieste Hospital, and S. Maria della Misericordia Hospital, Udine, Italy.

Background –Dilated cardiomyopathy (DCM) is the most common form of cardiomyopathy and frequently leads to heart failure. In cases without an environmental etiology, variants in over 50 genes have been linked to this DCM disease. However, genetic causes are only suspected in about ~30-40% of DCM cases. Therefore, it is crucial to continue the search for novel DCM disease-causing genes to understand this link between genetics and heart disease. One such potential gene is *Filamin-C* (*FLNC*), which encodes for a structurally important actin-binding protein located within the sarcomere of heart and skeletal muscle. Mutations in this gene are known to cause skeletal myopathies, and more recently has been associated with hypertrophic cardiomyopathy in the absence of skeletal muscle pathology. In a previous study (manuscript in preparation), we used Whole Exome Sequencing to discover a *FLNC* splicing mutation in two Italian DCM families and suggested haploinsufficiency of the resulting *FLNC* protein as the pathological mechanism. In this current study, we decided to pursue these findings in a larger DCM population. *Methods and Results* – To gain an understanding of the gene mutation spectrum in patients with DCM, we identified and analyzed 187 DCM affected individuals utilizing Illumina TruSight One Sequencing Panel, which queries 4,813 genes associated with clinical phenotypes. Following a bioinformatic filter, we detected six families with *FLNC* truncation mutations (three stopgain and three splicing), for a frequency of ~3.2%. These individuals tested presented with a DCM phenotype with lack of skeletal muscle abnormalities. In addition, these mutations are absent from the 1000 Genomes Project and NHLBI Exome Sequencing Project, which suggests novelty.*Conclusion* – *FLNC* truncation mutations have been inadequately understood with respect to cardiac muscle. This study provides additional findings that *FLNC* truncation mutations can lead to dilated cardiomyopathies without the presence of skeletal muscle defects. Currently, studies are ongoing to recapitulate the DCM phenotype seen in humans caused by *FLNC* truncations in a zebrafish model. As a result of this model, we can better understand the pathological mechanism of these truncation mutations.

609F

A unique case of multisystemic smooth muscle dysfunction syndrome with expanding genotype-phenotype correlations. *G. Bhat[1], R. Ji[2], W. Zou[2], D. Penny[2], F. Yuxin[2], P. Levy[1].* 1) Division of Medical Genetics, Montefiore Medical center, Albert Einstein College of Medicine, Bronx, NY; 2) John Welsh Cardiovascular Diagnostic Laboratory, Department of Pediatrics, Texas Children's Hospital, Baylor college of Medicine.

ACTA2 located at 10q23.31 is an actin isoform that encodes for vascular smooth muscle alpha-actin (#OMIM 102620). *ACTA2* mutations were initially found to be associated with familial thoracic aortic aneurysm, however in recent years the phenotypic spectrum has become increasingly diverse. Multisystemic smooth muscle dysfunction (MSMD [MIM 613834]) syndrome is on the severe end of the spectrum of *ACTA2* mutations with earlier onset and higher penetrance. Characteristic presentation includes hypoperistalsis, hypotonic bladder, pulmonary hypertension, congenital mydriasis, cerebrovascular abnormalities, stroke and PDA. Mileciwz et al in 2010 identified R179H mutation as causative in a cohort of 7 patients with MSMD syndrome. We report on a 27 year old female with MSMD syndrome caused by a missense R179C mutation in *ACTA2* and expand the genotype-phenotype correlations.Our proposita was born full term to a non-consanguineous couple of Puerto Rican ancestry. She had a complicated newborn course due to recurrent pneumonia, intestinal malrotation requiring surgery and PDA repair in infancy. Congenital mydriasis was noted at birth. She has persistent transient ischemic attack-like episodes first noted at 7 years of age, characterized by hand numbness and mutism. Her MRI of brain and angiogram showed bilateral cortical infarcts, with bilateral supraclinoid carotid artery occlusions, extensive collaterals, encephalomalacia, white matter signal abnormalities and extensive volume loss. CT of the thorax showed lucency of the left lung, emphysematous changes and areas of bronchiectasis. Echocardiogram showed right ventricular dilatation and pulmonary hypertension. The patient has a history of cholelithiasis and a unique finding of small left kidney, not previously described with MSMD syndrome. Sequence analysis identified a heterozygous missense mutation c.535C>T in one copy of the *ACTA2* gene. This nucleotide change predicts an amino acid change from arginine to cysteine (p.R179C). The only other patient with heterozygous p.R179C substitution in *ACTA2* associated with a MSMD phenotype was reported by Meuwissen et al in 2012. The patient had additional findings of malformations of corpus callosum and died at 3 years of age due to complications from PDA repair surgery. In conclusion, our case supports the hypothesis that the p.R179C mutation in *ACTA2* leads to a more extensive involvement with multiple comorbidities as compared to the more common p.R179H mutation.

610W

Genome-wide rare copy number variations contribute to genetic risk for transposition of the great arteries. *G. Costain[1,2], L. Ogura[1], C.R. Marshall[3], S.W. Scherer[3,4], C.K. Silversides[5], A.S. Bassett[1,5,6,7,8,9].* 1) Clinical Genetics Research Program, CAMH, Toronto, Ontario, Canada; 2) Residency Training Program in Medical Genetics, University of Toronto, Toronto, Ontario, Canada; 3) The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) McLaughlin Centre and Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 5) Division of Cardiology, Department of Medicine, University Health Network, Toronto, Ontario, Canada; 6) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 7) Department of Psychiatry, and Toronto General Research Institute, University Health Network, Toronto, Ontario, Canada; 8) Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 9) The Dalglish Family Hearts and Minds Clinic for 22q11.2 Deletion Syndrome, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada.

Background: Transposition of the great arteries (TGA) is an uncommon but severe congenital heart malformation of unknown aetiology. Rare copy number variations (CNVs) have been implicated in other, more common conotruncal heart defects like tetralogy of Fallot (TOF), but there are as yet no CNV studies dedicated to TGA. **Methods:** Using high-resolution genome-wide microarrays and rigorous methods, we investigated CNV in a group of prospectively recruited adults with TGA (n=102) from a single centre. We compared rare CNV burden to well-matched cohorts of controls and TOF cases, adjudicating rarity using 10,113 independent population-based controls and excluding all subjects with 22q11.2 deletions. We identified candidate genes for TGA based on rare CNVs that overlapped the same gene in unrelated individuals, and pre-existing evidence suggesting a role in cardiac development.**Results:** The TGA group was significantly enriched for large rare CNVs (2.3-fold increase, p=0.04) relative to controls, to a degree comparable with the TOF group. Extra-cardiac features were not reliable predictors of rare CNV burden. Smaller rare CNVs helped to narrow critical regions for conotruncal defects at chromosomes 10q26 and 13q13. Established and novel candidate susceptibility genes identified included ACKR3, IFT57, ITGB8, KL, NF1, RERE, SLC8A1, SOX18, and ULK1.**Conclusions:** These data demonstrate a genome-wide role for rare CNVs in genetic risk for TGA. The findings provide further support for a genetically-related spectrum of congenital heart disease that includes TGA and TOF.

611T

Analysis of rare variants and CNVs in non-syndromic tetralogy of Fallot. *J. Goodship[1], M. Miossec[1], D. Brown[1], I. Wilson[1], L. Sutcliffe[1], A. Topf[1], K. Devriendt[2], A. Rauch[3], D. Winlaw[4], F. Bu'Lock[5], S. Bhattacharya[6], M. Lathrop[7], B. Keavney[8], M. Santibanez-Koref[1].* 1) Inst Genetic Medicine, Univ Newcastle upon Tyne, Newcastle on Tyne, United Kingdom; 2) Paediatric Cardiology, University of Leuven, Leuven B-3000, Belgium; 3) Institute of Medical Genetics, University of Zurich, Zurich-Schwerzenbach CH-8603, Switzerland; 4) Heart Centre for Children, The Children's Hospital at Westmead, Sydney NSW 2145, Australia; 5) East Midlands Congenital Heart Centre, Glenfield Hospital, Leicester LE3 9QP, UK; 6) Department of Cardiovascular Medicine and Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK; 7) McGill University and Génome Québec Innovation Centre, Montréal (Québec) Canada H3A 0G1; 8) Institute of Cardiovascular Sciences, Manchester University, Manchester M13 9NT, UK.

Tetralogy of Fallot (TOF) is the commonest cyanotic congenital heart disease requiring surgery in infancy. The majority of cases do not occur as part of a syndrome. We have investigated rare nucleotide variants and copy number variants (CNVs) in over 800 non-syndromic TOF cases. For nucleotide variants we filtered for variants with MAF <1% in 1000 genomes and EVS, not shared between cases and 500 UK10K individuals, not present in more than 1% of individuals in a set and not falling within a segmental duplication. Clustering of variants in transcripts and in exons was ascertained using Poisson distribution. Correcting for the difference in sample numbers in the two groups, the number of clusters when considering synonymous variants was very similar in the TOF cases and UK10K individuals. In contrast, there was an excess of clusters in TOF cases when considering truncating, predicted deleterious and non-synonymous variants. Clusters were found in genes previously associated with TOF and novel candidate genes. CNVs were called for the same TOF cases on PCA corrected SNP array intensities (Cooper et al. Hum. Mol. Genet 2014), which enabled the joint analysis of case and control data while allowing for batch effects. Penncnv and quantiSNP were used for CNV calling on the corrected data and the controls for this work were as previously reported (Soemedi et al Am. J. Hum. Genet 2012). This enabled detection of additional smaller CNVs than reported in our earlier work.

612F

Genetic Causes for Congenital Heart Disease with Neurodevelopmental and Other Deficits. *J. Homsy[1,2], S. Zaidi[3], Y. Shen[4], J.S. Ware[1,5,6], J.G. Seidman[1], B.D. Gelb[7], E. Goldmuntz[8,9], M. Brueckner[3], R.P. Lifton[3,10], W.K. Chung[11], C.E. Seidman[1,12,13], Pediatric Heart Network and the Pediatric Cardiac Genomics Consortium.* 1) Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA; 2) Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA; 3) Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA; 4) Departments of Systems Biology and Biomedical Informatics, Columbia University Medical Center, New York, New York, USA; 5) NIHR Cardiovascular Biomedical Research Unit at Royal Brompton & Harefield NHS Foundation and Trust and Imperial College, London, London, UK; 6) National Heart & Lung Institute, Imperial College London, London, UK; 7) Department of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 8) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA; 9) Division of Cardiology, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA; 10) Howard Hughes Medical Institute, Yale University, New Haven, Connecticut, USA; 11) Departments of Pediatrics and Medicine, Columbia University Medical Center, New York, New York, USA; 12) Cardiovascular Division, Brigham & Women's Hospital, Harvard University, Boston, Massachusetts, USA; 13) Howard Hughes Medical Institute, Harvard University, Boston, Massachusetts, USA.

Congenital heart disease (CHD) is a severe birth defect that is often associated with extra-cardiac manifestations such as neurodevelopmental disabilities (NDD) and congenital anomalies (CA). Known genetic syndromes that cause CHD with extra-cardiac manifestations account for less than 10% of cases. We used whole exome sequencing to identify *de novo* mutations in 1,220 CHD case and 900 control parent-offspring trios. Using an adapted statistical framework for the analysis of *de novo* mutations, we found a marked excess of protein-damaging mutations in CHD cases, especially in genes highly expressed in the developing heart (enrichment 2.4x; p=1.6e-24) and brain (enrichment 2.2x; p5e-8). Mutations were enriched in genes involved in morphogenesis, chromatin modification, and transcriptional regulation, including *RBFOX2*, which regulates mRNA splicing. We found a significant overlap between genes mutated in our CHD cohort and in cohorts ascertained for NDD; moreover, CHD patients with these mutations were enriched for NDD. *De novo* mutations contributed to 19.5% of CHD with both NDD and CA but only 2.5% of isolated CHD. These findings reveal shared genetic contributions to CHD, NDD, and CA and suggest opportunities for improved assessment of neurodevelopmental prognosis and early therapeutic intervention in CHD patients.

613W

Utilization of Whole-Exome Sequencing to Identify Causative Mutations in Familial Congenital Heart Disease. *S. LaHaye[1,2,3], D. Corsemeier[4], J.L. Bowman[2,5], S. Fitzgerald-Butt[1,2,5], G. Zender[1], K. Bosse[1], K.L. McBride[1,2,5], P. White[4,5], V. Garg[1,2,3,5].* 1) Center for Cardiovascular and Pulmonary Research, Nationwide Children's Hospital, Columbus, OH; 2) The Heart Center, Nationwide Children's Hospital, Columbus, OH; 3) Department of Molecular Genetics, The Ohio State University, Columbus, OH; 4) Biomedical Genomics Core, Nationwide Children's Hospital, Columbus, OH; 5) Department of Pediatrics, The Ohio State University, Columbus, OH.

Congenital heart disease (CHD) is the most common type of birth defect, affecting ~8 out of every 1,000 live births, but there remains a limited understanding of the etiology for the majority of cases of CHD. Familial and population based studies have identified a strong genetic component to CHD, and familial cases of almost every type of CHD have been reported. The goal of this study was to identify pathogenic segregating variants in multiplex CHD families by whole-exome sequencing (WES). Target capture was performed using the Agilent SureSelectXT Target Enrichment kit, followed by WES on an Illumina HiSeq2500, on members from 9 families with Mendelian inherited CHD. Sequence alignment, post alignment processing, variant calling, and genotyping was performed with the Churchill pipeline. Tertiary analysis and annotation consisted of identifying rare (<1% mean allele frequency in the population) segregating variants, prioritizing by matching to a CHD gene list, and predicting pathogenicity via 12 different utilities. This strategy utilized a list of 69 CHD genes selected under strict criteria and allowed for prioritization of the variants. Predicted pathogenic mutations were identified in 3 of the 9 families. A splice donor site mutation was identified in *MYH11* (c.4599+1delC) in a family with autosomal dominant patent ductus arteriosus that is predicted to cause a 71 amino acid deletion, affecting functionality of the coiled-coil tail domain. A *GATA4* mutation, p.G115W, was identified in a family with autosomal dominant atrial septal defects (ASD) and is predicted to affect the transactivation ability of *GATA4*. A p.I263V mutation in *TLL1* was identified in a family with autosomal dominant ASD; as this mutation was located in the astacin-like metalloprotease catalytic domain it is predicted to affect the enzymatic efficacy of *TLL1*. In summary, our findings demonstrate the clinical utility of WES to successfully identify causative mutations in familial CHD and support the use of a CHD candidate gene list to allow for a more streamlined approach that enables rapid prioritization and identification of pathogenic variants from large WES data sets.

614T

Search for Novel Mutations Predisposing to Ventricular Fibrillation Without Overt Cause. *J.T. Leinonen[1], H. Swan[2], N. Junna[1], A. Lahtinen[3], M. Viitasalo[2], K. Kontula[3], E. Widen[1].* 1) Institute for Molecular Medicine Finland, FIMM, University of Helsinki, Helsinki, Finland; 2) Heart and Lung Center, Helsinki University Hospital, Helsinki, Finland; 3) Department of Medicine, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.

Ventricular fibrillation (VF) is a major cause of sudden cardiac death. In 5-10% of the cases the arrhythmia remains idiopathic, the patient presenting spontaneous VF in the absence of known ischemic, electrical, infectious or structural heart disease. While the pathology underlying idiopathic VF is likely variable and complex, the condition seems to have a heritable component. For example, a haplotype containing the gene *DPP6* associates with idiopathic VF in Dutch families, and a missense mutation in *SEMA3A* has been associated with idiopathic VF in a Japanese study.In the current study, we sought to identify mutations predisposing to idiopathic VF in Finland, by sequencing the exomes of a nation-wide cohort of patients resuscitated from out-of-hospital VF during 1999 – 2011. Standard clinical examinations in the primary hospitals had failed to identify any cause for the VF. The study comprised altogether 38 subjects (16 males and 22 females, age at diagnosis <55). Exome sequencing was carried out using Nimblegen SeqCap v2 kit.We restricted our analyses to 91 candidate genes previously associated with cardiac channelopathies, or cardiomyopathy. We screened the genes for rare (MAF < 0.0005 in the ExAC (http://exac.broadinstitute.org/) and Sequencing Initiative Suomi (http://www.sisuproject.fi/) databases) and likely damaging missense and nonsense variants. Based on these criteria, we identified mutations in 5/38 patients. Two of the patients had novel missense mutations in the CPVT-associated gene *RYR2,* whereas one patient carried a de novo missense mutation in *CACNA1C*, a gene previously associated with long QT-syndrome. One of the patients carried a rare nonsense mutation in the cardiomyopathy-associated limb-girdle muscular dystrophy gene *FKRP*, whereas the fifth patient carried a truncating mutation in the cardiomyopathy gene *TTN*. Interestingly, the subject with the *FKRP* mutation had family history of muscular dystrophy, and the *TTN* mutation-carrier had two cousins previously diagnosed with cardiomyopathy. Current work is underway to characterize the segregation of the identified mutations in the families, and to assess the current cardiac status of the patients and their relatives.Our data suggest that even if part of idiopathic VF patients may carry potentially pathogenic mutations in genes associated both with channelopathies and cardiomyopathy, such mutations are unlikely to be a major cause of idiopathic VF associated with sudden cardiac death.

615F

Whole exome sequencing identified known and novel genetic causes for pulmonary arterial hypertension. *L. Ma[1], C. Gonzaga-Jauregui[2], A. Sawle[3], Y. Shen[4], A. King[2], EB. Rosenzweig[5], JG. Reid[2], J. Overton[2], AR. Shuldiner[2], F. Dewey[2], WK. Chung[1].* 1) Molecular Genetics, Columbia Univ, New York, NY; 2) Regeneron Genetics Center Regeneron Pharmaceuticals Tarrytown, New York, USA; 3) Herbert Irving Comprehensive Cancer Center Columbia University Medical Center New York, New York, USA; 4) Departments of Systems Biology and Biomedical Informatics Columbia University Medical Center New York, New York, USA; 5) Department of Pediatrics Columbia University Medical Center New York, New York, USA.

Pulmonary arterial hypertension (PAH) is a rare disease characterized by distinctive changes in pulmonary arterioles that lead to progressive elevation of pulmonary artery pressure, pulmonary vascular resistance, right sided heart failure, and a high mortality rate. Mutations in *BMPR2* and a few other genes have been implicated in families with PAH, however ~20% of cases remain unexplained. To identify additional genetic causes of PAH, we performed whole exome sequencing in 60 PAH patients and when possible both parents, including 48 pediatric PAH probands. The study included 17 hereditary PAH, 43 idiopathic PAH, 8 congenital heart disease (CHD)-PAH, 2 hereditary hemorrhagic telangiectasia (HHT)-PAH, and 2 patients with syndromic PAH including either autism or a connective tissue disorder. We identified known and novel mutations in 18 of 48 (37.5%) patients with pediatric onset PAH. Seven of the 18 (38.9%) were due to *de novo* mutations. The genes identified, number of patients for each gene and their associated phenotypes are listed in Table 1. The high yield of whole exome sequencing in pediatric-onset PAH and the fact that novel genes were identified suggests that this is a useful strategy for gene identification and genetic diagnosis in children with PAH. Furthermore, the identification of novel genes for PAH provide insights into pathophysiology and potential new drug targets for treatment and prevention.Table 1. List of genes identified, number of patients for each gene and their associated diseases.

Gene	Number of Variants	Number of patients	Disease	Whether known/ novel PAH gene
ABCC8	5	5	Pediatric PAH, one autism-PAH	Novel
ACVRL1	1	1	HHT-PAH	Known
BMPR1B	1	1	Pediatric PAH	Known
BMPR2	2	2	Pediatric PAH	Known
EIF2AK4	1	1	Adult PAH	Known
ENG	1	1	Pediatric PAH	Known
SMAD6	1	1	Pediatric PAH	Known
TBX4	6	6	Pediatric PAH, one CHD-PAH	Known
ZFPM2	2	1	Pediatric PAH	Known

.

616W

Whole exome sequencing in 41 Czech families with inherited cardiovascular disorders: initial data from a representative Central European population. *M. Macek[1], N. Ptakova[3], A. Krebsova[2], L. Piherova[3], M. Kubanek[2], V. Stranecky[3], M. Macek Sr.[3], J. Kautzner[2], S. Kmoch[1], V. Melenovsky[1].* 1) Department of Biology and Medical Genetics, 2nd Faculty of Medicine and University Hospital Motol, Prague, Czech Republic; 2) Institute of Inherited Metabolic Disorders, 1st Faculty of Medicine and VFN, Prague, Czech Republic; 3) Department of Cardiology, Institute of Clinical and Experimental Medicine Prague, Czech Republic.

Whole exome sequencing (WES) fosters establishment of etiological diagnostics in hereditary forms of rare cardiovascular disorders, enables genetic stratification and may eventually facilitate introduction of precision medicine-based therapies. Altogether 96 families with ≥ 2 affected individuals were characterized based on established center care and using international clinical / laboratory standards. WES was carried out in a total of 41/96 families (TruSight Exome, Illumina, USA): 23/41 families suffered from dilated cardiomyopathy (DCM), 1/41 - non-compaction cardiomyopathy (LVNC), 14/41 - hypertrophic cardiomyopathy (HCM), 1/41 -arythmogenic cardiomyopathy (ACM) and 2/41 had unexplained cardiac arrest (UCA). Detected variants were confirmed by Sanger DNA sequencing and by segregation analysis. Pathogenic variants were revealed in 22/41 (64%) of studied families: 12/22 (54%) in DCM, 1/1 LVNC, 13/14 (92%) HCM, 1/1 ACM and in 1/2 familes with UCA. The most frequently affected gene in DCM was *TTN* (23%) and in HCM the *MYH 7* (29%). In case of LVNC, the likely pathogenic variants were identified in the *OBSC*, for ACM in *PKP2* and in the family with UAC in *SCN5A* genes, respectively. Although all disease genes were already identified, most variants were novel, thereby substantiating a population specific diagnostic strategy. In summary, obtained results are in accordance with other European studies and validate the integrated clinical care and genomic diagnostics in our country, whose genomic variation is representative of other Central European populations. Supported by: FNM 00064203, CZ.2.16/3.1.00/24022, NF-CZ11-PDP-3-003-2014, LD14073, IGA NT13770.to M.M. and IKEM 00023001 to V.M.

617T

Investigating the role of somatic mutation in congenital heart disease. *D.E. Miller[1, 2, 3], G. Twist[1], N.A. Miller[1], D.C. Bittel[1], E.G. Farrow[1], J.E. O'Brien[1], G.S. Shirali[1], R.S. Hawley[2, 3], S.F. Kingsmore[1, 4].* 1) Children's Mercy Hospital, Kansas City, MO; 2) Stowers Institute for Medical Research, Kansas City, MO; 3) University of Kansas Medical Center, Kansas City, KS; 4) University of Missouri Kansas City, Kansas City, MO.

Mutations that arise during development (congenital somatic mutations) are an understudied cause of human genetic disease. Although it is well understood that germline mosaicism produces offspring with new mutations not apparent in the parental somatic tissue, less is known about the impact of mosaicism within a particular individual. The falling cost of whole-genome and exome sequencing makes it feasible to sequence multiple tissues from a single individual in an effort to understand the impact of somatic mutations on the incidence of congenital disease. Hypoplastic left heart syndrome (HLHS) is a devastating condition in which the left ventricle of the heart does not fully develop. Although HLHS shows evidence of heritability, suggesting a genetic cause, very few genes or loci have been implicated in its development. To understand the role that somatic mosaicism may have in the development of HLHS, we generated exome (average depth of coverage: 180x) and whole-genome data (average depth of coverage: 50x) from both cardiac tissue and blood, as well as transcriptome data (cardiac tissue only, average depth of coverage: 350x) from four individuals with HLHS to determine if there is a role of somatic mutation in the development of this disorder. We do not find support for the hypothesis that a de novo variant present in cardiac tissue that is not found in blood may be causative for HLHS. Alternatively, we do find several examples of alleles found at low frequency in both blood and cardiac tissue that are suggestive of mutations that occurred during early development. Our study is a proof-of-concept pilot for the identification and analysis of de novo mutations present in one or few tissues in a single individual and forms the foundation of a larger analysis of the role of somatic mutation in congenital heart disease.

618F

Exome Analysis in 36 Unrelated Patients with Pregnancy-Associated or Peripartum Cardiomyopathy Demonstrates a Rare Variant Basis. *A. Morales[1], D.D. Kinnamon[1], D. Wheeler[1], M. Bamshad[2], D. Nickerson[2], R.E. Hershberger[2].* 1) The Ohio State University, Columbus, US; 2) University of Washington Center for Mendelian Genomics, Seattle, WA.

The underlying cause of peripartum cardiomyopathy (PPCM), a subtype of dilated cardiomyopathy (DCM), remains enigmatic with multiple etiologies suggested. PPCM is defined as DCM of unknown cause occurring in the last month of pregnancy or 5 months after delivery. Pregnancy associated cardiomyopathy (PACM) occurs earlier than the last month of pregnancy. To test the hypothesis that PPCM/PACM may have an underlying genetic etiology, we undertook exome sequencing of 37 PPCM/PACM DNA samples in our study (35 from unrelated cases and 2 from the same family). Exome sequencing was conducted at the University of Washington Center for Mendelian Genomics; 19 PPCM/PACM cases had been previously Sanger sequenced for 1-14 candidate genes. GATK (v.1.6) and GEMINI (v0.6.3.2) were used for sequence processing, variant calling and variant prioritization. Missense variants and indels were prioritized based on allele frequency <0.5% (1K Genomes or Exome Variant Server), GERP score >2 or conservation (Phastcons), and segregation among affected relatives with DCM or PPCM. Using a DCM variant adjudication scheme developed by our group, pathogenicity was assigned following ClinVar terminology: pathogenic (P), likely pathogenic (LP), or variant of uncertain significance (VUS). The analysis of variants from 40 previously identified DCM genes is presented here. Of 36 unrelated cases (27 PPCM, 9 PACM), 20 had familial DCM (f-DCM), and 16 had non-familial DCM (nf-DCM). The median age of onset was 26.5 years; 27 presented with heart failure; the mean LVEDD Z-score was 4.9, and the mean ejection fraction was 24%. A total of 43 unique variants were identified in 28 unrelated cases (2 P, 9 LP, 32 VUS). Pathogenic variants were identified in *PSEN2* and *SCN5A*. Likely pathogenic variants were identified in *LMNA, MYBPC3, MYH7, TNNT2*, and *TTN*. LP variants were identified in 2 nf-DCM cases. P/LP variants were identified in 10 f-DCM cases. Multiple variants were identified in 6 nf-DCM cases and 4 f-DCM cases. In 6 previously reported PPCM/PACM cases assayed with Sanger sequencing, no additional variants were found with stronger evidence favoring pathogenicity. We conclude that a proportion of PPCM/PACM, a subtype of DCM, has a rare variant genetic basis. Hence, a genetic evaluation is indicated for women diagnosed with PPCM/PACM, which includes genetic testing, as well as cardiovascular screening and predictive testing of at-risk relatives.

619W

Connective tissue disorders presenting severe arterial tortuosity in the perinatal period. *H. Morisaki[1], M. Horiuchi[2], J. Yoshimatsu[2], T. Yokoyama[3], M. Iwasa[3], Y. Takahashi[4], Y. Kawazu[5], N. Okamoto[5], T. Morisaki[1,6].* 1) Dept of Bioscience and Genetics, National Cerebral and Cardiovascular Ctr (NCVC), Osaka, Japan; 2) Div Perinatology and Gynecology, NCVC, Osaka, Japan; 3) Dept Pediatrics, Japanese Red Cross Nagoya Daini Hospital, Nagoya, Japan; 4) Dept Pediatrics, Chiba Univ. Grad Sch Medicine, Chiba, Japan; 5) Osaka Medical Ctr and Research Inst for Maternal and Child Health, Osaka, Japan; 6) Dept Molecular Pathophysiology, Osaka Univ Grad Sch Pharm Sci.

Arterial tortuosity (AT) is often accompanied with heritable connective tissue disorders (CTD), including arterial tortuosity syndrome caused by *SLC2A10* mutations, autosomal recessive cutis laxa type IB by *FBLN4* mutations, and Loeys-Dietz syndrome (LDS) by *TGFBR1/2* mutations. Diagnosis of these conditions is not easy in early childhood, though it is not difficult to diagnose them in adulthood. We experienced 5 cases with severe AT during prenatal and early infantile period, showing heart or respiratory failure as a main symptom. First case was a 2-year-old boy with compound heterozygous *SLC2A10* mutations, presenting dyspnea due to pulmonary emphysema after RSV infection. MRA showed annulo-aortic ectasia (AAE) and severe AT and elongation of the aorta and medium-sized arteries in both brain and abdomen. Second case was a 3-month-old boy with compound heterozygous *FBLN4* mutations, developing severe aortic regurgitation, dilatation of ascending aorta with isthmal narrowing and poststenotic dilatation, micro-thrombosis in the brain arteries. He underwent a successful surgical replacement of ascending aorta at age 2. Third case was a female fetus with *FBLN4* mutations presenting hydrops fetalis due to dysplastic tricuspid valve with an elongated aorta at 32 weeks gestation. The pregnancy ended in intrauterine death at 33wk. Autopsy revealed severe AT of total aorta with thickened wall and narrow lumen along with kinky aneurysm in subclavian and iliac arteries. The fourth case was a newborn baby with *TGFBR1* mutation, prenatally diagnosed as DORV and interrupted aortic arch with severe pulmonary artery dilatation, highly tortuous aorta and cervical arteries. He had emergency pulmonary banding followed by dilatation of descending aorta but died on 21st day. Fifth case was a 4-month-old girl with *TGFBR2* mutation, diagnosed as LDS with cleft palate, club foot and arachnodactyly. Ultrasound study showed severe AT, AAE and ASD/PDA. Her bronchi were depressed by dilated pulmonary artery. She subsequently died of respiratory failure due to cervical spine dislocation. None of these 5 cases presented any typical CTD features like cutis laxa, chest deformity or joint hypermobility during perinatal period, while AT was a sole symptom towards proper diagnosis. Prognosis was generally poor but early intervention was effective in some cases. It is important to consider CTD in such very young patients with AT.

620T

Whole exome sequencing, blood lipids, and cardiovascular outcomes in 31,000 participants in the Regeneron Genetics Center – Geisinger Health System (DiscovEHR) human genetics collaboration. *C. O'Dushlaine[1], J.B. Leader[2], C. Van Hout[1], S. Mukherjee[1], S.A. Pendergrass[2], D.R. Lavage[2], L. Habegger[1], O. Gottesman[1], A. Lopez[1], C. Gonzaga-Jauregui[1], S. Hamon[1], A. Verma[3], J. Wallace[3], S. Dudek[3], M.D. Ricthie[2], D.J. Carey[2], M.F. Murray[2], D.H. Ledbetter[2], J.D. Overton[1], J.G. Reid[1], A. Baras[1], A.R. Shuldiner[1], H.L. Kirchner[2], F.E. Dewey[1].* 1) Regeneron Genetics Center, Tarrytown, NY; 2) Geisinger Health System, Danville, PA; 3) The Center for Systems Genomics, The Pennsylvania State University, State College, PA.

Low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and triglyceride levels are heritable risk factors for cardiovascular disease. Sequencing the protein coding regions of the genome (exome) enables discovery of rare mutations that may exert large effects on lipid traits and cardiovascular disease risk.We sequenced the exons of 18,951 protein-coding genes in 31,058 participants of predominantly European ancestry in the Regeneron Genetics Center-Geisinger MyCode-DiscovEHR collaboration. Tests of association were performed for ~350,000 individual variants with minor allele count ≥ 20 and for non-synonymous and putative loss-of-function (LoF) variants aggregated by gene with median lifetime lipid levels extracted from electronic health records (EHR) of 22,400 individuals. Lipid-associated variants and genes were tested for associations with coronary artery disease (CAD) in 8,496 cases and 22,562 controls. Phenotypic associations were expanded with a phenome-wide association study (PheWAS) of ~1,700 discrete disease traits.Analysis of this sample as discovery (n=~12,400) and replication (n=~10,000) strata for lipid levels identified 50 replicated coding variant associations in 32 genes that were also associated with lipid levels in joint analysis at P < 1.4x10-7, including multiple coding variants not previously associated with lipid levels. Gene-based association tests identified collections of rare non-synonymous and LoF variants in 8 genes that were reproducibly associated with lipid levels in staged analysis and also in joint analysis at P<1.3x10-6. In addition to known coding variant associations, these results implicated novel protein-disrupting mutations in 4 known lipid loci (*ABCA1, LCAT, SCARB1, LIPC*). Two novel lipid-associated variants were associated with CAD risk at P < 0.05. Multivariate testing for lipid traits identified additional coding variants with evidence for substantial effects on all multiple traits. PheWAS of these loci revealed multiple additional known and previously undiscovered disease associations.Using whole-exome sequencing in over 31,000 individuals with EHR-derived clinical phenotypes, we confirmed previous coding variant associations with lipid levels, and identified new associations between collections of rare, protein-disrupting mutations, lipid levels, and coronary artery disease. These results implicate rare protein-disrupting variation in the genetic architecture of lipid traits and CAD.

621F

Identification of Novel Risk Genes for Venous Thromboembolism using Rare Coding Variant Burden Analysis in the GIFT and ELATE/DODS Cohorts. *A. Ozel[1], K. Desch[2], M. Germain[4], C. Willer[1], D. Siemeniak[3], D. Tregouet[4], P. Reitsma[5], D. Ginsburg[1,3], J. Li[1].* 1) Department of Human Genetics, University of Michigan - Ann Arbor, Ann Arbor, MI; 2) Department of Pediatrics, University of Michigan - Ann Arbor, Ann Arbor, MI; 3) HHMI, University of Michigan - Ann Arbor, Ann Arbor, MI; 4) INSERM, Università Pierre et Marie Curie, Paris, France; 5) Leiden University Medical Center, Leiden, Netherlands.

Venous thromboembolism (VTE) affects ~1 million individuals in the United States and is the third most common cause of cardiovascular death. It is a highly heritable disease with ~60% of the lifetime risk attributable to genetic factors. While genome-wide association studies (GWAS) validated the involvement of common variants in the*Factor 5*(*F5*) and*ABO*genes, which account for 30% of the overall attributable risk for VTE, they are not suited to discover rare causal variants. The unique biology of VTE and the complex interactions between the hemostatic and host defense systems lead us to hypothesize that a limited number of rare, high-impact coding variants might help explain the remaining portion of the inherited risk. In this study, we compared the per-gene aggregate rare variant burden between 392 VTE patients from the GIFT, ELATE/DODS cohorts and 342 non-VTE subjects of the same ancestry (European). After joint calling of coding and splicing variants using exome sequencing data of these samples, we conducted case/control gene burden tests while accounting for the technical heterogeneity between the controls and cases by using the each gene's synonymous variant count as the background variant discovery rate. We used multiple annotations to define deleterious variants: (1) PolyPhen2, (2) CADD >= 10, (3) CADD >= 20, (4) CONDEL, (5) PhyloPNH, and performed the burden analysis at several minor allele frequency (MAF) thresholds: 0.01, 0.05, 0.10, 0.15 and 0.50. We analyzed all genes as well as 95 candidate genes that were implicated in VTE according to prior literature. As a positive control, we confirmed the known association of the Factor 5 Leiden (R506Q) mutation with VTE (p=8.7E-19, odds ratio=8.7; 95 % CI: 4.8 – 16.8). Among the 95 candidate genes, rare variants (MAF <= 5%) in *F5*, *SERPINA1*, *GP1BA* and *HNF4A*had significantly higher burden (p<0.00053) in VTE cases with a majority of the deleterious variant definitions. Among all 20,000 genes, rare variant burden in*F5*, *SELP*, *VWF* and *GP1BA* rank among the top 3% with at least two of the definitions. Our results suggest the aggregate contribution of rare variants in both known and unknown genes to the overall VTE risk. Additional analyses are underway to use publicly available exome control data to replicate these findings.

622W

Loss-of-function *LOX* mutations cause thoracic aortic aneurysms and acute aortic dissections. *E.S. Regalado[1], D.-C. Guo[1], L. Gong[1], X. Duan[1], E.M. Hostetler[1], R.L.P. Santos-Cortez[2], G. Jondeau[3], C. Boileau[4], R. Moran[5], M.J. Bamshad[6], J. Shendure[6], D.A. Nickerson[6], S.M. Leal[2], D.M. Milewicz[1], University of Washington Center for Mendelian Genomics.* 1) Department of Internal Medicine, University of Texas Health Science Center at Houston, Houston, Texas, USA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 3) AP-HP, Hôpital Bichat, Centre National de Référence pour le syndrome de Marfan et apparentés, Paris, France; Università Paris 7, Paris, France; AP-HP, Hôpital Bichat, Service de Cardiologie, Paris, France; INSERM, U1148, Paris, France; 4) AP-HP, Hôpital Bichat, Centre National de Référence pour le syndrome de Marfan et apparentés, Paris, France; Università Paris 7, Paris, France; AP-HP, Hôpital Bichat, Laboratoire de Génétique moléculaire, Boulogne, France; INSERM, U1148, Paris, France; 5) Genomic Medicine Institute, Cleveland Clinic, Cleveland, Ohio, USA; 6) Department of Genome Sciences, University of Washington, Seattle, Washington, USA.

Aortic structural integrity is dependent on the activity of lysyl oxidases, which are extracellular copper enzymes that catalyze crosslinking of elastin and collagens. Mammals have five lysyl oxidase genes with highly similar catalytic domains, including the prototypic *LOX*. Deletion of *Lox* in mice leads to loss of 80% of lysyl oxidase activity in the aorta, and the mice die shortly after birth of aortic rupture. Exome sequencing was performed using DNA from two affected individuals from family TAA602 with autosomal dominant inheritance of thoracic aortic aneurysm and dissection. Using previously described filtering strategies, the possible causative variants were narrowed to 10 candidates. A variant in *LOX* was the best candidate for causing disease in TAA602 because of the established role of lysyl oxidase in aortic integrity. The variant, p.Ser280Arg, disrupts a highly conserved residue within the catalytic domain of the enzyme, is predicted to be damaging by 5 of 7 functional predictors, and segregates with aortic disease in the family. Exome and Sanger sequencing of 395 unrelated probands with familial thoracic aortic disease identified three additional rare *LOX* variants that were predicted to be damaging (p.Gly202X, p.Thr248Ile, p.Ser348Arg); one leads to haploinsufficiency and two are located in the catalytic domain and predicted to be damaging. Notably, the ESP database had no *LOX* nonsense or indel variants and only one rare variant in the catalytic domain in the European Americans. To determine if these variants in the catalytic domain disrupt enzymatic activity, expression plasmids of *LOX* wildtype (WT) and the missense variants were constructed and transfected into Hela cells and lysyl oxidase activity assayed. Based on equal protein amounts assayed, Hela cells transfected with p.Ser280Arg, p.Ser348Arg, and p.Thr248Ile mutant LOX constructs showed increased lysyl oxidase activities that were significantly lower compared to the WT LOX plasmid transfected cells by 58% (p=.002), 33% (p=.003), and 10% (p=0.003), respectively. The clinical disease in all families is characterized by fusiform aneurysms involving the root and ascending aorta and ascending aortic dissections and bicuspid aortic valves in 28% of mutation carriers. In conclusion, loss-of-function *LOX* mutations are a rare cause of familial thoracic aortic aneurysms.

623T

Clinically relevant variants identfied in thoracic aortic aneurysm patients by research exome sequencing. *J.A. Schubert[1,2], B.J. Landis[3], A.R. Shikany[3], R.B. Hinton[3], S.M. Ware[1,2].* 1) Indiana University School of Medicine, Indianapolis, IN; 2) University of Cincinnati College of Medicine, Cincinnati, OH; 3) Cincinnati Children's Hospital Medical Center, Cincinnati, OH.

Thoracic aortic aneurysm (TAA) is a genetically heterogeneous disease involving subclinical and progressive dilation of the thoracic aorta, which can lead to life-threatening complications such as dissection or rupture. Genetic testing is important for risk stratification and identification of at risk family members, and clinically available genetic testing panels have been expanding rapidly. However, when past testing results are normal, there is little evidence to guide decision making about when and in whom to pursue additional clinical genetic testing. Results from research based genetic testing can help inform this process. Here we present 10 TAA patients who have a family history of disease and who enrolled in research-based exome testing. Nine of these ten patients had previous clinical genetic testing, all of which was negative. We sought to determine the number of rare variants in 23 known TAA associated genes identified by research-based exome testing. In total, we found 13 rare variants in 7 patients, with no rare variants in these 23 genes in 3 patients. Likely pathogenic variants included one patient with a *TGFB2* variant and another with a *SMAD3* variant. These variants have been reported previously in individuals with similar phenotypes. In addition, novel variants in both *MYLK* and *MFAP5* were identified in a third patient. In total, clinically reportable rare variants were found in 5/10 (50%) patients, with at least 2/10 patients having likely pathogenic variants identified. These results highlight the importance of re-testing TAA patients with previous negative or inconclusive testing on a broader scale, as the field rapidly expands.

624F

Association of a single nucleotide polymorphism variant in the *SCUBE1* gene with Coronary Artery Calcium score in the *ClinSeq*® Study. *H. Sung[1], B. Suktitipat[1], K. Lewis[2], D. Ng[2], SG. Gonsalves[2], JC. Mullikin[3,4], LG. Biesecker[2], AF. Wilson[1], NISC Comparative Sequencing Program.* 1) Genometric Section, NIH/NHGRI/CSGB, Baltimore, MD; 2) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 3) Comparative Genomics Analysis Unit, Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 4) National Institutes of Health Intramural Sequencing Center (NISC), National Human Genome Research Branch, NIH, Bethesda, MD.

ClinSeq® is a large-scale medical sequencing study designed to investigate associations of rare sequence variants with traits related to Coronary Artery Disease (CAD). The study currently includes more than 1000 non-smoking participants between the ages of 45 and 65 with normal to severe coronary artery calcification scores. About 200 traits related to CAD were measured at the NIH Clinical Research Center in Bethesda, MD. Exome sequencing was performed with the Agilent SureSelect 38Mb and 50Mb capture kits for 387 and 325 individuals, respectively, at the NIH Intramural Sequencing Center. Single nucleotide variants (SNVs) common to both capture regions having a call rate > 98% and minor allele frequency (MAF) > 1% were used to check for cryptic relatedness and for population stratification by multidimensional scaling analysis. After removing samples failing this screening, 635 unrelated European Americans (EAs) remained. For each capture kit with EAs only, SNVs without at least one homozygote or a call rate less than 50% were excluded. Then samples sequenced with the two capture kits were merged, leaving 439,807 shared SNVs. Of these SNVs, 74% and 46% had MAFs < 0.01 and < 0.001, respectively. SNVs with MAF < 0.01 were collapsed into a single derived variant for each genomic region defined using hotspot boundaries. Collapsed variants were coded as the proportion of minor alleles occurring within each region; common variants were coded as the proportion of minor alleles (scaled from 0 to 1). Tests of association of each variant with Coronary Artery Calcium (CAC) score (available in 550 samples, excluding samples with stents), were performed on log-transformed CAC with simple linear regression, adjusting for age, sex, BMI and cholesterol lowering medication use. The derived collapsed variant incorporating rs73420094 (MAF=0.0056) in the exon and rs193288922 (MAF=0.0018) in the intron of the *SCUBE1* gene was found to have a suggestive significant association at the level of 1e-05. All seven samples having a non-reference allele (A) of rs73420094 have a CAC score within the top 14% of the distribution. The next most significant variant, rs3751058 in the *SCUBE2* gene, was found to have an association at the 1e-04 level. Inhibition of *SCUBE1* was found by Wu et al. to protect against thrombosis in mice and *SCUBE2* is found by Ali et al. to be involved in atherosclerotic plaque progression via Hedgehog signal transduction.

625W

Arterial Tourtosity in two Turkish pediatric patient; with novel homozygous missence mutations in the SLC2A10 gene. *F. Uysal[1], S.G Temel[2,3], O.M Bostan[1], D. Proost[4], L.V Laer[4], B. Loeys[4], E. Cil[1].* 1) Uludag University, Faculty of Medicine, Department of Pediatric Cardiology, Bursa, Turkey; 2) Near East University, Faculty of Medicine, Department of Histology & Embryology, Nicosia, North Cyprus; 3) Uludag University, Faculty of Medicine, Department of Histology & Embryology, Bursa, Turkey; 4) Center for Medical Genetics, Antwerp University Hospital and University of Antwerp, Antwerp, Belgium.

Arterial tortuosity syndrome (ATS) is a rare autosomal recessive connective tissue disorder, mainly characterized by tortuosity and elongation of the large and medium-sized arteries with predisposition to stenosis and aneurysm. ATS is caused by mutations in the *SLC2A10* gene, encoding for the facilitative glucose transporter 10 (GLUT10) and is described typically in pediatric patients. We report two unrelated boys (2 and 1 year old) with ATS who initially presented with a cardiac murmur and respiratory distress, respectively. In the first proband echocardiography at the age of one month demonstrated dilatation of the aortic root (15mm, Z-score=) and tortousity in the descending aorta. The main clinical findings included elongated face, saggy cheeks, micrognathia, malar hypoplasia, joint hypermobility and hyperextensible skin. Follow-up echocardiography showed an increase of the diameter of the aortic root to 32mm and 35 mm at the age of 8 months and 2 years old, respectively. Angiography and computerized tomography angiography showed fusiform aneurysmal dilatation of ascending aorta, marked tortuosity of both pulmonary arteries and thoracic aorta, stenosis in vena cava inferior. For the second proband, echocardiography at the age of one week dilatation in ascending aorta and anomaly suspicion in pumonary artery (pumonary sling). He had an ileus and operated at the age of one month. The main clinical findings included elongated face, saggy cheeks, migrognathia, pectus excavatum and hyperextensible skin. Anjography showed dilatation in ascending aorta, tortuosity of arch branching arteries and 'S' like tortuosity in descending aorta and normal pulmonary arteries. Sequencing of the *SLC2A10* gene in the probands revealed the presence of two novel pathogenic homozygous missense variants (c.727C>A/c.395C>A). Heterozygous *SLC2A10* mutations were confirmed in both mothers and fathers. The pathogenic variants lead to a p.Gln243Lys and p.Arg132Gln change in the seventh transmembrane domain and endofacial loop of GLUT10, respectively. The first proband underwent aortic root replacement surgery. ATS resembles Loeys-Dietz and Marfan related disorders, so timely differential diagnosis is extremely important for early diagnosis and intervention of aneurysms to prevent serious vascular complications.

626T

Genetic testing practices in pediatric cardiomyopathy: identifying opportunities to positively impact diagnosis and family-based risk stratification. *S.M. Ware[1], S.E. Lipshultz[2], S.D. Colan[3], L. Shi[4], C.E. Canter[5], D. Hsu[6], D.A. Dodd[7], M.D. Everitt[8], P.F. Kantor[9], L.A. Addonizio[10], J.L. Jefferies[11], J. Rossano[12], E. Pahl[13], P. Rusconi[14], J. Schubert[15], T. Lee[16], E. Miller[11], M. Tariq[1], J.D. Wilkinson[2], J.A. Towbin[17].* 1) Pediatrics, Indiana University, Indianapolis, IN; 2) Pediatrics, Wayne State University, Detroit, MI; 3) Pediatrics, Harvard Medical School, Boston, MA; 4) Biostatistics, New England Research Institutes, Watertown, MA; 5) Pediatrics, Washington University, St. Louis, MO; 6) Pediatrics, Children's Hospital at Montefiore, Bronx, NY; 7) Pediatrics, Vanderbilt Heart & Vascular Institute, Nashville, TN; 8) Pediatrics, Denver Children's Hospital, Denver, CO; 9) Pediatrics, Stollery Children's Hospital, Edmonton, AB, Canada; 10) Pediatrics, New York-Presbyterian Morgan Stanley Children's Hospital, New York, NY; 11) Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 12) Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA; 13) Pediatrics, Northwestern University, Chicago, IL; 14) Pediatrics, University of Miami Hospital, Miami, FL; 15) University of Cincinnati College of Medicine, Cincinnati, OH; 16) Pediatrics, Columbia University, New York, NY; 17) Cardiology, Le Bonheur Children's Hospital, Memphis, TN.

Consensus guidelines recommend genetic testing in children and adults for hypertrophic cardiomyopathy (HCM), and consideration of testing for dilated or restrictive cardiomyopathy (DCM, RCM). Cardiac surveillance is recommended in first degree relatives, and genetic testing for a known familial mutation (KFM) is indicated. Despite these recommendations, payer reimbursement rules, family choice, and practice barriers to testing and screening exist, and uptake in the pediatric population, where the disease is rare and causes are heterogeneous, is unclear. We evaluated genetic testing practices for 154 consecutively enrolled pediatric patients with cardiomyopathy at 11 sites as part of a large, prospective NHLBI-funded study. Inclusion criteria included patients with familial or idiopathic disease or unconfirmed myocarditis. Patients with known genetic syndromes, inborn errors of metabolism, mitochondrial disease, or neuromuscular disease were excluded. A pedigree was obtained, and clinical genetic testing reports were collected if applicable. All patients provided blood samples for research-based exome sequencing. In this preliminary sample of the target cohort, 48% (n=74) had previous clinical genetic testing, with 25% (n=41) of the cohort having a likely pathogenic or pathogenic variant. At enrollment, 38% had a family history of cardiomyopathy. The rate of clinical testing by phenotype was 72% for HCM, 36% for RCM, 45% for DCM, and 7% for LVNC/unknown. All variants reported by clinical genetic testing were also identified by research-based exome. Filtering exome results for 35 genes available on clinical cardiomyopathy panels revealed 49 additional clinically actionable results, 38 of which were in previously untested patients. Overall, 79/154 patients had positive findings (51%), including 75% with HCM, 71% with RCM, and 35% with DCM. Importantly, the 38 patients with newly identified genetic disease had 148 first degree relatives for whom KFM testing can now be offered, allowing risk stratification and further cascade screening. This is the first multisite evaluation of clinical genetic testing practices in the pediatric cardiomyopathy population. The results demonstrate testing is currently underutilized, show a high diagnostic yield across phenotypes, and indicate an opportunity to significantly impact cascade screening and risk stratification for family members.

627F

Identification of Second-hit Mutations in Known CHDs Causative genes in 22q11DS Patients with Conotruncal Heart Defects by Whole-exome Sequencing. *T. Guo[1], JH. Chung[1], D. McDonald-McGinn[2], WR. Kates[3], T. Wang[4], B. Emanuel[2,5], B. Morrow[1].* 1) Department of Genetics, Albert Einstein College of Medicine, Bronx, NY; 2) 2Division of Human Genetics, Children's Hospital of Philadelphia and Department of Pediatrics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania; 3) 3Department of Psychiatry and Behavioral Sciences, and Program in Neuroscience, SUNY Upstate Medical University, Syracuse, New York; 4) Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York; 5) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

The 22q11.2 deletion syndrome (22q11DS, velo-cardio-facial/DiGeorge) is the most common microdeletion disorder, occurring in 1:4,000 births. Most of the patients have a similar deletion on chromosome 22q11.2, but approximately 60% have congenital heart defects (CHD), mostly of the conotruncal type, while 40% have normal cardiac anatomy. We hypothesize that 22q11DS patients with CHD have second-hit mutations in known CHD genes reported in the literature. To test this, whole exome sequencing was performed on DNA samples from 184, 22q11DS subjects including 89 cases with CHD and 95 controls without. Rare variants (MAF < 0.001 in 1000 Genome and ESP6500 projects) that are predicted to be damaging in any of the 5 specific predictions (SIFT, Polyphen2 HVAR, MutationTaster, MutationAssessor, FATHMM and scaled CADD) were retained for analysis. We compiled 66 pathogenic, CHD-causing CNVs (copy-number variations) from the Copy Number Variation in Disease database (PMID: 22826268). These 66 CHD-causing CNVs encompass 782 genes which we tested for rare deleterious variants. A total of 259 genes had at least one deleterious mutation in either a case or control with 22q11DS. Among the 259 genes, we identified 6 cases (5 with tetralogy of Fallot (TOF); 1 with aortic arch artery anomalies) with deleterious heterozygous mutations in the *FRY* (FURRY) gene, but none in controls (Fisher extract p value=0.01). *FRY* is located on chromosome 13q13.1-q13.2, which is the CHD critical region of the 13q deletion syndrome (also called monosomy 13q syndrome). Among 35 reported patients with chromosomal deletions of 13q13.1-q13.2, 8/11 cases were reported to have conotruncal heart defects (PMID: 20598760). Furry (Fry) is an evolutionarily conserved protein of ~300 kDa that contains multiple HEAT/Armadillo-like repeats in the N-terminal region. Mammalian Fry plays crucial roles in chromosome alignment and bipolar spindle organization during mitosis, critical for asymmetric cell division, regulation of cell shape and early embryonic development, implicating it as a strong candidate gene within the region to explain the basis of CHD in patients with chromosome 13q13.1-q13.2 deletions and as a modifier of this phenotype in 22q11DS.

628W

Sequence Data Processing and Analysis of the first 20,000 Human Genomesin the NHLBI TopMed Program. *G.R. Abecasis[1], D. Applebaum-Bowden[2], K. Barnes[3], T. Blackwell[1], J. Blangero[4], E. Burchard[5], A. Correa[6], P. Ellinor[7], C. Fox[2], S. Gabriel[7], W. Gan[2], C. Jaquish[2], H.M. Kang[1], S. Kathiresan[7], C. Laurie[8], S. McGarvey[9], B. Mitchell[10], D. Nickerson[8], J. O'Connell[10], G. Papanicolau[2], P. Qasba[2], R.S. Vasan[11], S. Redline[12], K. Rice[8], E. Silverman[13], J.D. Smith[8], P. Srinivas[2], S. Weiss[14], J. Wilson[6] for the National Heart Lung and Blood Institute's TopMed Program.* 1) University of Michigan, Ann Arbor, MI; 2) National Heart Lung and Blood Institute, Bethesda, MD; 3) Johns Hopkins University, Baltimore, MD; 4) University of Texas Rio Grande Valley, Rio Grande, TX; 5) University of California San Francisco, San Francisco, CA; 6) University of Mississippi Medical Center, Jackson, MS; 7) Broad Institute of Harvard and MIT, Cambridge, MA; 8) University of Washington, Seattle, WA; 9) Brown University, Providence, RI; 10) University of Maryland, Baltimore, MD; 11) Boston University School of Medicine, Framingham, MA; 12) Harvard Medical School, Boston, MA; 13) Brigham and Women's Hospital, Boston, MA; 14) Partners Health Care, Boston, MA.

The National Heart Lung and Blood Institute's Trans-Omics for Precision Medicine (TopMed) initiative aims to stimulate discovery of the fundamental mechanisms that underlie heart, lung, blood and sleep disorders by integrating high quality phenotypes, whole genome sequence data and other genomic information. The first phase of the effort will generate deep whole genome data for ~20,000 diverse samples (a majority with non-European ancestries traditionally under-represented in genomic research) and represents a collaboration between diverse groups of scientists, 4 genome sequencing centers, a project data coordination center, an informatics research center, staff at the National Institutes of Health and the National Center for Biotechnology Information (NCBI). Future phases of the program will include additional cohorts, genome sequencing and other omics data.Here, we describe efforts at the project's Informatics Research Center (IRC)to integrate sequence data across participating studies and sequencing centers and their initial results. As of abstract submission, >19,800 samples are prioritized for sequencing, with data for >1,000 samples received at the IRC. Sequencing is expected to be complete by February 2016; sequence data for >5,000 samples is expected by October 2015. To date, we have achieved an average sequencing depth of 36.9 (after duplicate removal), resulting in >98% of the genome covered at depth ≥10. We have evaluated alternatives for transfer and storage of sequence data, including file formats, quantization of base quality scores and others, to achieve reasonable costs for data processing and to allow the project to scale to even more samples. Deep sequencing, PCR-free sample preparation, and harmonized data processing already allow us to achieve <0.2% difference in SNP callsets when a sample is sequenced at different sites. Although more differences are observed for indels, these are concentrated in low complexity regions of sequence, such as short tandem repeats.A primary goal of the project is to allow investigators exploring diverse study designs and phenotypes to benefit from access to high-quality sequence data and analysis. There are also many opportunities from combined analysis of the data being generated, including a diverse catalog of genetic variation, a resource of controls for association studies, and an imputation reference panel. We discuss some of these opportunities and associated logistical and technical challenges.

629T

Identifying low frequency and rare coding variation influencing cardiometabolic traits through whole exome sequencing of 20,000 Finns: the FinMetSeq Study. *A.E. Locke[1], K. Meltz Steinberg[2], S. Service[3], V. Ramensky[3], M. Pirinen[4], H.M. Stringham[1], A.U. Jackson[1], M. Kurki[5], D.E. Larson[2], L.J. Scott[1], R.S. Fulton[2], D.C. Koboldt[2], J. Nelson[2], T.J. Nicholas[2], S. Ripatti[4,6,7], V. Salomaa[8], A. Palotie[4,5,7], M. Laakso[9], N. Freimer[3], R.K. Wilson[2], M. Boehnke[1].* 1) Center for Statistical Genetics and Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2) McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO; 3) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California Los Angeles, Los Angeles, CA; 4) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 5) Broad Institute of MIT and Harvard, Cambridge, MA; 6) Department of Public Health, Hjelt Institute, University of Helsinki, Helsinki, Finland; 7) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK; 8) National Institute for Health and Welfare, Helsinki, Finland; 9) Department of Medicine, Kuopio University Hospital and University of Eastern Finland, Kuopio, Finland.

Due to recent population bottlenecks, low frequency, deleterious variants are enriched in the Finnish population. The resulting increase in power and the large population studies ongoing in Finland provide an excellent resource for complex trait genetics.We are sequencing the exomes of 20,000 individuals from two large population-based longitudinal studies of metabolic and cardiovascular disease in Finland: METSIM and FINRISK. Currently, we have completed sequencing for the first 10,628 individuals. Using an integrated variant discovery and genotyping approach across three sites -- McDonnell Genome Institute, University of Michigan, and UCLA -- we identified more variable sites (>1.5M SNPs) and achieved greater overall genotype accuracy than any single sequence analysis pipeline.We performed single-variant and gene-based association tests using EMMAX for >30 cardiometabolic phenotypes, identifying 46 apparently independent associations at exome-wide significance ($p<5x10-7$) at 26 loci in 9 different traits. While most observed associations, known and novel, were common, seven index SNPs had a minor allele frequency (MAF) <5%, and four <1%. We replicated low frequency non-synonymous variants in *PCSK9* (rs11591147) associated with decreased levels of total and LDL cholesterol and ApoB levels, in *HNF4A* (rs1800961) associated with lower HDL cholesterol, and a non-synonymous variant in *LPA* (rs3798220) with MAF~1% associated with total cholesterol. All of these variants are equally or more frequent in our cohorts than in other European populations.Finally, we identified a non-synonymous variant with MAF<0.5% in *LCAT* associated with significantly lower HDL cholesterol, which is known to convert cholesterol and phosphatidylcholines to cholesterol esters on HDL particles. Mutations in *LCAT* are also known to cause Fish-eye and Norum diseases, both of which are characterized by significantly altered lipid profiles and corneal opacity. This particular variant has not been implicated in either Fish-eye or Norum disease, so we plan to re-contact carriers to identify vision problems and more fully examine their lipid profiles.The population history of the Finnish people, the diverse array of phenotypes, and the large-scale but focused nature of this exome sequencing study allow us a unique opportunity to identify coding variation influencing cardiometabolic traits, and we plan to follow up candidate variants by collecting additional phenotypes.

630F

POSSIBLE PATHOLOGICAL EVENTS INVOLVED IN DILATED CARDIOMYOPATHY. *M.L. Satyanarayana[1], C. Narsimhan[2], N. Pratibha[1].* 1) Department of Genetics, Osmania University, Hyderabad, Telangana State, India; 2) CARE Hospitals, Hyderabad, Telangana State, India.

Dilated cardiomyopathy (DCM) is characterized by ventricular chamber dilation and diminished cardiac contractility, the most common cause of chronic heart failure (CHF) and cardiac transplantation. The underlying etiology of DCM varies due to genetic factors, viral (myocarditis), mitochondrial, and metabolic disorders. To date, candidate gene screening led to the identification of numerous heterozygous mutations in more than 40 different genes. The present study is an attempt to elucidate the interactions between the sarcomeric/cytoskeletal and the modifier gene(s), by screening *MYH7, MYBPC3, ANKRD1, EDN1, NOS3, IL-1B, TNF-A, HSP-70* gene variations, in order to understand the underlying pathogenesis of DCM, accounting to the phenotypic and genetic heterogeneity.The *MYH7 (G377R, E1258D)* mutations were predicted to disrupt actin-myosin binding and cross-bridge function, as localized to the LMM region. *MYBPC3 (V158M)* mutations resulted in the disruption of the sarcomeric assembly, posttranslational modifications which can cause alterations in myosin crossbridge architecture leading to **'defective force production'**. The novel protein truncation *(E583X)* mutation of *MYBPC3* produces a truncated protein **lacking titin and myosin binding sites.** *MYH7* mutations *(E1258D, A1379T and K1537N)* disrupting the sarcomere assembly with cytoskeletal protein network resulting in **reduction of force transmission**. Further, *ANKRD1 (D304H)* mutation resulted in the loss of one of the ankyrin repeats, hence altering the sarcomere interaction with cytoskeletal proteins **titin** and **myopalladin**, leading to reduced force transmission.*Injury to the myocardial cell* may results in the release of cytokines *(IL-1Δ,TNF-α)*, which can be detrimental to cardiac function, leading to overt congestive heart failure in a DCM phenotype. *Oxidative stress* and *heightened peripheral vasoconstriction* are other key features, wherein *NOS3* and/or *EDN1* gene variations can cause abnormal and excessive remodeling of the peripheral vasculature by imposing pressure on the ventricles. *Apoptosis* is believed to play an important role in the continuous loss of myocardial cells during chronic heart failure. Hence compensatory mechanisms like anti-apoptotic pathway can also be triggered, wherein *HSP70* plays a key role.Thus, suggesting the implication of sarcomeric/cytoskeletal gene mutations with epistatic interactions of modifier genes, accounting to complexity and multifactorial mode of inheritance of DCM.

631W

Effects of rare coding variants at *NOS1AP* and other genes encoding intercalated disc proteins on the electrocardiographic QT interval. *A. Kapoor[1], K. Bakshy[1], L. Xu[1], P. Nandakumar[1], D. Lee[1], E. Boerwinkle[2], M.L. Grove[2], D.E. Arking[1], A. Chakravarti[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, 21205, USA; 2) Division of Epidemiology, Human Genetics and Environmental Sciences, University of Texas Health Science Center, Houston, Texas, 77030, USA.

We have previously shown that *NOS1AP* (MIM 605551), the major genetic regulator of the QT interval (MIM 610141) explaining ~1% of its variation, is localized to cardiomyocyte intercalated discs (ID) in the human heart and that its overexpression in cardiomyocytes leads to altered cellular electrophysiology. We also showed that noncoding sequence variants mapping in and around (±10kb) 175 genes encoding ID proteins are enriched for associations greater than that observed for genome-wide comparisons. Specifically, at false discovery rate (FDR) of 5% we identified 27 ID genes at which noncoding variants were significantly associated with QT interval. Thus, we sought to assess whether coding variants at these 28 ID genes, including *NOS1AP*, affect QT interval variation. We used data on exome sequencing in 4,469 ARIC cohort individuals of European ancestry to focus on 2,393 rare (MAF <0.01) missense, splice-site and stop-gain variants and tested their effect on standardized QT interval residuals corrected for heart-rate, age, sex and genotypes of 34 known common QT interval GWAS hits. Single variant association tests performed using PLINK identified 6 missense variants (3 in *TTN* (MIM 188840) and 1 each in *ANKRD30A* (MIM 610856), *LRFN2* (MIM 612808) and *SIPA1L1* (MIM 610484)) at FDR of 5%: all were observed as singletons and thus need testing in larger sample sizes for replication. Gene-based burden tests performed using seqMeta identified 3 marginally significant genes: *CAV2* (MIM 601048) (P=0.038), *LRFN2* (P=0.046) and *TLN1* (MIM 186745) (P=0.025). *TTN* encodes for titin that is a large and abundant protein of striated muscle and plays a key role in its assembly and functioning. Rare coding mutations in *TTN* are known to be associated with cardiomyopathies. *CAV2* encodes caveolin-2 that is a caveolin family member and related to *CAV1* and *CAV3*. Caveolin proteins have been proposed to act as scaffolding proteins within caveolar membranes and the 3 caveolins are known to interact in cardiac tissue. Rare coding mutations in *CAV3* are known to be associated with long QT syndrome and cardiomyopathies. *SIPA1L1* encodes a Rap GTPase activating protein that plays a role in non-canonical Wnt signaling and development. Common variants at *SIPA1L1* locus are known to be associated with QRS interval duration. In summary, using a functionally enriched gene set we have identified association of QT interval with 6 potential candidates that need to be replicated using larger sample sizes.

632T

Rare Exome Sequence Variants in *CLCN6* Reduce Blood Pressure Levels and Hypertension Risk. *B. Yu[1], S.L. Pulit[2,3], C. Newton-Cheh[3,4], A.C. Morrison[1] on behalf of the CHARGE Blood Pressure Working Group and NHLBI Exome Sequence Project.* 1) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA; 2) Department of Medical Genetics, University Medical Centre Utrecht, The Netherlands; 3) Broad Institute of Harvard and MIT, Cambridge, MA, USA; 4) Cardiovascular Research Center and Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA.

Genetic effects of rare variants on blood pressure (BP) are not yet well characterized. We performed whole exome sequencing on DNA samples from 9,950 and 4,547 individuals of European and African ancestry to examine the impact of rare variants on hypertension and four BP traits: systolic and diastolic BP (SBP, DBP), pulse pressure (PP), and mean arterial pressure (MAP). Single variant tests of common variants (minor allele frequency, MAF, ≥1%, conducted in ~170,000 common variants, statistical significance $P \leq 2.9 \times 10\text{-}7$) and gene-based tests of rare variants (MAF<1%, conducted in ~17,000 genes, statistical significance $P \leq 2.9 \times 10\text{-}6$) were evaluated for each trait and ancestry separately, followed by multiethnic meta-analyses. Rare coding variants (defined as splicing, stop-gain, stop-loss, nonsynonymous variants, or indels) in *CLCN6* were associated with lower DBP (cumulative MAF = 1.3%, b = -3.20, $P = 4.1 \times 10\text{-}6$), and the effect was consistent in both ancestries. These effects persisted after conditioning on a nearby common variant (rs17367504), known to be associated with BP. The variants in *CLCN6* were also associated with lower SBP (b = -4.11, $P = 2.8 \times 10\text{-}4$), MAP (b = -3.50, $P = 8.9 \times 10\text{-}6$), and reduced risk of hypertension (odds ratio = 0.72, P = 0.017). The association with DBP was corroborated by two independent samples of European ancestry. By meta-analyzing the two-stage discovery samples, *CLCN6* was shown to be associated with lower DBP at exome-wide significance (cumulative MAF = 1.1%, b = -3.30, $P = 5.0 \times 10\text{-}7$). Our findings implicate the effect of rare coding variants in *CLCN6* in BP variation, and by highlighting a single gene at a prior known locus offer new insights into BP regulation.

633F

First evidence of association between MTHFR C677T polymorphism and Rheumatic Mitral Value Disease. *S. Justin Carlus[1], Khalid. Al Harbi[2], Ghadeer. Saleh Mossad AlHarbi[3], Atiyeh. Abdallah[3].* 1) King Faisal Specialist Hospital & Research Center, Jeddah, Saudi Arabia; 2) Pediatric Cardiology, College of Medicine, Taibah University, Madina, Saudi Arabia; 3) Centre for Genetics and Inherited Diseases (CGID), Taibah University, Madina, Saudi Arabia.

Background: Rheumatic heart disease (RHD) is a serious complication of rheumatic fever (RF). *Plasma homocysteine (Hcy) levels are* increased in RHD patients. MTHFR catalyzes the irreversible conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate and plays a vital role in Hcy metabolism.However, the role of *MTHFR* polymorphisms in RHD has not been evaluated to date.*We therefore* conducted a case-control study to test the hypothesis that *MTHFR* C677T polymorphisms are associated with the risk of developing RHD.**Methods:** Eighty-six patients with RHD and 130 matched controls without a history of RHD were eligible for study. The diagnosis of RHD was made according to modified Jones' criteria and echocardiography. Using echocardiography, RHD patients were further divided into mitral valve lesion (MVL) and combined valve lesion (CVL) groups. *MTHFR* C677T polymorphisms were genotyped by DNA sequencing. The chi-squared test was used to evaluate differences in genotypes.**Results:** Control genotypes were in Hardy-Weinberg equilibrium. The C677T homozygous genotype (OR=4.09; 95% CIs 1.16-14.44; *P*=0.020) and recessive model (TT vs. CC+CT; OR=4.05; 95% CIs 1.17-14.04; *P*=0.019) were significantly associated with MVL RHD.**Conclusions:** This is the first study to investigate the association between the *MTHFR* C677T polymorphism and risk of RHD. The *MTHFR*C677Tpolymorphism is associated with RHD in patients with MVLs, perhaps via an Hcy-mediated cytokine effect.

634W

A phenome-wide scan to detect pleiotropic effects of the loss of function R46L variant in *PCSK9*. *M.S. Safarova[1], E.E. Austin[1], M. de Andrade[2], B.J. Coombes[2], T.G. Lesnick[2], M.S. Williams[3], M.D. Ritchie[4], K.M. Borthwick[5], S.J. Hebbring[6], E.B. Larson[7], A. Scrol[7], G.P. Jarvik[8,9], J.C. Denny[10], D.J. Schaid[2], I.J. Kullo[1], the electronic MEdical Records and GEnomics (eMERGE) Network.* 1) Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN; 2) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 3) Genomic Medicine Institute, Geisinger Health System, Danville, PA; 4) Center for Systems Genomics, Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA; 5) The Henry Hood Center for Health Research, Geisinger Health System, Danville, PA; 6) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI; 7) Group Health Research Institute, Center for Health Studies, Seattle, WA; 8) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA; 9) Department of Genome Sciences, University of Washington, Seattle, WA; 10) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN.

Objectives. Proprotein convertase subtilisin/kexin type 9 (PCSK9) plays a key role in regulating plasma low-density lipoprotein cholesterol (LDL-C) levels. Since drugs targeting PCSK9 are expected to be in clinical use in the near future, we undertook an agnostic scan of the phenome to identify pleiotropic effects of a *PCSK9* variant (R46L) known to be associated with lower LDL-C levels.**Methods.** We performed a phenome-wide association study (PheWAS) of the missense mutation R46L (rs11591147) in *PCSK9* with 1802 binary clinical phenotypes in 45,654 individuals (44.5% males) from the electronic MEdical Records and GEnomics (eMERGE) Network. A comprehensive range of phenotypes available in the electronic health record (EHR) was mapped using a hierarchical ICD-9 code-based PheWAS software (R statistical package PheWAS). A control set for each phenotype was constructed by selecting all patients that did not have the phenotype or closely related phenotypes. After standard quality control steps, R46L was imputed using IMPUTE2. Logistic regression with adjustment for age, sex, and race was used to test for associations between R46L and individual phenotypes. Bonferroni correction was applied to account for the testing of multiple phenotypes.**Results**. The study cohort included European Americans (34,960; 76.6%), African Americans (7731; 16.9%), Asian/Pacific Islanders (252; 0.6%), and individuals with unknown ethnicity (2711, 5.9%). The imputation accuracy score (INFO) for R46L was 0.73 and the overall frequency of this variant was 0.64%. In PheWAS analyses, we replicated the association of this variant with hypercholesterolemia (odds ratio (OR) 0.56; p=8.6x10-6). Additionally R46L was associated with osteoporosis (OR 1.78; p=1.3x10-5) and acute bronchitis/bronchiolitis (OR 2.63; p=1.2x10-5).**Conclusion**. In an EHR-based study, PheWAS analysis found evidence of pleiotropy for the *PCSK9* R46L variant. In addition to replicating the known association with lipid traits, we discovered novel associations of R46L with skeletal and pulmonary traits. Further validation of these findings as well as *PCSK9*-wide level analyses are underway.

635T

Leveraging the uniqueness of family studies in the search for coding variants for Blood pressure: An exome array study in African American cohorts. *X. Zhu[1], H. Wang[1], P. Nandakumar[2], F. Tekola-Ayele[3], B. Tayo[4], E. Ware[5], C. Gu[6], M. Fornage[7], S. Kardia[8], C. Rotimi[3], R. Cooper[4], D. Rao[6], A. Morrison[9], G. Ehret[2], A. Chakravarti[2].* 1) Epidemiology & Biostatistics, Case Western Reserve Univ, Cleveland, OH; 2) Johns Hopkins University School of Medicine, McKusick - Nathans Institute of Genetic Medicine, Baltimore, MD; 3) Center for Research on Genomics and Global Health, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 4) Public Health Science, Stritch School of Medicine, Loyola University Chicago, Maywood, IL; 5) School of Publish Health, University of Michigan, Ann Arbor, MI; 6) Division of Biostatistics, Washington University School of Medicine, St. Louis, Missouri; 7) Institute of Molecular Medicine and Human Genetics Center, University of Texas Health Science Center, Houston, TX; 8) Department of Epidemiology, University of Michigan, Ann Arbor, MI; 9) Human Genetics Center, University of Texas Health Science Center, Houston, TX.

High blood pressure (BP) is more prevalent, and is a major risk factor for cardiovascular disease (CVD), in African Americans than in any other US population. Although many large genome-wide association studies (GWAS) of BP have been conducted in individuals of European ancestry, studies of African Americans have lagged behind. Moreover, Family based linkage studies have been long ignored as GWASs became increasingly popular. However, family-based studies have greater power when rare variants underlie complex traits. Consequently, we performed genome-wide linkage analysis of blood pressure traits using the African-American families genotyped on the Illumina HumanExome BeadChip v1.1, for 242,901 variants in the Family Blood Pressure Program. We identified a genomic region on chromosome 1q with MLOD score 3.8 for pulse pressure, a region already implicated in prior studies of European ancestry families. Additional analyses using combined linkage and association studies identified 96 variants potentially accounting for the observed linkage evidence. Replication analysis in three independent African ancestral cohorts confirms the association evidence of these variants with all three BP traits (systolic, diastolic and pulse BP, $P < 0.039$). Single variant analysis and sliding window-based analysis both suggest the enrichment of BP variants in replication cohorts, with the most significant evidence observed in a window consisting of five variants with systolic BP (P=0.00044). Our study suggests that the chromosome 1q region harbors multiple genes contributing to BP variation and that family-based studies can be powerful in the search for rare functional variants within specific coding genes.

636F

Somatic/mosaic mutations are an important cause of sporadic vascular anomalies. *M. Vikkula[1,2,3], P. Brouillard[1], A. Mendola[1], J. Soblet[1], M. Schlögel[1], M. Amyere[1], N. Limaye[1], L.M. Boon[1,2].* 1) Human Molecular Genetics, de Duve Institute, Università catholique de Louvain, Brussels, Belgium; 2) Center for Vascular Anomalies, Division of Plastic Surgery, Cliniques universitaires Saint Luc, Brussels, Belgium; 3) WELBIO (Walloon Excellence in Lifesciences and Biotechnology), de Duve Institute, Università catholique de Louvain, Brussels, Belgium.

Purpose: Vascular anomalies are localized defects of the lymphatic or vascular system. They are most commonly observed on the skin, but can affect any body part. They are divided according to affected vessel type into arterial, capillary, venous, lymphatic and combined malformations. The etiology has remained unknown.Based on observed rare familial inheritance of some malformations, we have been able to identify underlying germline changes in several genes, including TIE2 for mucocutaneous venous malformations (VMCM), glomulin for glomuvenous malformations (GVM), and RASA1 for capillary malformation-arteriovenous malformation (CM-AVM). These familial cases are commonly characterized by multifocality of lesions that, in general, are small in size. Often, additional, small lesions develop with time. We have explained this by Knudson's double-hit theory, and demonstrated somatic second-hits in all three entities. These can be intragenic point mutations or deletions, or chromosomal anomalies, including deletions and acquired uniparental isodisomy.**Methods**: The frequent discovery of somatic 2nd hits made us study sporadic patients for eventual somatic changes. We screened frozen tissues from >100 venous malformations (VM), 30 capillary malformations (CM) and 20 lymphatic malformations (LM) for mutations in functional candidate genes. We used Ion Torrent PGM as technique to allow us to reach high vertical coverage (>1000x).**Results**: In the majority of the lesions, we have pinpointed a somatic mutation: 50% of the VMs have an activating TIE2 mutation, 50% of the CMs have an activating GNAQ mutation, and 80% of the LMs have an activating PIK3CA mutation. Syndromic forms, such as the CLOVES syndrome, is also characterized by a somatic PIK3CA mutation. Interestingly, some sporadic patients with multifocal lesions are mosaic in blood for an underlying mutation that seems to need a somatic second-hit. Moreover, in BRBN, also characterized by multifocal lesions, the somatic mutations in distant lesions are shared, even if the patient's blood is not mosaic for the underlying mutation.**Conclusion**: We conclude that somatic mutations are a common pathophysiologic cause of sporadic vascular anomalies, whether isolated or syndromic. In sporadic patients with multifocal lesions, mosaicism may play a role, yet in other, progenitor cell migration may underlie sharing of somatic mutations in distant sites.

637W

Large-scale analysis of population titin truncations reveals these variants are rarer than previously estimated. *O. Akinrinade[1], J. Koskenvuo[2,3], T. Alastalo[1,2].* 1) Children's Hospital, Institute of Clinical Medicine, University of Helsinki and Helsinki University Central Hospital Finland; 2) Blueprint Genetics, Helsinki, Finland; 3) Department of Clinical Physiology and Nuclear Medicine, HUS Medical Imaging Center, Helsinki University Central Hospital and University of Helsinki, Finland.

Background – Truncating titin (TTN) mutations account for 25% and 18% of familial and sporadic cases of dilated cardiomyopathy (DCM) respectively. However, distinguishing true pathogenic *TTN* truncating variants (TTNtv) from benign variants in clinical diagnostics and in research setting is challenging, as these variants are also present in control populations. We carried out systematic analyses of TTNtv in publicly available control populations, including, for the first time, data from Exome Aggregation Consortium (ExAC), in a bid to establish a more accurate estimate of prevalence of different TTNtv and to foster better clinical interpretation of these findings.**Methods and Results** – Using data from 1000 Genomes Project, Exome Sequencing Project (ESP) and ExAC, we estimated the prevalence of TTNtv in the population. Interestingly, 52-54% of population TTNtv were located either in exons with low expression in the left ventricle (LV) or exons present in few isoforms of the gene. The frequency of truncations affecting all transcripts in ExAC was 0.36% and 0.19% for those affecting the A-band. In the A-band region, the prevalences of frameshift, nonsense and consensus splice site variants were 0.057%, 0.090%, and 0.047% respectively. *Cga/Tga* (arginine/nonsense – R/*) transitional change at CpG mutation hotspots was the most frequent type of *TTN* nonsense mutation accounting for 91.3% (21/23) of arginine residue nonsense mutation (R/*) at *TTN* A-band region. Non-essential splice-site variants had significantly lower proportion of private variants and higher proportion of low-frequency variants compared to essential splice-site variants (respectively).**Conclusion** – Considering the rarity of A-band titin splice-site, nonsense and frameshift mutations in the population, we conclude that TTNtv are major causes of dilated cardiomyopathy. Our findings highlight the importance of the use of appropriate control population such as ExAC in filtering low frequency benign variants. Existing data supports classification of *TTN* truncations especially in A-band region likely pathogenic, and any individual tested positive should be considered to have elevated risk for dilated cardiomyopathy.

638T

Association of variation in 82 pharmacogenes with low-density lipoprotein cholesterol levels in the eMERGE-PGx project. *I.J. Kullo[1], E.E. Austin[1], M.S. Safarova[1], D.S. Kim[2,3], A.S. Gordon[2], P.D. Robertson[3], D.S. Hanna[3], D.S. Carrell[4], A. Scrol[5], M. de Andrade[6], S.J. Bielinski[6], K.F. Doheny[7], P.K. Crane[8], R. Li[9], S. Stallings[10], S.S. Verma[11], J. Wallace[11], M.D. Ritchie[11], M.O. Dorschner[3], E.B. Larson[5], D.A. Nickerson[3], J.G. Linneman[12], C. McCarty[13,14], D. Berg[14], G.P. Jarvik[2,3], D.J. Schaid[6], D.R. Crosslin[2,3], the electronic MEdical Records and GEnomics (eMERGE) Network.* 1) Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN; 2) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA; 3) Department of Genome Sciences, University of Washington, Seattle, WA; 4) Group Health Research Institute, Group Health Cooperative, Seattle, WA; 5) Group Health Research Institute, Center for Health Studies, Seattle, WA; 6) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 7) Center for Inherited Disease Research, Johns Hopkins University School of Medicine, Baltimore, MD; 8) Division of General Internal Medicine, University of Washington, Seattle, WA; 9) Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 10) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 11) Center for Systems Genomics, Department of Biochemistry and Molecular Biology, Pennsylvania State University, Pennsylvania, PA; 12) Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, WI; 13) Essentia Institute of Rural Health, Duluth, MN; 14) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI.

Background and hypothesis. In the electronic MEdical Records and GEnomics (eMERGE)-PharmacoGenomics (PGx) study, ~9000 participants have been sequenced with the targeted Pharmacogenomics Research Network sequence platform (PGRNseq). This platform was designed to sequence the coding regions, UTRs, 2kb upstream, and 1kb downstream for 84 pharmacogenes. Common variants in five genes (*LDLR*, *HMGCR*, *NAT2*, *ABCA1,* and *APOA1*) on this platform are associated with low-density lipoprotein cholesterol (LDL-C) levels in genome wide association studies.**Aim.** We investigated whether any of the 82 Pharmacogenes (in particular the five candidate genes listed above) were associated with LDL-C levels in the eMERGE PGx study.**Methods.** We ascertained the median LDL-C level in the electronic health records (EHRs) for 3256 participants with lipid panel and targeted sequencing data. Population stratification factors were generated from a pruned set of all sequenced variants (minor allele frequency ≥ 0.05; linkage disequilibrium r2 <0.50). Covariate adjustments were made for age, sex, statin use, and population stratification. Gene level statistical significance was quantified with the adaptive sum of powered score (aSPU) test. This global test identifies significant associations of both rare and common variants in the presence of numerous non-associated rare variants. The adaptive feature of the test allows it to maintain high power across varying genetic architectures.**Results.** The median LDL-C level was available for 3256 participants across five eMERGE sites (Marshfield Clinic Research Foundation, Mayo Clinic, Northwestern University, University of Washington/Group Health, and Vanderbilt University). The mean (SD) age of participants was 56.5 (9.8) years, 48.6% were male. The mean (SD) LDL-C was 114.3 (29.3) mg/dL. The top five genes (aSPU *P*-value) ranked by statistical significance were *HMGCR* (0.001), *LDLR* (0.005), *TCL1A* (0.017), *NR3C1* (0.017), and *CRHR1* (0.019). Twelve of the 82 genes were found to be significant at a 0.05 level. Variation in *NAT2, ABCA1,* and *APOA1* was not significantly associated with LDL-C levels (aSPU *P*-values 0.612, 0.519, and 0.729, respectively)**Conclusion.** We replicated the association of *HMGCR* and *LDLR* with LDL-C levels and identified additional novel associations. Further studies are underway to include the larger anticipated sample size of ~9,000 individuals for gene and variant level analyses.

639F

Defining the Genetic Landscape of Pulmonary Arterial Hypertension Using Whole-Genome Sequencing. *S. Gräf[1,2], M. Bleda[1], M. Haimel[1], C. Hadinnapola[1], N. Morrell[1] on behalf of the PAH National Cohort and the NIHR BRIDGE-PAH Consortium.* 1) Department of Medicine, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom; 2) Department of Haematology, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom.

Pulmonary arterial hypertension (PAH) is a devastating rare disease with an estimated incidence of 2–3 per million of the adult population per year. Elevated blood pressure in the lung manifests from increased narrowing and stiffening of small pulmonary arteries. Individuals tend to present with symptoms late in the disease progression and die on average only 3–5 years after diagnosis from right heart failure. About 75% of familial cases and approximately 20% of sporadic cases without a known family history can be attributed to deleterious mutations in the bone morphogenetic receptor type 2 gene. Although several additional disease genes have been identified recently using candidate gene approaches or whole-exome sequencing, a genetic cause for the majority of cases remains to be identified.We have established an international consortium comprising PAH specialist centres in the UK and abroad with the aim of sequencing the whole genomes of more than 1,000 unrelated adults and children diagnosed with PAH as part of the NIHR BioResource–Rare Diseases (BRIDGE) 10,000 genomes project. When the patients are recruited to the study extensive phenotype data is recorded from the time of diagnosis. Participants are then followed up every 6 months for 5 years to monitor disease progression. In the case of familial PAH we are also recruiting family members to the longitudinal study to assess genetic penetrance and to determine additional environmental or genetic factors involved in this disease. Here we present the first account derived from the comparison of more than 400 sequenced PAH cases with more than 3,000 sequenced BRIDGE control samples.

640W

Testing causality in the association of plasma cortisol with risk of coronary heart disease: a Mendelian randomisation study. *A.A. Crawford[1, 2], N.J. Timpson[2], G. Davey Smith[2], B.R. Walker[1].* 1) University/BHF Centre for Cardiovascular Science, University of Edinburgh, Edinburgh, UK; 2) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK.

Coronary heart disease (CHD) is one of the leading causes of death in the developed world. Elevated morning plasma cortisol is associated with multiple cardiovascular risk factors in metabolic syndrome. Epidemiological studies have also reported a positive association between plasma cortisol and coronary heart disease (CHD), although not all estimates are statistically significant (OR 1.10 (95% CI 0.97 to 1.25)) (Davey Smith *et al.* Circulation 2005). Importantly, observational studies are unable to infer causality and results may be confounded. A two-sample Mendelian randomisation approach was used to estimate the causal effect of plasma cortisol on risk of CHD. Our genetic instrument for plasma cortisol comprised of three SNPs which were associated with plasma cortisol in the recent Cortisol Network (CORNET) genome wide association meta-analysis (GWAMA) (n=12,597) (Bolton *et al.* PLoS Genetics 2014). We investigated the association between our genetic instrument for plasma cortisol and risk of CHD in up to 22,223 cases/64,762 controls from the publicly available CARDIOGRAM consortium. Each standard deviation change in genetically elevated plasma cortisol was associated with an odds ratio of 1.27 (95% CI: 1.01 to 1.60) for CHD. The inconsistent results from observational studies may be explained by: the inverse association between cortisol and obesity, which confounded the positive association of cortisol with other cardiovascular risk factors; and the use of single 'snapshot' plasma cortisol measurement rather than cumulative measure of cortisol exposure provided by genetic prediction. A bi-directional Mendelian randomisation analysis between plasma cortisol and BMI may yield greater clarity. To improve the strength of our genetic instrument an expanded CORNET GWAMA is currently underway with the aim of recruiting 30,000 individuals. However, initial results suggest that measurements of cortisol may add value to predictions of CHD risk.

641T

Possibly Pathogenic Copy Number Variations Identified in Patients with Hemorrhagic Stroke. *A. Donatti[1], R. Secolin[1], F.R. Torres[1], L.E. Ferreira[2], P.H.C. França[2], V. Nagel[2], N.L. Cabral[2], I. Lopes-Cendes[1].* 1) University of Campinas, Campinas, SP, Brazil; 2) University of Joinville Region, Joinville, SC, Brazil.

Background: Stroke is the third most common cause of death or disability worldwide. Hemorrhagic stroke (HS) occurs in 15%-20% of patients and it features the disruption of cerebral blood vessels as main clinical characteristic. There are only a few genetic studies related to HS, due to its low population frequency and high mortality rate. To our knowledge no previous study examined the possible role of copy number variations (CNVs) in HS.**Objective:** To determine if CNVs could be involved in the etiology of HS.**Methods:** We obtained DNA sample from peripheral blood of 46 patients with HS and 41 age matched individuals without HS. All samples were obtained from the Joinville Stroke Biobank, as part of a population-based epidemiologic study. We detected CNVs using the Genome-Wide Human SNP 6.0 DNA chips (Affymetrix Inc.). We evaluated the presence of population stratification in the sample by principal component analysis (PCA) using 'affy' and 'limma' packages in R software. We analyzed CNVs data using Bayesian Robust Linear Model using Mahalanobis (BRLMM) and Canary algorithms by Genotype Console®Software (Affymetrix Inc.). Furthermore, we built gene networks using METACORE TM software to analyzed possible interaction among genes found within CNVs.**Results:** PCA analysis showed that the total sample did not present population stratification, which allows unbiased comparison between patients and controls. Using the matched controls as reference, we identified 378 CNVs present **only** in patients, including 139 losses and 239 gains. The CNVs encompassed 105 genes located within duplicated regions and 57 genes within deletions. According to gene network analysis, we found CNVs in several genes with functions related to blood pressure regulation, such as *DEFB103B*, *DEFB4B*and *PRODH.* We also found CNVs in genes related to vascular processes, coagulation and hypertension, and lipids metabolism (*ANXA8*, *HGSNAT* and *NOMO1*).**Conclusion:** We describe here patients with HS and CNVs in genes related to vascular process, coagulation, hypertension and lipid metabolisms. Furthermore, these CNVs were not detected in age-matched control individuals.Supported by CNPq and BRAINN-CEPID/FAPESP, BRAZIL.

642F

Concordance in classification of hypertrophic cardiomyopathy variants is markedly higher among expert centers than among clinical labs. *A. Furqan[1], P. Arscott[2], F. Girolami[3], A. Cirino[4], M. Michels[5], S. Day[2], I. Olivotto[6], C. Ho[4], E. Ashley[7,8], C. Caleshu[7,9], SHaRe Consortium.* 1) California State University, Stanislaus, Turlock, CA; 2) University of Michigan Cardiology, Ann Arbor, MI, USA; 3) Genetics Unit, Careggi University Hospital Florence, Italy; 4) Cardiovascular Division, Brigham and Women's Hospital, Boston, MA, USA; 5) Department of Cardiology, Thoraxcenter, Erasmus Medical Center, Rotterdam, the Netherlands; 6) Referral Center for Myocardial Diseases, Azienda Ospedaliera Universitaria Careggi, Florence, Italy; 7) Stanford Center for Inherited Cardiovascular Disease, Stanford, CA, USA; 8) Division of Cardiovascular Medicine, Stanford Medical Center, Stanford, CA, USA; 9) Division of Medical Genetics, Stanford Medical Center, Stanford, CA, USA.

Background: It has become increasingly clear that there are often marked differences in how variants in Mendelian disease genes are classified by different laboratories. Efforts to improve variant classification strategies will be aided by a thorough understanding of why differences in classifications exist.**Aim:** To gain such insights, we examined discordance in sarcomere variant classifications in SHaRe, a consortium of multiple international centers with expertise in cardiomyopathy genetics. **Methods and results:** We evaluated the frequency of disagreement in variant classifications between centers submitting to SHaRe and compared it to the frequency of disagreement in classifications among clinical laboratories submitting to ClinVar. Of the 536 sarcomere variants in SHaRe, 120 were submitted by ≥1 center and 17 (14%) of those were classified differently by ≥2 centers. The frequency of discordance in ClinVar was three times higher; there were 2355 sarcomere variants, 680 had classifications by ≥1 clinical lab and 307 (45%) were discordant. We then assessed the severity of discordance as severe (i.e. (likely) benign vs. (likely) pathogenic), moderate (i.e. VUS vs. pathogenic or benign), or modest (i.e. VUS vs. likely benign or likely pathogenic). The majority of discordant classifications in SHaRe were modest (10/17), with 1/17 severe and 6/17 moderate. In contrast, discordant classifications in ClinVar were more severe; about half were moderate (155/307), 145/307 were modest and 7/307 were severe. To identify the sources of discordance in SHaRe, we compared the data and rationale that each center used in making their classifications. The frequent cause of discordance was differences in privately held data, either at the testing lab (12/17) or the clinical center (11/17). Additional causes of discordance included differences in published data used (8/17), in interpretations of the same published data (2/17), and in classification criteria (5/17).**Conclusions:** Taken together, these data suggest that variant classification can be improved by evaluation at specialized centers. The lower and less severe discordance among cardiomyopathy genetics centers may be due to classification of variants by experts or to more thorough follow-up of genetic findings including segregation analysis. Our findings also underscore the importance of sharing of privately held data as this was a frequent cause of discordance.

643W

Detection of eQTLs in Human Left Atrium and their association with Atrial Fibrillation Using Hidden Expression Factor Analysis. *M. Heydarpour, M.I. Sigurdsson, L. Saddic, T.W. Chang, S.K. Shernan, J.D. Muehlschlegel, S.C. Body, Center for Perioperative Genomics.* Anesthesiology, Brigham and Women's Hospital, Boston, MA.

Introduction: Risk alleles for atrial fibrillation (AF) have been identified by genome wide association study in ambulatory and surgical populations, notably found in or near the *PITX2*, *KCNN3* and *ZFHX3* genes. However, these associations have, to date, failed to be supported by biological mechanisms. We hypothesized that these variants may alter human left atrial gene expression as a mechanism in the genesis of AF. **Methods:** In 62 Caucasian patients in sinus rhythm, left atrial (LA) tissue samples were excised prior to mitral valve surgery and analyzed using mRNASeq. Quality control of gene expression data was performed by MatrixeQTL (Shabalin, 2012). Patients were genotyped using the HumanOmni2.5 array. Twenty one patients (34%) had AF postoperatively, and were generally older with a higher frequency of hypertension. We performed gene-expression differential analysis between 21 AF patients and 41 no-AF to detect genes with significant differences. Whole genome eQTL analysis was executed for 17,925 genes and 1,243,383 genetic-markers. A hidden expression factor analysis (HEFT) method (Gao et al. 2014) was used for detecting eQTLs while accounting for age, gender, BMI, and cardiopulmonary bypass time. Hidden factors selected by assessing the eigen spectrum of the overall gene expression covariance (>5% of total variance). **Results:** We identified 407 genes which significantly differential expressed [log2(expr-AF/expr-noAF) >0.5] between patient with AF and no-AF. 5,628 cis-eQTLs and 21,406 trans-eQTLs significant (P<1.0E-8) were identified through whole genome eQTL analysis. After integrated these results with 407 genes, we recognized 192 cis-eQTLs in 15 genes were overlapped. The most significant cis-eQTLs were identified for *ERAP2* (5q15), *MYOZ1* (10q22.1), *HLA-DQB1* (6p21.3), and *HLA-DQA1* (6p21.3). SNPs in each region were in high linkage disequilibrium (LD). We did not detect any significant eQTL for genes previously associated with AF such as *PITX2*, *KCNN3* and *ZFHX3.* The *SCN10A* gene previously associated with AF in ambulatory populations and PR interval was also significant (P=1.0E-10) in this study.**Conclusion:** We present a comprehensive eQTL analysis of gene expression in the human left atrium. These observations may help establish mechanisms for how common genetic variation predisposes to AF. The method that we used (HEFT) is able to reveal significant cis-eQTLs that are not detectable by linear regression, while providing strict false positive control.

644T

Co-Aggregation of Depression and Cardiovascular Diseases. *S. Knight[1,2], H.T. May[1], H. Coon[2,3], B.D. Horne[1,2].* 1) Intermountain Heart Institution, Intermountain Medical Center, Murray, UT; 2) Division of Genetic Epidemiology, University of Utah, Salt Lake City, UT; 3) Department of Psychiatry, University of Utah, Salt Lake City, UT.

Introduction: Studies have found greater risk of coronary artery disease (CAD) in individuals with depression, and greater risk of suicides in individuals who had a myocardial infarction (MI). These findings, and the fact that depression, CAD, and MI have strong genetic components, suggest possible co-aggregation of these diseases. The purpose of this project was to test for significant co-aggregation of depression with CAD and MI using a large genealogy registry, the Intermountain Genealogy Registry.Methods: The Intermountain Genealogy Registry (IGR) is a database of genealogical information from over 23 million individuals who lived or whose blood relatives lived in the Western Intermountain region of the United States. The IGR is linked to clinical records of over 650,000 (20%) patients in the Intermountain Healthcare system, an integrated healthcare delivery system which provides the majority (60%) of healthcare services in Utah. Two-way Genealogy Indices of Familiality (GIF) were generated for depression and CAD, and for depression and MI. One-way GIFs were also computed for depression, CAD, and MI individually. The GIF statistic is a measure of average relatedness (10,000x average kinship coefficient for all pairs) of individuals with a specified diagnosis in a population or a set of diagnoses. Significance was determine empirically using GIFs for age and sexed matched control sets (n=1000) of Intermountain patients in the IGR. Patients linked to the registry without ancestors were excluded, as these would not provide information for the GIF. The average GIF of the controls sets are reported. Results: There were 11,712 patients in IGR who had depression, 70,631 with CAD, and 16,113 with MI. After excluding founders, the sample sizes decreased to 7,780, 45,312, and 10,361, respectively. The one-way GIFs for each disease were 0.061 (control GIF=0.056; p=0.015) for depression, 0.085 (control GIF=0.048; p<0.001) for CAD, and 0.075 (control GIF=0.069; p<0.001) for MI. The two-way GIF for depression and CAD was 0.065 (control GIF=0.051; p<0.005), and for depression and MI it was 0.069 (control GIF=0.060; p<0.001).Conclusions: Significant co-aggregation of depression with both CAD and MI was found in a large clinical population genealogy. This supports previous evidence that these diseases may have a shared genetic cause and suggests further joint study of their genetic basis.

645F

Gene Frequencies of the *GSTT1* and *GSTM1* genes in Ramgarhia Sikh population of Amritsar district and association with Coronary Artery Disease. *N. Mahajan, G. Gandhi.* Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India.

Introduction: Cardiovascular diseases (CVDs), the diseases of the heart and blood vessels, are predominant causes of morbidity and mortality and it has been estimated that by 2030 there will be >50% deaths due to cardiovascular diseases among non-communicable diseases. The types of CVD include heart beat irregularities, cardiac failure, coronary artery disease and sudden cardiac death. In coronary artery disease, damage of the coronary arteries lead to atherosclerosis/arteriosclerosis and so the disease is known as atherosclerotic heart disease or coronary artery disease (CAD) or coronary heart disease. Detoxification of metabolites produced by oxidative stress within the cell is carried out by **genetic variants of xenobiotic-metabolizing enzymes, such as** Glutathione S-transferases (GSTs) which protects the cells against injury and **may be associated with the risk of coronary artery disease (CAD)**. For the present study the contribution of polymorphism in the *GSTT1* and *GSTM1* genes in the development of coronary artery disease in *Ramgarhia Sikh* population of Amritsar district has been investigated.**Methodology:** DNA samples from CAD patients (n=200) and age-, sex- matched healthy controls (n=200) were genotyped for *GSTT1* and *GSTM1* polymorphism using multiplex polymerase chain reaction (PCR). Genotypic frequency of CAD patients were compared with genotypic frequency of controls using chi-square test.**Results**: Results of the present study showed significant (p<0.05) difference for genotypic frequencies of *GSTT1* and *GSTM1* genotype in cases as compared to controls. *GSTT1* null genotype was 2.3-folds higher in controls as compared to CAD patients, whereas *GSTM1* null genotype was 1.7-folds higher in CAD patients as compared to controls.**CONCLUSION**: *GSTT1* null genotype showed protection against CAD whereas *GSTM1* null showed susceptibility to CAD in *Ramgarhia Sikh* population of Amritsar.

646W

Differential genetic expression in ECG characterized normokinetic and akinetic/dyskinetic myocardium zones in humans. *O.A. Makeeva[1,3], V.E. Babokin[2,4], E.V. Kulish[1], V.M. Shipulin[2], V.P. Puzyrev[1].* 1) Research Institute of Medical Genetics, Tomsk, Russia; 2) Research Institute of Cardiology, Tomsk, Russia; 3) Research Institute for Complex Issues of Cardiovascular Diseases, Kemerovo, Russia; 4) Moscow Regional Research and Clinical Institute (MONIKI), Moscow, Russia.

The aim of the study was to investigate expression profiles in different zones of human left ventricular myocardium characterized by echocardiography (ECG) as (1)akinetic, (2)dyskinetic, (3)hypokinetic, and (4) normokinetic in patients who survived myocardial infarction.**Methods:** Myocardial biopsy was taken with diagnostic purpose during cardiosurgery for aneurism resection and left ventricular reconstruction. Total of 24 biopsy samples were investigated and divided into groups according to ECG characteristics.Total RNA was extracted using RNAeasy kit (Qiagen,Germany). Quality of RNA samples was evaluated using Agilent Total RNA 6000 chip (Agilent Technologies). 400 ng of total RNA were amplified with Illumina TotalPrep™ RNA Amplification Kit (Ambion). Whole genome expression analysis was conducted using Illumina HT-12v4 BeadChip. Data collection and analysis was performed using Illumina GenomeStudio, Gene Expression Module.**Results:** Samples from different zones of myocardium were compared in respect to differential expression. Samples from normokinetic zones differed from dyskinetic in expression of *FAM43B*, *FAP, GRB14*,and *SERPINEF1*. Normokinetic samples differed from akinetic in expression of the *GRB14* only. When akinetic and dyskinetic samples were combined in one group and compared to normokinetic, 3 genes with differential expression were detected: *CNN1, DIO2* and *GRB14*.**Discussion:** ECG characteristics well correspond to functional status of myocardium. Normokinetic zones present preserved wall thickness and normal contractility; hypokinetic zones correspond to reduced wall thickening and/or reduced longitudinal shortening in systole. Akinetic/dyskinetic segments are suggestive of myocardial scar due to old infarction.Normal myocardium had higher expression of *GRB14*. *GRB14* encodes a growth factor receptor-binding protein that interacts with insulin receptors and insulin-like growth-factor receptors. *GRB14* plays a role in signaling pathways that regulate growth and metabolism. Increased *GRB14* expression was previously described in adipose tissues of model animals and obese humans. *GRB14* associated with HDL-cholesterol, triglycerides, and fasting insulin gene expression levels in human tissues.Reviled differences in gene expression can lead developing new drugs to help post-infarction treatment and rehabilitation. Investigations are needed to explore whether the stimulation of *GRB14* expression in post-infarction myocardium can improve patients outcome.

647T

Heritability of age of onset of cardiovascular disease using large-scale Finnish health registry and genome-wide data. *C. Benner[1], M. Pirinen[1], V. Salomaa[3], J. Palmgren[1,2], S. Ripatti[1,4].* 1) Institute of Molecular Medicine Finland, Helsinki, Finland; 2) Karolinska Institutet, Stockholm, Sweden; 3) National Institute for Health and Welfare, Helsinki, Finland; 4) Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

Whole genome sequencing is currently being performed on thousands of Finns as part of the Sequencing Initiative Suomi project (SISu). Combining genomic profiles of tens of thousands of Finns with health registry data provides a unique resource in the world to accomplish breakthroughs in modern human genomics with possibly important implications on public health. In particular, the information can be utilized to determine to what extent phenotypic variation in age of diseaseonsetis attributable to genetic effects. Recent methodological development has enabled heritability estimation for quantitative traits in population cohorts by estimating the proportion of genetic sharing among pairs of individuals from genome-wide marker data. The possibility to use population cohorts rather than family structures to assess heritability opens up a multitude of new avenues for genetic research of age of disease onset. However, connecting age of onset to big genomics data form large-scale population cohorts has so far not been feasible with hitherto existing methods.

We introduce a novel method that deals with prospective follow-up information from health registry data and permits heritability estimation and association testing based on millions of genetic markers and health registry events in tens of thousands of individuals with more than 20 years of follow-up time. Our approach implements a very flexible piecewise constant hazard model that contains an individual-specific Gaussian random effect with an arbitrary covariance structure. Computationally, we analyze the model in a Bayesian context by using a Markov Chain Monte Carlo algorithm. We compared our method on simulated data to a mixed effects Cox model, as implemented in the R package ‘coxme’. Compared to the mixed effects Cox model, our method produced lower mean squared error of variance and regression parameters under varying proportion of variance explained in the genotype, different number of simulated causal variants and sample size as well as high censoring rates. Using over 10000 unrelated Finns from the National FINRISK study, we also estimated that the heritability of cardiovascular disease (defined as coronary heart disease and ischaemic stroke) in terms of relative risk of disease due to genetic effects is 27% (95% CI: 6-51%).

648W

Genetic polymorphism of *P2RX7* can be a predictive biomarker for responsibility to long-term effect of infliximab against Crohn's disease. *M. Yoshimura[1], D. Mukai[1], Y. Fukumitsu[1], T. Inamine[1], S. Kondo[1], S. Urabe[2], K. Matsushima[2], H. Isomoto[2], T. Ishida[3], R. Uehara[2], T. Honda[2], H. Minami[2], H. Machida[2], Y. Akazawa[2], C.C. Chen[2], N. Yamaguchi[2], K. Ohnita[2], F. Takeshima[2], K. Nakao[2], K. Tsukamoto[1].* 1) Dept Pharmacotherapeutics, Nagasaki Univ Grad Sch Biomed Sci, Nagasaki, Japan; 2) Dept Gastroenterology and Hepatology, Nagasaki Univ Grad Sch Biomed Sci, Nagasaki, Japan; 3) Oita Red Cross Hospital, Oita, Japan.

Purpose: Infliximab (IFX) is a chimeric anti-tumor necrosis factor-a monoclonal antibody exerting the therapeutic effect against Crohn's disease (CD). However, some CD patients, who treated with IFX, showed the secondary loss of response to IFX at long-term treatment of 1 year although they have achieved remission at short-term treatment of 10 weeks. Therefore, in order to identify any biomarkers to predict the therapeutic effect of IFX, we investigated a candidate gene-based association study between single nucleotide polymorphisms (SNPs) in a cell surface ATP receptor, P2X7 receptor (P2RX7, encoded by *P2RX7*) and response to IFX at the 1-year treatment period in Japanese CD patients.**Methods:** A total of 116 CD patients, who showed response to IFX at the 10-week treatment period in Oita Red Cross Hospital or Nagasaki University Hospital, were divided into two groups, responders and non-responders, based on the presence of IFX effect at the long-term period of 1 year after IFX administration. Eleven tag SNPs in *P2RX7* were analyzed by PCR-restriction fragment length polymorphism or -high resolution melting curve analysis. The frequencies of alleles and genotypes of each SNP between responders and non-responders at the 1-year treatment period were compared in three different inheritance models using chi-squared or Fisher's exact test.**Results:** The genetic analyses indicated that the frequency of a minor allele homozygous G/G genotype of rs3751143 in *P2RX7* in the minor allele recessive model was significantly decreased in responders in comparison to that in non-responders (P = 0.001, odds ratio = 0.143), implicating ~7.0-fold secondary loss of response to IFX. Conversely, a T/T or T/G genotype of rs3751143 indicated 7.0-fold response to IFX.**Conclusions:** Our results suggest that the G/G genotype of rs3751143 in *P2RX7* may affect the gain-of-function of P2RX7, thus leading to the activation of the signals in the Ras signaling pathway in the intestines of CD patients. The activation of Ras signals can lead to the perpetuation of the chronic intestinal inflammatory process, thereby resulting in the secondary loss of response to IFX at the 1-year treatment period. Thus, *P2RX7* appear to be a genetic determinant for response to IFX at the 1-year treatment period against Japanese CD patients, and to be a target molecule for new therapeutic drugs overcoming the secondary loss of response to IFX treatment.

649T

Association between *KCNJ6* (*GIRK2*) gene polymorphism rs2835859 and postoperative analgesia, pain sensitivity, and nicotine dependence. *D. Nishizawa[1], K. Fukuda[2], S. Kasai[1], Y. Ogai[1], J. Hasegawa[1], N. Sato[3,4], H. Yamada[4], F. Tanioka[5], H. Sugimura[4], M. Hayashida[6], K. Ikeda[1].* 1) Psychiatry and Behavioral Science (Addictive Substance Project), Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 2) Department of Dental Anesthesiology, Tokyo Dental College, Tokyo, Japan; 3) Department of Clinical Nursing, Hamamatsu University School of Medicine, Hamamatsu, Japan; 4) Department of Tumor Pathology, Hamamatsu University School of Medicine, Hamamatsu, Japan; 5) Department of Pathology, Iwata City Hospital, Iwata, Japan; 6) Department of Anesthesiology & Pain Medicine, Juntendo University School of Medicine, Tokyo, Japan.

Background: G-protein-activated inwardly rectifying potassium (GIRK) channels are expressed in many tissues and activated by several Gi/o protein-coupled receptors, such as opioid and dopamine receptors, and thus are known to be involved in the modulation of opioid-induced analgesia, pain, and reward. **Methods**: We focused on a GIRK channel subunit that plays a pivotal role in the brain, GIRK2, and investigated the contribution of genetic variations of the *GIRK2* (*KCNJ6*) gene to individual differences in the sensitivity to opioid analgesia using a two-stage analysis in 355 healthy patients who were scheduled to undergo cosmetic orthognathic surgery for mandibular prognathism at Tokyo Dental College Suidoubashi Hospital and provided informed, written consent for the genetics studies, followed by additional association analyses using other samples. **Results**: In our initial linkage disequilibrium (LD) analysis, a total of 27 single-nucleotide polymorphisms (SNPs) were selected in the whole exon and intron region as well as 5'- and 3'- flanking regions (~10kbp) of the *KCNJ6* gene. Among them, the rs2835859 SNP was selected in the exploratory study as a potent candidate SNP for the further confirmatory association study. The association with opioid analgesic sensitivity was confirmed for this SNP, in which the carriers of the C allele of this SNP required less analgesics compared with non-carriers after painful cosmetic surgery. We also found that the carriers of the C allele of this SNP were less sensitive to both cold and mechanical pain in healthy volunteers. Moreover, we found that homozygous carriers of the C allele of this SNP showed higher score in Tobacco Dependence Screener (TDS), an index of nicotine dependence, and required a greater number of trials until attaining successful smoking cessation. **Conclusions**: The rs2835859 SNP could serve as a marker that predicts sensitivity to analgesic and pain and severity of nicotine dependence.

650W

Genetic polymorphisms in the neuroplasticity-related genes contribute to the therapeutic effect of antidepressants for major depression. *Y. Kawafuchi[1], T. Kurokawa[1], Y. Arata[1], K. Keya[1], T. Araki[1], S. Ijichi[1], T. Inamine[1], S. Kondo[1], N. Kurotaki[2], K. Nishiwaki[3], K. Tsukamoto[1].* 1) Dept Pharmacotherapeutics, Nagasaki Univ Grad Sch Biomed Sci, Nagasaki, Japan; 2) Dept Neuropsychiatry, Nagasaki Univ Grad Sch Biomed Sci, Nagasaki, Japan; 3) Nishiwaki Hospital, Nagasaki, Japan.

Purpose: Major depression (MD) is a common psychiatric disorder in substantial proportion of the worldwide. As MD is a multifactorial disease, a variety of genetic and environmental factors contribute to its etiology and development. Neuroplasticity and neuroinflammation contribute to the pathogenesis of MD as well as responsibility to antidepressants. We investigated a candidate gene-based association study between single nucleotide polymorphisms (SNPs) in neuroplasticity-related genes (*GRIN1*, *MAPK1*, *ARC*, *FGFR1*, *FRS2*, *FRS3*, *BDNF*) and response to antidepressant(s) at the short-term treatment of 8 weeks in Japanese MD patients.**Methods:** A total of 105 MD patients were treated with antidepressants including serotonin reuptake inhibitors (SSRIs) and serotonin noradrenaline reuptake inhibitors (SNRIs) in Nishiwaki Hospital, and were divided into two groups, responders and non-responders, based on the presence of SSRI or SNRI effect at the 8-week treatment period. Thirty-four tag SNPs among 7 candidate genes were analyzed by PCR-restriction fragment length polymorphism, -direct DNA sequencing, or -high resolution melting curve analysis. The frequencies of alleles and genotypes of each SNP between responders and non-responders to SSRIs or SNRIs were compared in three different inheritance models using chi-squared or Fisher's exact test.**Results:** The frequency of a heterozygous C/T or a minor homozygous T/T genotype of rs6474354 in *FGFR1* in the minor allele dominant model was significantly increased in responders to SSRIs in comparison to that in non-responders ($P = 0.044$, odds ratio (OR) = 6.61), showing ~6.6-fold response to SSRIs. Conversely, the frequency of a minor homozygous T/T genotype of rs3804281 in *FRS3* in the minor allele recessive model was significantly decreased in the responders to SSRIs as compared with that in non-responders ($P = 0.007$, OR = 0.08), indicating ~12.5-fold loss of response to SSRIs. **Conclusions:** Our results suggest that the activation of the signals in the FGF signaling pathway due to these polymorphisms of *FGFR1* and *FRS3* may accelerate synaptic plasticity and neurogenesis induced by SSRIs, thereby resulting in good response to SSRIs. Thus, *FGFR1* and *FRS3* appear to be genetic determinants of responsibility to SSRIs in Japanese MD patients. Moreover, these molecules in the FGF signaling pathway may become targets for novel drug discovery for MD patients.

651T

Genetic risk factors for β-lactam antibiotic-induced cutaneous adverse drug reactions in Japanese population. *T. Ozeki[1], T. Mushiroda[1], A. Takahashi[2], M. Kubo[3].* 1) Laboratory for Pharmacogenomics, RIKEN Cen. Integr. Med. Sci., Yokohama, Kanagawa, Japan; 2) Laboratory for Stastical Analysis, RIKEN Cen. Integr. Med. Sci., Yokohama, Kanagawa, Japan; 3) Laboratory for Genotyping Development, RIKEN Cen. Integr. Med. Sci., Yokohama, Kanagawa, Japan.

β-lactam antibiotics, e.g. penicillin and cephem, are known to show incidences of cutaneous adverse drug reactions (cADRs) including Stevens-Johnson syndrome (SJS), toxic epidermal necrolysis (TEN) and drug-induced hypersensitivity syndrome (DIHS). To identify a gene(s) susceptible to β-lactam antibiotic-induced cADRs, we conducted a genome-wide association study (GWAS) and subsequent HLA typing in 92 cases of β-lactam antibiotic-induced cADRs and 879 subjects of a general population in Japanese. Among the SNPs analyzed in the GWAS, suggestive SNPs were located in the *HLA* loci. Thus, we genotyped the individual *HLA-A*, *-B*, *-C* and *–DRB1* alleles and found that *HLA-B*44:03* and *HLA-DRB1*13:02* was present in 31.5% (29/92) and 25.6% (23/90) of the cADR cases, but in 12.1% (106/879) and 10.5% (92/879) of the general population controls, respectively (odds ratio (OR) = 3.4 and 2.9, 95% confidence interval (CI) = 2.1-5.4 and 1.7-4.9, $P = 3.91 \times 10\text{-}6$ and $1.24 \times 10\text{-}4$, respectively).

652W

PCSK9 variants are associated with LDL-C response to statin therapy in African-Americans. *Q. Feng[1], WQ. Wei[2], CP. Chung[3], R. Levinson[4], L. Bastarache[2], JC. Denny[2], CM. Stein[1,5].* 1) Division of Clinical Pharmacology, Department of Medicine, Vanderbilt University, Nashville, TN; 2) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 3) Division of Rheumatology, Department of Medicine, Vanderbilt University, Nashville, TN; 4) Vanderbilt Genetic Institute, Vanderbilt University, Nashville, TN; 5) Department of Pharmacology, Vanderbilt University, Nashville, TN.

Background Statins (HMG-CoA reductase inhibitors) lower low-density lipoprotein cholesterol (LDL-C) and prevent cardiovascular disease (CVD). Despite the overall benefits of statins, there is wide individual variation in LDL-C response. Drugs targeting proprotein convertase subtilin/kexin type 9 (PCSK9) lower LDL-C and are likely to be used with statins. PCSK9 plays a critical role in the degradation of LDL receptors; statins up-regulate both LDL receptors and PCSK9. Therefore, individuals with loss-of-function (LOF) *PCSK9* variants would be expected to have an increased LCL-C response to statins because LDL receptor degradation would be reduced.The gene encoding PCSK9 is highly polymorphic and several LOF and gain-of-function (GOF) variants have been identified. However, the relationship between *PCSK9* variants and statin response remains unclear. We examined the hypothesis that functional *PCSK9* variants affect statin response. **Methods** We studied 669 African Americans (AAs) treated with a statin in clinical practice. The LDL-C response to statin treatment (difference between natural log of baseline and on-treatment LDL-C) was extracted from the electronic health record. Genotypes for 11 *PCSK9* variants known to be functional were obtained using the MetaboChip. Association analysis was performed with adjustment for age, sex, baseline LDL-C, principal components of ancestry, statin type and dose with a Bonferroni-adjusted significance threshold of $p< 0.0045$. We also analyzed the collective effect of 11 known *PCSK9* functional variants using a burden test.**Results** One LOF variant, rs11591147 (p.R46L) was significantly associated with LDL-C response to statin (p=0.00236). In the 3 patients with the minor allele of the variant there was a 55.6% greater LDL-C reduction compared to non-carriers. This SNP has previously been associated with LDL-C response to rosuvastatin in 6989 Caucasians (JUPITER) with locus-wide significance. Another functional variant, rs298362261(p.N425S), which has been associated with baseline LDL-C in a large AA cohort, was associated with statin response with a borderline p-value (p=0.0064). The collective effect of the 11 known *PCSK9* functional variants was also of borderline significance (P=0.06).**Conclusion** Some functional *PCSK9* variants are associated with the LDL-C response to statin therapy. Further analysis in larger populations of African Americans will be important for a more thorough evaluation of the effect size of these SNPs.

653T

Identification of genetic and environmental factors that influence time to achieve target INR and time in therapeutic range with warfarin. *M.R. Botton[1], P.P. Viola[1], M.R. Meireles[1], P. Zuchinali[2], E. Bandinelli[1], L.E. Rohde[2], T.L.L. Leiria[3], J.Y.Y. Salamoni[2], A.P. Garbin[2], M.H. Hutz[1].* 1) Departamento de Genética, UFRGS, Porto Alegre, Rio Grande do Sul, Brazil; 2) Serviço de Cardiologia, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brasil; 3) Instituto de Cardiologia - Fundação Universitària de Cardiologia, Porto Alegre, Rio Grande do Sul, Brasil.

Warfarin is an oral anticoagulant widely prescribed to prevent and treat thromboembolic disorders. It has a narrow therapeutic window and must have its effect controlled. Prothrombin test, expressed as INR values, is used for follow up. Time to reach target INR and time in therapeutic range (TTR) are important outcomes and several factors might influence them. The aim of this study was to identify genetic and environmental factors that can potentially influence time to reach target INR and TTR. A total of 422 patients using warfarin were investigated. Genotypes for VKORC1 -1639G>A, CYP2C9*2 and CYP2C9*3 were detected by TaqMan assays by real-time PCR. Furosemide co-medication and presence of chronic renal failure were associated with more time to reach target INR whereas VKORC1 -1639GA and -1639AA genotypes were associated with less time to reach INR. Regarding TTR, glibenclamide co-medication, male gender and presence of CYP2C9*2 allele were associated with more percentage of TTR, while amiodarone and enalapril co-medication were associated with less percentage of TTR. Our results suggest that TTR is not influenced by the same factors that alter time to reach INR. Other studies must be developed to confirm these results.

654W

SNP-based HLA tagging, imputation, and association in adverse drug reaction of epilepsy patients from Hong Kong. *S.S. Cherny[1 2 3], H. Gui[2], M. Kwok[4], P.C. Sham[1 2 3], L. Baum[1], P. Kwan[4 5].* 1) Department of Psychiatry, The University of Hong Kong; 2) Centre for Genomic Sciences, The University of Hong Kong; 3) The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong; 4) Department of Medicine & Therapeutics, The Chinese University of Hong Kong; 5) Department of Neurology, Royal Melbourne Hospital, Melbourne, Australia.

Human leukocyte antigen (HLA) genes control the regulation of the human immune system and are involved in infection and autoimmune-related complex diseases. Some HLA alleles (B*15:02, A*31:01 and B*57:01) have been associated with increased risk of cutaneous adverse drug reaction (cADR) in epilepsy patients taking antiepileptic drugs like carbamazepine and abacavir. However, traditional serotyping or Sanger sequencing for HLA genotyping may not be cost-effective for clinical application. A more efficient way may be to look for single nucleotide polymorphisms (SNP) that can tag HLA risk alleles with high sensitivity and specificity. Since haplotype structure is population-specific due to different linkage disequilibrium patterns, an investigation on pairwise SNP-HLA relationships was conducted in the Hong Kong population. This analysis arose from our program of research on epilepsy genetics and pharmacogenomics, for which we have built an in-house database including classical HLA serotyping, genome-wide SNP arrays, exome sequencing, and whole-genome sequencing. The discovery panel employed included 144 patients with both genome-wide SNP-array data and HLA*B or HLA*A serotypes, which enable a greedy search of highly correlated pairs of HLA risk alleles and SNPs through use of pairwise LD statistics. Common variants rs9265348, rs2532923, and rs114025781 were found to tag HLA risk alleles B*15:02, B*58:01, and B*13:01, respectively, with 100% sensitivity and >95% specificity. No HLA-A risk allele was found to be tagged by any single SNP. The results are being replicated in an independent sample, for which both HLA and SNP genotypes can be called directly from exome or whole genome sequencing. In addition, SNP2HLA was used to construct a Hong Kong population-specific reference dataset for HLA imputation. Using the GWAS sample, complete HLA-A and HLA-B allele information was inferred for a total of 408 epilepsy patients whose cADR status upon taking aromatic antiepileptic drugs was recorded. To determine how HLA genes affect cADR individually and jointly, this data will be combined with Caucasian populations from the International League Against Epilepsy Consortium on Complex Epilepsies for an overall trans-ethic meta-analysis.

655T

Differential allele and haplotype frequencies in the ADME genes *ABCB1* and *NAT2* in Mexican population. *A.V. Contreras, V. Bonifaz-Peña, J.C. Fernández-López, M.E. Tejero.* National Institute of Genomic Medicine, Mexico City, Mexico.

Genetic polymorphisms in genes related to drug absorption, distribution, metabolism and excretion (ADME genes) influence how humans respond to commonly used drugs. However, the effect of the polymorphisms to drug response can differ in magnitude or can be absent depending on the ethnic origin of the population being studied. Most of the Mexican population results from an admixture of any of 65 Native American groups with European ancestry individuals, and to a lesser extent, Africans. To characterize the distribution of relevant genetic markers for pharmacogenomics in *ABCB1* and *NAT2* in Mexican population, we genotyped eight SNPs and inferred haplotypes from individual genotypes. The sample included 250 self-identified Mestizo individuals (MES) from 5 states located in geographically distant regions in Mexico and 45 Native American Zapotecos (ZAP). We evaluated the extent of genetic differentiation (measured as Fst values) of the genetic markers genotyped between Mexican and parental populations. We observed alleles showing large frequency differences in both genes studied. The *ABCB1* 2677G>A/T (rs2032582) TT and the *ABCB1* 3435C>T (rs1045642) TT genotypes was enriched in ZAP (frequency: 46.7 and 53.3%, respectively) compared to MES (frequency: 14.3 and 20.2%, respectively). Moreover, comparative analysis with parental populations evidenced different allele frequencies, including the *ABCB1* 1236C>T (rs1128503) TT genotype (frequency: 11.86% in CEU, 48.88% in ZAP, and 22.32% in MES). In MES and ZAP, the *NAT2*4* (rapid phenotype) was highly prevalent (frequency: 39.2 and 51.5%, respectively), and the frequency showed regional variations related to ancestry. Haplotype distribution for *NAT2* alleles associated with slow phenotype, such as *NAT2*5B*, *NAT2*6A* and *NAT2*7B*, also showed differences between distant regions in Mexico and parental populations. Our results indicate differences in the distribution of allele and haplotype frequencies by the admixture history of Mexican population. Thus, these analyses may be of relevance to better design pharmacogenomic studies leading to a more rational use of drugs in the Mexican population.

656W

Molecular Markers of Blood Pressure Response to Thiazide Diuretics Identified Through Whole Transcriptome RNA-Seq Analysis. *A. Costa Sa[1, 2], A. Webb[3], Y. Gong[1], C.W. McDonough[1], T.Y. Langaee[1], S.T. Turner[4], A.L. Beitelshees[5], A.B. Chapman[6], E. Boerwinkle[7], J.G. Gums[8], S. Scherer[9], R.M. Cooper-DeHoff[1, 10], W. Sadee[3], J.A. Johnson[1, 2, 10].* 1) Center for Pharmacogenomics and Department of Pharmacotherapy and Translational Research, College of Pharmacy, University of Florida, Gainesville, FL; 2) Graduate Program in Genetics and Genomics, Genetics Institute, University of Florida, Gainesville, FL; 3) Department of Pharmacology, College of Medicine, Ohio State University, Columbus, OH; 4) Division of Nephrology and Hypertension, Mayo Clinic, Rochester, MN; 5) Division of Endocrinology, Diabetes and Nutrition, University of Maryland, Baltimore, MD; 6) The Renal Division, Department of Medicine, Emory University, Atlanta, GA; 7) Division of Epidemiology, University of Texas at Houston, Houston, TX; 8) Department of Community Health and Family Medicine, University of Florida College of Medicine, Gainesville, FL; 9) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 10) Division of Cardiovascular Medicine, Department of Medicine, University of Florida, Gainesville, FL.

Hypertension (HTN) is the most significant risk factor for cardiovascular disease, affecting about 1 billion individuals worldwide. Thiazide diuretics are among the most commonly prescribed first line anti-HTN medications, however less than 50% of treated patients achieve BP control. Therefore, the purpose of this study is to identify novel molecular determinants of thiazide diuretics BP response by exploring transcriptome profiling in hypertensive participants treated with hydrochlorothiazide (HCTZ) and chlorthalidone from PEAR (Pharmacogenomics Evaluation of Antihypertensives Response) and PEAR-2 studies, respectively. This study assessed whole genome expression levels in whole blood samples from 150 participants with extremes of BP response (responders and non-responders) to thiazide diuretics (50 whites from PEAR; 50 whites and 50 blacks from PEAR-2). RNA-Seq generated reads (100 bp) were aligned to the human reference genome (Hg19) with TopHat2. Abundance comparisons between responders and nonresponders were carried out using Cufflinks and reported as fragments per kilobase per million reads (FPKM). A t-test was performed to evaluate significance of the difference in FPKM (responders/non-responders). In addition, genome-wide genotyping was determined in 228 white participants, from PEAR and 417 participants from PEAR-2 via Illumina Human Omni 1M-Quad Chip and 2.5M-8 BeadChip, respectively. At FDR adjusted p-value <0.05, we identified 14, 20 and 55 genes that were differentially expressed in relation to HCTZ BP response in whites, and chlorthalidone BP response in whites and blacks, respectively. For each differentially expressed gene in PEAR or in PEAR-2 whites and blacks, we attempted replication in the other race and drug groups. *CEBPD* and *TSC22D3* were differentially expressed in all 3 cohorts, while *SERINC5, PPP2R5C, TFCP2* and *FOS* identified in the PEAR-2 whites participants' cohort, replicated only in PEAR-2 blacks. Of note, rs1004249 was the top significant genetic variant in *SERINC5* where variant allele carriers had poor chlorthalidone BP response vs non-carriers (P=1.8x10-4; b=-3.1 mmHg). Additionally, rs1203488 in *PPP2R5C* was also associated with chlorthalidone BP response in whites (P=3.7x10-4; b=-4.0 mmHg). Further evaluation of these genes may provide new insights into molecular mechanisms underlying BP response to thiazide diuretics.

657T

Distribution of allele frequencies for clinically relevant pharmacogenes in Mexican population. *J.C. Fernandez-Lopez[1], V. Bonifaz-Peña[1], A.V. Contreras[2], A. Hidalgo-Miranda[3].* 1) Computational Genomics Consortium, Instituto Nacional de Medicina Genómica, Mexico City; 2) Nutrigenetics and Nutrigenomics Laboratory, Instituto Nacional de Medicina Genómica, Mexico City; 3) Oncogenomics Consortium, Instituto Nacional de Medicina Genómica, Mexico City.

Pharmacogenomics markers can differ in magnitude or be completely absent depending on the population being assessed. Recently, the frequency distribution of relevant polymorphisms in Absorption, Distribution, Metabolism, and Elimination genes have been described in Mexican mestizo using pharmacogenomic microarray technology (Affymetix DMET array). In this work, we explore allele frequencies and genetic differentiation (*FST* statistic) of 25 polymorphisms (SNPs) in 8 genes: *CYP2C19, CYP2C9, CYP2D6, CYP3A5, DPYD, SLCO1B1, TPMT* and *VKORC 1.* According to the pharmacogenomics knowledge resource (PharmGKB) these genes have been reported to have the highest clinical level of evidence for pharmacogenetics implementation (Clinical Annotation Level 1A). We analyzed these 25 SNPs on three data sets: 320 unrelated Mexican mestizo with mainly Native American and European ancestry (57.4% and 39.1%, respectively); HapMap samples of 60 Europeans, 208 Africans individuals; and 45 Native Americans Zapotecas from Oaxaca, Mexico. Our results showed that allele frequencies and *FST* values for most of SNPs analyzed were similar for Europeans and Mexican mestizo. However, some SNPs showed high allele frequency differences, *CYP3A5* rs776746 with 6% in Europeans and 21% in Mexican mestizos, and *FST* value of 0.05 (p=1.99x10-05). As a matter of fact, this farmacogenomic marker has been reported to have the highest *FST* value between European and African populations (*FST*=0.64). Other genetic variants with important allele frequency differences were *CYP2C19* rs12248560, 22% in Europeans and 9% in Mexican mestizos, and *FST*=0.03 (1.94x10-4); the *CYP2C9* rs1799853, 5% in Mexican mestizos and 15% in Europeans, and *FST*=0.026 (p=4.4x10-4); *CYP2D6* rs3892097 24% in Europeans and 11% in Mexican mestizos, and *FST*=0.026 (p=3.3x10-3). In summary, the allele frequencies for drug related pharmacogenes studied here are in many cases two or four-fold higher between Europeans and Mexican mestizos. This information may help clinicians decide on dosages for populations with large Native American ancestry, as Mexico and other Latin-American countries.

658W

Influence of genetic variability on platelet aggregation in clopidogrel-treated patients of the International Clopidogrel Pharmacogenomics Consortium (ICPC). *J. Lewis[1,2] on behalf on the International Clopidogrel Pharmacogenomics Consortium.* 1) Endocrinology, Diabetes & Nutrition, University of Maryland, Baltimore, Baltimore, MD; 2) Program for Personalized and Genomic Medicine, University of Maryland, Baltimore, Baltimore, MD.

Dual antiplatelet therapy (DAPT) is critical in preventing recurrent cardiovascular events in patients with coronary artery disease (CAD) and other related disorders. While inhibition of platelet aggregation is most commonly achieved with aspirin and clopidogrel, other agents are available (i.e. ticagrelor and prasugrel) and allow for the possibility of more effective and personalized patient care. To better understand the genetic determinants of clopidogrel efficacy, we evaluated the impact of 32 previously reported single nucleotide polymorphisms (SNPs) on ADP-stimulated platelet aggregation in 4175 CAD patients of the International Clopidogrel Pharmacogenomics Consortium (ICPC). Consistent with prior investigations, we observed strong association between the loss-of-function *CYP2C19**2 [rs4244285] variant and increased on-clopidogrel ADP-induced platelet aggregation (P = 1.97 x 10-52, b = 0.46). In contrast, on-clopidogrel ADP-stimulated platelet aggregation was substantially lower in carriers of the loss-of-function *CES1* G143E variant (rs71647871; P = 4.27 x 10-17, b = -0.72) as well as the *CES1* intronic variant rs2244613 (P = 7.84 x 10-4, b = -0.09). While CAD patients who carried the well-described gain-of-function *CYP2C19**17 variant (rs12248560) initially appeared to have better clopidogrel response (P = 7.04 x 10-9, b = -0.17), this association did not remain statistically significant after adjustment for *CYP2C19**2 (P = 0.02, b = -0.07), suggesting that linkage disequilibrium between these SNPs was driving the association with *CYP2C19**17 (D' = 1.0, r2 = 0.05). Genetic variants in other candidate genes (e.g. *PON1* [Q192R (rs662) and L55M (rs854560)], *ABCB1* C3435T [rs1045642] and C1236T [rs1128503]), *ITGB3* [rs5918]), and other CYP enzymes including *CYP2B6*, *CYP2C9*, and *CYP1A2* were not associated with ADP-stimulated platelet aggregation in clopidogrel-treated patients after correction for multiple testing. Taken together, in addition to *CYP2C19**2, our data suggest that genetic variants in *CES1*, encoding the primary enzyme responsible for degrading clopidogrel into biologically inactive carboxylic acid metabolites, significantly influence platelet aggregation in patients on clopidogrel therapy.

659T

Investigating Genetic Variants of *IL-1β*, *IL-2*, *IL-6*, *TSPO*, and *BDNF* in Association with Treatment Response to Duloxetine and Placebo Treatment in Patients with Major Depression. *V.S. Marshe[1,2], M. Maciukiewicz[1], A.K. Tiwari[1], N. Freeman[1], S. Rotzinger[3], J.A. Foster[4], J.L. Kennedy[1,2,5], S.H. Kennedy[3,5], D.J. Müller[1,2,5].* 1) Pharmacogenetics Research Clinic, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Institute of Medical Science, Faculty of Medicine, University of Toronto, Ontario, Canada; 3) Department of Psychiatry, University Health Network, Toronto, Ontario, Canada; 4) Department of Psychiatry and Behavioral Neurosciences, McMaster University, Hamilton, Ontario, Canada; 5) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada.

Major depressive disorder (MDD) is a prevalent psychiatric disorder treated with antidepressant medication such as duloxetine. In addition, placebo treatments have been shown to improve depressive symptoms in a subgroup of patients. This study examined the role of genetic variation of inflammatory markers (*IL-1b*, *IL-2*, *IL-6*, and *TSPO*) including brain-derived-neurotrophic factor (*BDNF*) in response to duloxetine and placebo. Twenty single nucleotide polymorphisms (SNPs) across *IL-1b*, *IL-2*, *IL-6*, *TSPO* and *BDNF* were genotyped in 215 patients receiving duloxetine and 235 patients receiving placebo for 8 weeks. Samples were obtained through a partnership between the Canadian Biomarker Integration Network for Depression (CAN-BIND) and Lundbeck. Interleukin SNPs (r2 = 0.8, MAF > 0.05) covered ~100% of the common genetic variation. For ANCOVAs, we used quantitative and binary response variables. Quantitative response was defined as percentage change in MADRS score from baseline to endpoint. Binary response versus non-response was defined by at least 50% of reduction of MADRS scores from baseline. Two SNPs, rs2066992 (p=0.047) and rs10242595 (p=0.028), in the *IL-6* gene were associated with response to duloxetine after 6 weeks of treatment. *IL-6* variant rs2066992 was also significantly associated with response to placebo after 6 weeks (p=0.026). When dichotomizing response into response vs. non-response, *IL-6* variant rs10242595 was also found to be associated with response to duloxetine (p=0.003), but not placebo. Therefore, SNPs across *IL-6* may play a role in response to duloxetine and placebo.

660W

Spindle and Kinetochore Associated Complex Subunit 2 (SKA2) May Play a Role in Response to Citalopram. *A.J. Lisoway[1], C.C. Zai[1,3], A.K. Tiwari[1], D.J. Müller[1,2,3], Z.A. Kaminsky[4], J.L. Kennedy[1,2,3].* 1) Neurogenetics Section, Campbell Family Research Institute, Centre for Addiction and Mental Health, Toronto, Canada; 2) Institute of Medical Science, University of Toronto, Toronto, Canada; 3) Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD; 4) Department of Psychiatry, University of Toronto, Toronto, ON.

Background: Major Depressive Disorder (MDD) has a strong genetic component and is characterized by a number of physiological impairments, including diminished ability of the HPA axis to mediate stress response. The current process used to determine pharmacological treatments is markedly inefficient, with more than 50% of antidepressant treated patients failing to reach remission. Genetic and epigenetic variation in SKA2 (Spindle And Kinetochore Associated Complex Subunit 2) is implicated in mediating HPA axis function, and has recently been associated with suicidal behaviour. We hypothesized that genetic variation in SKA2 may play a role in predicting response to antidepressant medication.**Methods:** 492 Caucasian MDD patients were selected from the STAR*D sample. Change in HAMD-17 score was used to measure response to citalopram. Linear regression analysis was used to model the relationship between seven single-nucleotide polymorphisms (SNPs) in SKA2 and antidepressant response.**Results:** Marker rs7208505 was not significantly associated with antidepressant response (p=0.573) or baseline score on the HAMD-17 suicidality item (p=0.409). rs9892425, located in the 5' region of SKA2, was nominally associated with antidepressant response (p=0.017, pcorr=0.051). This finding was significant for males (p=0.003, pcorr=0.009), but not for females (p=0.396, pcorr=1.000). A trend was observed using a three marker haplotype window encompassing rs9892425 (p=0.068).**Conclusions:** The results provide some evidence that SKA2 genetic variation may be a predictor of therapeutic response to antidepressant medication, particularly in male patients with MDD. Further work incorporating epigenetic information is warranted in larger samples.

661T

Genetic Risk, Smoking Cessation, and the Clinical Benefits of Pharmacotherapy. *L. Chen[1], T. Baker[2], L. Bierut[1].* 1) Psychiatry, Washington University School of Medicine, St. Louis, MO; 2) Tobacco Research and Intervention, University of Wisconsin, School of Medicine, Madison, WI.

Background: Multiple genetic variants have been associated with the risk of nicotine dependence and heavy smoking with genome wide significance. In this study, we tested whether or not these variants could predict the probability of smoking abstinence in a treatment trial and identify those individuals who derive greater clinical benefit from cessation pharmacotherapy.Methods: Using a randomized controlled trial of smoking cessation (N = 1282) we studied the association of a genetic risk score based on 7 genetic variants with smoking abstinence at end of treatment, adjusting for traditional clinical risk factors. We then investigated the genetic prediction of smoking abstinence stratified by treatment condition.Findings: In smokers receiving placebo, the genetic variants significantly improved the prediction of the smoking abstinence at end of treatment. The area under curve (AUC) was 0.56 with clinical predictors (age, sex, cigarettes smoked per day), and increased significantly to 0.71 (p=0.018) when genetic variants were added to the clinical predictors. In contrast, these genetic variants do not improve the prediction of cessation. Smokers with the highest genetic risk derive the greatest benefit from the cessation pharmacotherapies tested.Interpretation: A genetic risk score identified individuals at increased risk for failed smoking cessation amongst those given placebo. People with the highest burden of genetic risk derived the largest relative and absolute clinical benefit from cessation pharmacotherapy.

662W

A common missense variant of *LILRB5* is associated with statin intolerance: A GoDARTS study. *M.K. Siddiqui, A. Veluchamy, R. Tavendale, F. Carr, C. Maroteau, E.R. Pearson, H.M. Colhoun, A.D. Morris, E. Dow, J. George, A. Doney, C.N.A. Palmer.* Division of Cardiovascular & Diabetes Medicine, University of Dundee, Dundee, United Kingdom.

There are approximately 12 million statin users in the United Kingdom. Approximately 9% of users present with intolerance to statins, manifesting as muscle ache, fatigue or more seriously, muscle breakdown leading to myopathy. Creatine phosphokinase (CK) levels are used as a biomarker of statin-induced muscle damage. Variants in *LILRB5* and *CKM* were shown to be associated with CK levels irrespective of statin usage. This study aims to analyse the association of these variants with statin intolerance.Genotype information was gathered for two missense variants, rs12975366 (*LILRB5:* Asp247Gly) and rs11559024 (*CKM*: Glu83Gly) in the GoDARTS study cohort of Scottish Caucasian individuals in the Tayside and Fife areas. We found both variants were associated with CK levels in statin users and non-users, but that only Asp247Gly was associated with our definition of statin intolerance (raised CK and accompanying prescribing changes) : OR 0.48, p value 7.7 x 10 -5 and 95% CI (0.28, 0.65).This demonstrates that the *LILRB5* locus is associated with statin intolerance-related raised CK levels, whereas *CKM* is purely a marker of constitutional serum CK levels. This study also highlights the probability that statin intolerant individuals may not have raised CK levels for genetic reasons.This study presents a novel genetic factor associated with statin intolerance and raises considerations for the usage of CK as a biomarker. It encourages further investigation into the physiology of statin-induced muscle damage, inter-individual variability in CK levels and response to muscle damage.

663T

Fucosyltransferase 2 polymorphism (FUT2, rs492602) associates with vitamin B12 deficiency, but may be independent with proton pump inhibitor or metformin usages, diabetes mellitus, and thyroid diseases. *H. Mo[1], L. Bastarache[1], J.A. Pacheco[2], Y. Xu[1,3], J. Pathak[4], W.K. Thompson[5], J.C. Denny[1,6].* 1) Biomedical Informatics, Vanderbilt University, Nashville, TN; 2) Center for Genetic Medicine, Feinberg School of Medicine, Northwestern University, Chicago, IL; 3) Department of Biostatistics, Center for Quantitative Sciences, Vanderbilt University School of Medicine, Nashville, TN; 4) Department of Health Sciences Research and Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN; 5) Center for Biomedical Research Informatics, NorthShore University HealthSystem, Chicago, IL; 6) Department of Medicine, Vanderbilt University, Nashville, TN.

GWAS have suggested plasma vitamin B12 concentration is associated with SNP rs492602-C in fucosyltransferase 2 (*FUT2*)(Nat Genet 2008;40:1160). Medical conditions (e.g., diabetes mellitus [DM], thyroid diseases, and use of proton pump inhibitors [PPI], and metformin) are also found to associate with B12 deficiency. However, it has not been established if this SNP associates with B12 deficiency as a disease state, nor is it known if the effect on B12 is a result of interactions between medications or other diseases, rather than representing an independent effect. Thus, we performed an association analysis between rs492602 and B12 deficiency, and also did an interaction analyses between the SNP and related medical conditions. **Methods:** We used rs490602 data from 29,929 European-ancestry individuals genotyped on the Illumina Exome array with traits defined by VUMC electronic health records. We ascertained B12 deficiency cases and matched them with controls (as described in Lam 2013,JAMA,310:2435), and their related comorbid status (DM, thyroid diseases) and medication usages (PPIs, metformin). Associations between B12 deficiency and the SNP or medical conditions were computed using logistic regression, and each interaction was tested as a gene-environment interaction term (GxE). **Results:** In our genotyped population, we identified 524 B12 deficiency cases, and 3509 controls. SNP rs492602-T associates with higher risk of B12 deficiency (OR=1.40, p=5.1e-7), and with abnormally low plasma B12 levels (OR=1.35, p=3.1e-7). We found associations between 3 of the 4 medical conditions and B12 deficiency, and we did not find an interaction with rs492602-C and any of the medical conditions.

Associations between B12 deficiency and medical conditions

	OR	p	GxE OR	GxE p
DM	1.64	1.4e-6	1.05	0.73
Thyroid diseases	1.29	0.012	0.98	0.86
PPI use	1.32	0.07	1.07	0.76
Metformin use(in DM patients)	1.83	0.0028	0.86	0.60

We also tested B12 deficiency and these medical conditions adjusted with rs492602, but none of the new ORs or p-values dramatically changed. **Conclusion:** We found rs492602-T associates with clinically significant B12 deficiency, and replicated known associations between B12 and 2 diseases and 1 medication. We did not find any evidence of rs492602 either confounding or potentiating the association between B12 deficiency and DM, thyroid diseases, PPI use, or metformin use. Thus, we suggest *FUT2* polymorphism is an independent factor of vitamin B12 deficiency.

664W

Applying genetics in inflammatory disease drug discovery. *L. Folkersen[1], S. Biswas[2], K.S. Frederiksen[3], P. Keller[3], B. Fox[2], J. Fleckner[3].* 1) Department of Systems Biology, Copenhagen, Denmark; 2) Department of Molecular Immunology, Novo Nordisk, Seattle, WA, USA; 3) Department of PharmacoGenetics, Novo Nordisk, Novo Nordisk Park, Måløv, Denmark.

Recent groundbreaking work in genetics has identified thousands of small-effect genetic variants throughout the genome that are associated with almost all major diseases. These genome-wide association studies (GWAS) are often proposed as a source of future medical breakthroughs. However, with several notable exceptions, the journey from a small-effect genetic variant to a functional drug has proven arduous, and few examples of actual contributions to drug discovery exist. Here, we discuss novel approaches of overcoming this hurdle by using instead public genetics resources as a pragmatic guide alongside existing drug discovery methods. Our aim is to evaluate human genetic confidence as a rationale for drug target selection.

665T

Novel genetic loci for Resistant Hypertension discovered through a Genome-wide Association approach (GWAS) in the INternational VErapamil SR-Trandolapril STudy (INVEST) and the Secondary Prevention of Subcortical Strokes (SPS3) Study. *N. El Rouby[1], CW. McDonough[1], Y. Gong[1], LA. McClure[2], BD. Mitchell[3,4], RB. Horenstein[3,5], RL. Talbert[7], A. Takahashi[8], T. Tanaka[8], M. Kubo[8], CJ. Pepine[9], RM. Cooper-DeHoff[1,9], OR. Benavente[6], AR. Shuldiner[3,4,5], JA. Johnson[1,9].* 1) Department of Pharmacotherapy and Translational Research and Center for Pharmacogenomics, University of Florida, Gainesville, FL; 2) Department of Biostatistics, School of Public Health, University of Alabama, Birmingham, AL; 3) Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD; 4) Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, MD; 5) Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD; 6) Department of Neurology, University of British Columbia, Vancouver, British Columbia, Canada; 7) College of Pharmacy, University of Texas at Austin, Austin, Texas; 8) RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 9) Division of Cardiovascular Medicine, Department of Medicine, University of Florida, Gainesville, FL, USA.

Resistant hypertension (RHTN) is a clinically important pharmacogenomic phenotype defined as a blood pressure (BP) ≥140/90 mm Hg despite use of ≥ 3 antihypertensive medications or BP <140/90 mm Hg using ≥ 4 drugs. RHTN is associated with increased risk for adverse cardiovascular outcomes. We sought to identify novel genetic variants of RHTN through a genome wide association study (GWAS). A discovery cohort of hypertensive participants (N=657 whites) from INVEST-GENES were included as RHTN cases (defined above) or controls defined as BP<140/90 using ≤3 drugs. Participants were randomized to a β-blocker strategy (BB) or a calcium channel blocker strategy (CCB) and had genotype data from the Illumina OmniExpressExome chip. Multiple logistic regression was conducted using an additive genetic model, adjusting for clinical predictors of RHTN, principle components for ancestry and BP treatment arms. Suggestive signals ($p < 1x10-5$) were tested for replication in 263 whites from SPS3-GENES. Three SNPs in the discovery phase met the suggestive p, and two were tested for replication (third SNP absent in SPS3 dataset). The Bonferroni-adjusted p was set at 0.025 to account for multiple comparisons. A combined meta-analysis was also performed between INVEST-GENES and SPS3-GENES for replicated SNPs or SNPs that had consistent association in the two datasets with nominal significance. An intronic SNP in the AQP4-antisense RNA 1 *(AQP4-AS1)* was identified in INVEST-GENES and met the criteria for replication. The A allele of rs630495 was associated with reduced risk for RHTN (OR=0.55, p=1.9 x10-6 in INVEST-GENES; OR=0.6, p=0.023 in SPS3-GENES; meta-p= 3.3 x10-7). rs630495 is an eQTL for the aquaporin-4 (*AQP4)* gene in brain tissue according to the GTex Portal database. *AQP4* is a member of the aquaporin family and functions at the basolateral side of the renal collecting tubule. *AQP4* is highly expressed in the brain and is thought to play an important role in edema production in cerebrovascular diseases. The other SNP near *UPK2* gene (rs647769) was associated with increased risk of RHTN in INVEST-GENES (OR= 2.0, p= 1.6 x 10-7) and was directionally consistent though not significant in SPS3-GENES (OR=1.4, p=0.069). The meta-analysis OR was 1.8 and p= 1.5 x 10-7. In conclusion, we identified *AQP4-AS1* as a novel gene region associated with RHTN, and *UPK2* as a suggestive gene region. These associations, if further validated, may help identify those patients at risk for RHTN.

666W

A Genome-wide Association and Admixture Mapping Study of Bronchodilator Response in African Americans with Asthma. *M.L. Spear[1, 4, 5], M. Pino-Yanes[2, 3], C. Eng[4], S. Huntsman[4], D.G. Torgerson[4], E.G. Burchard[4, 5] on behalf of the SAGE and GALA II Investigators.* 1) Biomedical Sciences Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA; 2) Research Unit, Hospital Universitario N.S. de Candelaria, Tenerife, Spain; 3) CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain; 4) Department of Medicine, University of California, San Francisco, San Francisco, CA; 5) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, San Francisco, CA.

Short-acting b2-adrenergic receptor agonists (SABAs) are the most commonly prescribed asthma medications. Response to SABAs is measured as bronchodilator response (BDR), which varies among racial/ethnic groups in the US. Common variants have been associated with BDR through genome-wide association studies (GWAS), with studies performed in European Americans and Latinos. However, knowledge of genetic variation that contributes to BDR has never been studied in African Americans with asthma. We performed a GWAS in African American children from the Study of African Americans, Asthma, Genes & Environments (SAGE). We included 757 individuals with asthma, genotyped with the Axiom LAT1 array (World Array 4, Affymetrix, Santa Clara, CA) followed by imputation using 1000 Genomes phase 3 data. We performed linear regression adjusting by age, sex, BMI and genetic ancestry to test for an association between BDR and genotype at single nucleotide polymorphisms (SNPs). To complement this analysis, we performed genome-wide admixture mapping to identify regions whereby local African or European genetic ancestry is associated with BDR. We identified a region of candidate SNPs on chromosome 3 as being associated with BDR. For variants with minor allele frequency (MAF)≥5%, the minimum p-value was 10-6; including low frequency variants (1%≤MAF≤5%) we identified SNPs with $p < 5x10-8$ as being associated with differences in BDR. This region is located proximal to *KCNH8*, a member of a family of genes whose functions include regulating smooth muscle cell contraction. Candidate SNPs in *SPATS2L* that previously associated with BDR in Caucasians showed a similar trend of association in African Americans. From our admixture mapping analysis, we identified a peak on chromosome 2 (p=0.0003) where African ancestry was associated with lower BDR. We replicated this effect in Latinos from the Genes-environments and Admixture in Latino Americans (GALA II)study (p=0.005). Overall, our GWAS for BDR in African Americans with asthma identified regions containing common and low frequency variants as being associated with differences in BDR for follow-up study.

667T

Pharmacogenetics of Acute Coronary Syndrome. *P. Yin[1], A. Jorgensen[1], A. Morris[1], R. Turner[2], R. Fitzgerald[2], R. Stables[3], A. Hanson[2], M. Pirmohamed[2].* 1) Department of Biostatistics, INSTITUTE OF TRANSLATIONAL MEDICINE, Liverpool, United Kingdom; 2) Department of Molecular and Clinical Pharmacology, INSTITUTE OF TRANSLATIONAL MEDICINE, Liverpool, United Kingdom; 3) Liverpool Heart and Chest Hospital, Liverpool, United Kingdom.

Coronary heart disease (CHD) is a major global public health burden. It is a recurrent disease that is treated by multiple drugs, including aspirin, clopidogrel, statins, beta-blockers, and ACE inhibitors. There is significant inter-patient variability in the response to cardiovascular drugs, which may have a heritable basis. We have therefore undertaken a genome-wide association study (GWAS) to identify loci associated with response to cardiovascular drugs in 1470 patients recruited to a UK prospective pharmacogenetic study of acute coronary syndrome (PHACS).Index hospital admission was defined non-ST elevation acute coronary syndrome. Patients were treated with a range of cardiovascular drugs, including statins (93%) and clopidogrel (83%), and followed up prospectively for up to 48 months. Approximately 8% of the patients had another cardiovascular event including myocardial infarction (MI) or stroke, in some cases resulting in death. We defined a binary outcome corresponding to the occurrence of any of these events in one year after hospital discharge. We began by considering clinical risk factors for drug response. Several risk factors were related: age (p=3.2x10-7), prior MI (p=0.0048), diabetes (p=0.032), ACE inhibitor use pre-admission (p=0.0072), aspirin use at discharge (p=0.0066) and gender (p=0.12).Patients were genotyped using the Illumina OmniExpress array. After quality control, the genotype scaffold was imputed up to the 1000 Genomes Phase I reference panel (all ancestries, March 2012 release). We tested for association of SNPs with outcome under an additive dosage model in a logistic regression framework after adjusting for the clinical factors identified above and principal components to account for population structure. Variants mapping to the *LRPPRC* gene demonstrated strong evidence of association: lead SNP rs65544733, minor allele frequency 0.37, odds ratio (95% CI) 1.96 (1.49-2.63), p=2.9x10-7. When stratifying the analysis by drug, the association with this variant was strongest in patients treated with statins (p=6.5 x10-8). *LRPPRC* encodes a leucine-rich protein with, as yet, unknown role. Mutations in the gene have previously been associated with Leigh syndrome, which causes lowered levels of cytochrome C oxidase, a key enzyme in aerobic metabolism. Our study highlights that variants mapping to *LRPPRC* are associated with response to cardiovascular drugs in CHD patients, in particular those treated with statins.

668W

Using polygenic risk scores to guide antipsychotic dosage in a geriatric schizophrenia population. *N. Hettige[1,2], C. Cole[1], V. De Luca[1,2].* 1) Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada.

Antipsychotic medications are commonly used to treat elderly patients with delirium, agitation and psychosis due to Alzheimer's disease, and schizophrenia. Schizophrenia is a debilitating mental healh disorder that once diagnosed, remains throughout the lifetime. The first line of treatment for individuals with schizophrenia is to administer antipsychotic medication which aims to reduce the severity of psychotic symptoms, such as hallucinations and paranoia. As individuals with schizophrenia may be required to take antipsychotic medication for the duration of their life, antipsychotic dosage must be adjusted accordingly due to age-related neurobiological changes or symptom severity. The purpose of our study was to calculate the polygenic risk scores for the risk alleles that reached genome-wide significance from the recent genome-wide association study by the Psychiatric GWAS Consortium for Schizophrenia. We hypothesized that individuals with a high risk score would potentially represent increased symptom severity and therefore require higher antipsychotic dosage. Polygenic risk scores were used to predict whether individuals with a higher score also require higher antipsychotic dosage. In our preliminary sample of 83 European Caucasian individuals, we found that the risk score was not significantly predictive of antipsychotic dosage. Incorporating the polygenic score in our model, however, better explained the variance in dosage compared to age and sex alone. The polygenic risk score may be a useful way of translating the knowledge aquired from GWA studies to predict clinical outcomes and related endophenotypes.

669T

Partitioning the polygenic inheritance of paclitaxel-induced peripheral neuropathy. *E.A. Khramtsova[1], H.E. Wheeler[1], D.L. Kroetz[2], K. Owzar[3], H.L. McLeod[4], N.J. Cox[5], M.E. Dolan[1], B.E. Stranger[1], L.K. Davis[1].* 1) Department of Medicine, The University of Chicago, Chicago, IL; 2) Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA; 3) Department of Biostatistics and Bioinformatics, Duke University, Durham, NC; 4) Personalized Medicine Institute, Moffitt Cancer Center, Tampa, FL; 5) Division of Genetic Medicine, Vanderbilt University School of Medicine, Nashville, TN.

Peripheral neuropathy is one of the most common side effects of paclitaxel treatment affecting 20-40% patients receiving the chemotherapy. Previous studies have identified several single nucleotide polymorphisms (SNPs) and genes associated with increased susceptibility to peripheral neuropathy development, but individual SNP replication is lacking, so we sought to look for polygenic traits. Whole genome estimates of heritability have been challenging due to small sample size, although a previous study has shown that if gene sets are analyzed separately, the axonogenesis GO Term set had significant estimates of heritability close to 20%. We performed an analysis of paclitaxel-induced neuropathy heritability in the CALGB40101 (Alliance) trial, but in contrast to previous analysis of these datasets, we performed a more stringent population stratification analysis and included the X chromosome in the analysis. Principal component analysis of genetic data revealed a small group of closely-related individuals which formed a separate cluster from all of the individuals of European ancestry. These 10 individuals in the CALGB40101 cohort were excluded from further analysis. Thus, our study sample consisted of 204 cases with neuropathy grade 2 and 3, and 640 controls. Using GCTA, we estimated the whole genome heritability of paclitaxel-induced peripheral neuropathy in this cohort to be 0.49 (se=0.42, P=0.114). Partitioning heritability by allele frequency revealed that the minor allele frequency group of >0.4-0.5, which comprises 18.7% of total SNPs, contributes the most to heritability (h2=0.45, se=0.25, P=0.038). Partitioning heritability by chromosome showed that chromosomes 2 (h2=0.22, se=0.12, P=0.037) and 10 (h2=0.19, se=0.10, P=0.032) contribute the most to heritability. We did not find any contribution to heritability from the X chromosome. Genome-wide association analysis of neuropathy cases versus controls did not reveal any genome-wide significant SNP-trait associations. However, the strongest signals (P<10-6.5) were for SNPs in the *ESPNL* and *UBE2F-SCLY* genes. While chromosomes 2 and 10 explained the greatest proportion of phenotypic variance, there were no SNPs with P<10-8 on those chromosomes, suggesting that individual effect sizes are small but cumulatively important. Taken together, our findings are consistent with a polygenic model for chemotherapy-induced neuropathy.

670W

A Variant in *UGT2A1/2* is Associated with Sex-Specific Clopidogrel Response. *A.S. Fisch, J.P. Lewis, L.M. Yerges-Armstrong, A. Parihar, H. Xu, J.R. O'Connell, K.A. Ryan, R.B. Horenstein, B.D. Mitchell, A.R. Shuldiner.* Program for Personalized and Genomic Medicine, Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD.

Coronary artery disease (CAD) is the leading cause of death in the United States, with reported event rate disparities across sexes. Clopidogrel is an antiplatelet agent for high-risk CAD treatment. Clopidogrel response variability is established, and previous investigations show a genetic component to the variability, including the well-described *CYP2C19*2* variant. However, *CYP2C19*2* accounts for 12% of the ~70% heritability of clopidogrel response, suggesting undiscovered genetic determinants of the trait. To identify these variants, we genotyped 647 individuals from the Pharmacogenomics of Anti-Platelet Intervention (PAPI) Study using the Affymetrix Drug Metabolizing Enzymes and Transporter (DMET) chip. We then examined the effect of 643 polymorphic SNPs with MAF ≥ 0.01 (multiple testing-corrected significance threshold of P=7.43x10-5) on clopidogrel response, defined as maximal platelet aggregation change between pre- and post-clopidogrel treatment for one week (300 mg loading dose, 75 mg maintenance dose). Associations were analyzed with a multivariable linear regression model adjusting for age, sex, and relatedness. We observed significant associations between clopidogrel response and *CYP2C19*2* (rs4244285; P=5.34x10-13) as well as the novel intronic variant rs11249454 in UDP glucuronosyltransferase 2A1 and 2A2 (*UGT2A1/2*), with the minor allele associated with a 24% decrease in clopidogrel response (b=-10.3±2.4; P=2.92x10-5). Based on the roles of UGT2A1/2 in androgen and estrogen metabolism, a sex stratified association analysis was performed, showing significance in 325 PAPI women (P=8.40x10-6) and no association in 322 men (P=0.21). The results of a SNP x sex interaction analysis were significant (P = 0.049). Sex-specific differences remained after *CYP2C19*2* adjustment (Pwomen=8.12x10-5; Pmen=0.14). This study suggests that genetic variation in *UGT2A1/2* affects clopidogrel response in a sex-dependent manner. These findings are consistent with other groups' data regarding sex-specific effects of sex hormone metabolizing UGTs, and may connect to prior reports of altered *UGT2A1/2* expression by substrate level and biological sex. In light of previous findings regarding sex hormone effects on platelet aggregation, UGT2A1/2 may be a key regulator of sex-specific clopidogrel response. Continued studies are needed to replicate our findings and understand mechanisms by which rs11249454 affects UGT2A1/2 function and, ultimately, clopidogrel response.

671T

Prevalence of CYP2D6*4 allele in a Northern Mexican Mestizo population for the assessment of tamoxifen metabolism. *J.E. Gaytán-Arocha[1], K.L. Valdés-Morales[1], R.D. Arellano-PérezVertti[1], D. Delgadillo-Guzmán[1], R.I. de la Cruz-Granados[1], J.R. Argüello[1,2], F.F. González-Galarza[1].* 1) Faculty of Medicine, Autonomous University of Coahuila, Torreon, Coahuila, Mexico; 2) Institute of Science for Genomics Medicine, Torreon, Coahuila, Mexico.

In recent years, the use of different technologies such as microarrays has allowed scientists to perform cost-effective association studies in different fields, including pharmacogenomics. Tamoxifen has been a widely studied drug that is used in breast cancer therapy which needs the CYP2D6 enzyme to be metabolized into its active form. From the 155 alleles that have been identified in the CYP2D6 gene at present, CYP2D6*4 is the most common allele with inactive enzyme activity. Thus, in this study, we analyzed one-hundred randomly-selected individuals from the North Area of Mexico that were SNP genotyped using the Illumina array technology with more than 600,000 SNPs. To evaluate the allele coverage of the DNA chip, we used an in-house software tool to perform the sequence alignment. We also calculated genotype and allele frequencies and evaluate the CYP2D6 gene diversity in Mexicans compared to other populations. Interestingly, the allele CYP2D6*4 was present in 12.9% of the individuals. We believe that the use of these findings may provide a useful resource for scientists and assist physicians when deciding to prescribe tamoxifen.

672W

A longitudinal genome-wide association study of anti-tumor necrosis factor response among Japanese patients with rheumatoid arthritis. *D. Jawaheer[1], I. Hallgrimsdottir[2], K. Honne[3], C. Wu[4,5], R. Sebro[6], N. Jewell[7], T. Sakurai[8], M. Iwamoto[3], S. Minota[3].* 1) Children's Hospital Oakland Research Institute, Oakland, CA, USA; 2) Integrative Biology, UC Berkeley, CA 94720, USA; 3) Rheumatology and Clinical Immunology, Jichi Medical University, Shimotsuke, Tochigi, Japan; 4) Institute of Clinical Research, University of Southern Denmark, Odense, Denmark; 5) Dept of Obstertics & Gynecology, Odense University Hospital, Odense, Denmark; 6) Radiology Dept, University of Pennsylvania, Philadelphia, PA 19104, USA; 7) School of Public Health, UC Berkeley, CA 94720, USA; 8) Inoue Hospital, Takasaki, Gunma, Japan.

Introduction: Studies of patients with rheumatoid arthritis (RA) to identify genetic biomarkers of anti-tumor necrosis factor (TNF) response have used response at a single follow up time point as the phenotype with which single nucleotide polymorphisms (SNP) associations have been tested. There has been little overlap in findings across studies, most of which have been conducted in Caucasian populations. We report here the first genome-wide association study (GWAS) to identify genetic biomarkers of anti-TNF response among Japanese RA patients, using anti-TNF response at 2 time-points for a more reliable clinical outcome over time.Methods: Disease Activity Scores based on 28 joint counts and C-reactive protein (DAS28CRP3) were assessed at baseline (before initial therapy), and after 3 and 6 months in 487 Japanese RA patients starting anti-TNF therapy for the first time or switching to a new anti-TNF agent. A genome-wide panel of single nucleotide polymorphisms (SNPs) was genotyped and additional SNPs were imputed. Using change in DAS28 scores from baseline (ΔDAS28) at both 3 and 6 months as the response phenotype, a longitudinal genome-wide association (GWA) analysis was conducted using Generalized Estimating Equations (GEE) models to accommodate the repeated measures of the outcome, adjusting for baseline DAS28, time since initiation of therapy, type of anti-TNF agent and concomitant methotrexate.Results: A total of 4,253,138 autosomal SNPs passed quality thresholds for association analysis. Suggestive evidence of association ($p<1x10-6$) with ΔDAS28 was observed at 3 chromosomal regions (6q15: rs284515, p=6.6x10-7; 6q27: rs75908454, p=6.3x10-7 and 10q25.3: rs1679568, p=8.1x10-7), extending to numerous SNPs in linkage disequilibrium (LD) across each region. Potential candidate genes in these regions include *MAP3K7* (6q15) a key player in TNFα-mediated inflammatory pathway signaling, *GFRA1* (10q25.3) which was associated with anti-TNF response in an independent study, and *WDR27* (6q27).Conclusion: In this first GWAS of anti-TNF response among Japanese RA patients using a longitudinal analysis approach, three genomic regions demonstrated suggestive association with response to anti-TNF therapy. Using more than one assessment of response enhanced the power to detect these associations.

673T

Meta-analysis of the genome wide association studies (GWAS) on the intolerance of angiotensin converting enzyme inhibitors (ACEIs). *C.N.A Palmer[1], S.H Mahmoudpour[2], M.K. Siddiqui[1], A. Veluchamy[1], F.W. Asselbergs[3], P.C. Souverein[2], C.E. de Keyser[4], A. Hofman[4], B.H. Stricker[4], A. de Boer[2], A.H. Maitland-van der Zee[2], PREDICTION-ADR.* 1) Medical Research Institute, University of Dundee, Dundee, United Kingdom; 2) 1Div. of Pharmacoepidemiology & Clinical Pharmacology, UIPS, Utrecht University, Utrecht, the Netherlands; 3) Div. of Heart & Lungs, Dept. of Cardiology, UMC Utrecht, Utrecht, The Netherlands; 4) Dept. of Epidemiology, Erasmus MC, Rotterdam, The Netherlands;.

BackgroundACEIs are frequently used to treat hypertension and heart failure. Cough and angioedema are the two main adverse drug reactions (ADRs) associated with ACEI use that occur in up to 20% of the patients and are the main reason of therapy discontinuation1.**Aims**To identify single nucleotide polymorphisms (SNPs) associated with switching of an ACEI to an angiotensin receptor blocker (ARB) as a marker for ADRs.**Methods**A cohort of patients starting ACEIs was identified within the Rotterdam Study in the Netherlands and the GoDARTS study in Scotland. Cases were subjects that switched from an ACEI to an ARB while controls were subjects who used ACEIs for at least 2 years and did not switch. The validity of using switching as a marker for ACEI-induced adverse drug reaction (ADR) was investigated in a subset of users that had the primary care records available. A GWAS using an additive model was performed and results were meta-analyzed using METAL2. **Results/Conclusions** In total 5109 ACEI starters were included in the study of which 959 were cases. The validation of switch as marker for ACEI-induced ADRs showed the positive predictive value of 90.5% for at least possible ADRs within a subset of 1132 patients. Ten SNPs within four genes reached the GWAS significance level in the meta-analysis. The strongest associated SNP was located on chromosome 17q25 (MAF=0.16, OR=1.52 [95%CI: 1.32-1.76], p=6.2x10-9). These results indicate a substantial contribution of genetic variation in determining the risk of ACEI-induced ADRs, and warrant further studies in larger populations.**References:**Morimoto T, Gandhi TK, Fiskio JM, Seger AC, So JW, Cook EF, et al. An evaluation of risk factors for adverse drug events associated with angiotensin-converting enzyme inhibitors. J Eval Clin Pract 2004; 10: 499-509.Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190-2191.

674W

Identifying a Novel Hypertension Genetic Signature Influencing the Blood Pressure Response to Hydrochlorothiazide Treated Patients. *M.H. Shahin[1], A.C. Sa[1], A. Webb[2], Y. Gong[1], T. Langaee[1], C.W. McDonough[1], A.L. Beitelshees[3], A.B. Chapman[4], J.G. Gums[1], S.T. Turner[5], R.F. Frye[1], S.E. Scherer[6], W. Sadee[2], R.M. Cooper-DeHoff[1], J.A. Johnson[1].* 1) Pharmacotherapy and Translational Research, University of Florida, Gainesville, Florida; 2) Program in Pharmacogenomics, Department of Pharmacology, The Ohio State University, Columbus, Ohio; 3) Department of Medicine, University of Maryland, Baltimore, Maryland; 4) Department of Medicine, Emory University, Atlanta, Georgia; 5) College of Medicine, Mayo Clinic, Rochester, Minnesota; 6) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas.

Hypertension is a global public health burden and a significant contributor to heart attack, stroke, and kidney failure, making its control of critical importance. Hydrochlorothiazide (HCTZ) is among the most commonly prescribed first line antihypertensives in the US, yet, less than 50% of HCTZ treated patients achieve blood pressure (BP) control. A recent meta-analysis, including gene expression profiles of ~7000 individuals, revealed 34 genes explaining ~9% of the inter-individual variability in BP (PMID:25785607). We hypothesized that those genes might also influence the BP response to HCTZ and give us more insight in the BP lowering mechanism underlying this therapy. The primary analysis included clinical data and biological samples from 228 white hypertensives recruited as part of the Pharmacogenomic Evaluation of Antihypertensive Response (PEAR) trial, with BP determined at baseline and after 9 weeks of HCTZ treatment. A total of 216 Single nucleotide polymorphisms (SNPs), within the candidate regions (transcript ± 2kb) of the 34 BP susceptible genes, were extracted from the Illumina Omni 1M-Quad Chip. A Bonferroni threshold of 2.3x10-4 (0.05/216) was used to account for multiple comparisons of the SNP association analysis. Illumina© HiSeq 2000 was used to conduct RNA-Seq on PEAR white baseline blood samples of extreme HCTZ BP response (25 responders and 25 non-responders). RNA-sequencing reads were aligned to the reference genome (homo sapiens Hg19) with TopHat2, and gene expression levels were calculated using cufflinks/cuffdiff. We adjusted for age, gender, and baseline BP in all analyses. First, out of the 216 tested SNPs, we identified rs6750487 SNP within the *ARHGAP15* (Rho GTPase activating protein 15) gene as a significant predictor of HCTZ BP response where A allele carriers had a better BP response than non-carriers (DSBP/ΔDBP:-19.8/ -11.4 vs -7.5/-4.1 mmHg, respectively, DSBP p=8x10-6 and ΔDBP p=1x10-4). Additionally, *ARHGAP15* baseline expression levels were significantly correlated with HCTZ BP response (DSBP r=-0.37 p=9x10-3, ΔDBP r=-0.36 p=1x10-2). This study aligns with recent studies showing the importance of *ARHGAP15* in hypertension and BP regulation. Additionally, our results suggest that *ARHGAP15* might be an important determinant of HCTZ BP response.

675T

Robust Heritability Estimation of Anti-TNF Treatment Response Phenotypes in Rheumatoid Arthritis. *K.A. Standish[1,2], C.C. Huang[4], M. Curran[4], N.J. Schork[1,2,3].* 1) Human Biology, J. Craig Venter Institute, La Jolla, CA; 2) University of California, San Diego; 3) The Translational Genomics Institute, Phoenix, AZ; 4) Janssen R&D, LLC.

Several classes of drug have been approved to treat rheumatoid arthritis (RA); however, there is heterogeneity in patients' response to different treatments, including anti-TNF agents. Identifying an effective treatment early can slow disease progression and result in better overall outcomes for patients. Pharmacogenetics studies aim to identify genetic predictors of treatment response in order to better understand the underlying mechanisms of response variability and to inform more efficient treatment decisions. Previous studies attempting to estimate the heritability of anti-TNF treatment response have certain limitations. While patients can be stratified into "responder" and "non-responder" groups, continuous response phenotypes (e.g., disease activity state, swollen or tender joint counts) provide a more robust alternative; however, many genetic association methods make assumptions about the distributions of the phenotype variables. In particular, these parametric approaches assume that the data are normally distributed and heteroscedastic. In reality, the response phenotypes often do not conform to these assumptions. Furthermore, previous studies relied on a single pre-treatment measurement and a single follow-up measurement to determine the degree of response, failing to account for variability among repeated measurements. Here, we address these limitations as we make use of a longitudinal clinical data set and whole-genome sequencing in a cohort of rheumatoid arthritis patients treated with an anti-TNF drug. First, we find that transforming data to better conform to the basic assumptions of parametric association methods increased the heritability estimates for several metrics and resulted in greater levels of significance. Next, we show the variability between single measurements and confirm an increase in power when making use of repeated measurements before and/or after treatment. We find consistently higher heritability estimates when making use of averaged response metrics. Additionally, we demonstrate that the use of more subjective measurements, such as swollen and tender joint counts, can be confounded by placebo response. Finally, we validate our findings using permutation testing to calculate p-values relating to each heritability estimate. Overall, our work provides an example of the importance of robust estimation of phenotype heritability to pharmacogenetics studies.

676W

***CYP2D6* Gene Variants and Effectiveness of Adjuvant Tamoxifen in Breast Cancer: A Population-Based Case-Control Study.** *C.S. Richards[1], K.E. Richert-Boe[2], K.A.B. Goddard[2], C. Chen[2], S. Punj[1], D. Schwarzkopf[2], M. Kalter[2], S. Weinmann[2].* 1) Department of Molecular and Medical Genetics, Knight Diagnostic Laboratories, Oregon Health & Science University, Portland, OR; 2) Center for Health Research, Kaiser Permanente Northwest, Portland, OR.

Tamoxifen, a cornerstone of adjuvant therapy for hormone-receptor-positive breast cancer, is metabolized to the active metabolite endoxifen through enzymatic activity of cytochrome P450 2D6. *CYP2D6* has numerous alleles that affect metabolizing phenotype. Among women who take tamoxifen, those homozygous for inactive alleles (poor metabolizers) have lower levels of serum endoxifen than those with two functional alleles (extensive metabolizers). We conducted a population-based case-control study in the Kaiser Permanente Northwest integrated health plan to evaluate the hypothesis that, after adjuvant tamoxifen treatment for breast cancer, women with *CYP2D6* genotypes associated with poor metabolism of tamoxifen have an elevated risk of breast cancer recurrence compared to women with *CYP2D6* genotypes associated with extensive metabolism of tamoxifen. Study subjects were diagnosed between 1980 and 2011 with hormone-receptor positive breast cancer who received at least 180 days of adjuvant tamoxifen treatment and for whom stored formalin-fixed paraffin-embedded (FFPE) normal tissue was available for laboratory analysis. The Oregon Health & Science University Molecular Genetics Laboratory extracted genomic DNA from stored FFPE tissue blocks and performed allelic discrimination assays and pyrosequencing to accurately determine *CYP2D6* variant status for the alleles, *3, *4, *5, *10, *17, and *41. We also collected data from medical records and from pharmacy, laboratory, tumor registry, and membership health plan databases. Statistical analysis will include multivariable logistic regression analysis to assess CYP2D6 functional status and activity score in relation to breast cancer recurrence. We identified 358 cases with breast cancer recurrence. Randomly selected controls (833), without breast cancer recurrence, were matched to cases on tumor stage, diagnosis year, diagnosis age, race/ethnicity, and patterns of health plan membership. All assays have been completed and study subjects have been categorized according to CYP2D6 metabolizer phenotype (poor, intermediate, extensive) and activity score (0-2). Based on the ethnicities in our study population, the *CYP2D6* allele frequencies are in Hardy-Weinberg equilibrium, and the frequencies of the predicted metabolizer phenotypes also fall within the expected range. Preliminary logistic regression results will be presented.

677T

Assessment of Knowledge and Comfort of Healthcare Providers in the Development of a User-Friendly Pharmacogenomics Report. *C.A. Campbell[1], A.E. Kwitek[1], M. Kimble[1], D.L. Kolbe[1], C. Nishimura[1], M.A. Mansilla[1], S.O. Mason[1], T. Bair[1], K.L. Knudtson[1], O. Shchelochkov[1], D.J. Murry[2], M. Sorenson[3], M.J. Brownlee[3], R.J.H. Smith[1].* 1) Iowa Institute of Human Genetics, University of Iowa, Iowa City, IA; 2) University of Iowa College of Pharmacy, Iowa City, IA; 3) University of Iowa Hospitals and Clinics Pharmacy Department, Iowa City, IA.

As genetic tests are ordered more frequently by non-geneticists, it is imperative the results are understandable and useful. The Iowa Institute of Human Genetics (IIHG) offers a clinical pharmacogenetics (PGx) test, so our goal was to create a result report that is an informative educational tool to aid providers in the use of PGx in the management of their patients. The purpose of this study was to ask healthcare providers about PGx test reports to see whether the medical and genetic terminology used in the reports is understandable to the average healthcare provider. The original report contained a brief summary of results and recommendations. Clinical recommendations are based on the Clinical Pharmacogenetics Implementation Consortiumguidelines and written in collaboration with pharmacists at the University of Iowa. The report also contained appendices of result and methodology details, and educational materials. A genetic counselor conducted brief interviews with eight surgeons who had previously used the PGx test on 53 patients. Responses were transcribed and analyzed. Questions covered understanding of results, content preferences, points of confusion, PGx implementation challenges, and comfort explaining results. Providers were also shown results reports from other clinical testing labs for comparison. The report was modified, and additional interviews are being conducted with pharmacists, residents, and internists. Results of the interviews indicated all of the providers understood the report. In general, providers; 1) liked the report's brief summary and clinical recommendations; 2) appreciated education information in the appendices; and 3) were comfortable explaining results to patients. One interviewee stated, the report has "to speak to both the clinician and layperson. The clinician is going to read this like a layperson, but needs to understand what it means for the medicine so he can appropriately treat his patient." Providers expressed frustration with; 1) non-intuitive pharmacogenetics nomenclature; and 2) lack of insurance reimbursement. In summary, stakeholder interviews aided in the development a user-friendly PGx report that can serve as an educational resource for non-genetics providers. Feedback on return of genetic test results, and utilization of clinical pharmacogenetics testing from non-geneticists providers is critical if pharmacogenetics testing is to be implemented as part of a precision medicine program.

678W

Development and comparison of warfarin dosing algorithms in stroke patients. *S. Cho[1], K. Lee[2], J. Choi[3], K. Lee[1].* 1) Departments of Laboratory Medicine, Gangnam Severance hospital, Yonsei University College of Medicine, Seoul, Korea; 2) Departments of Neurology, Gangnam Severance hospital, Yonsei University College of Medicine, Seoul, Korea; 3) Departments of Laboratory Medicine, Severance hospital, Yonsei University College of Medicine, Seoul, Korea.

Purpose: The genes for cytochrome P450 2C9 (*CYP2C9*) and vitamin k epoxide reductase complex subunit 1 (*VKORC1*) have been identified as important genetic determinants of warfarin dosing and have been studied. We developed warfarin algorithm for Korean patient with stroke and compared the accuracy of warfarin dose prediction algorithms based on the pharmacogenetics.**Material and Methods**: A total of 101 patients on stable maintenance dose of warfarin were enrolled. Warfarin dosing algorithm was developed using multiple linear regression analysis. The performance of all the algorithms was characterized with coefficient of determination, determined by linear regression, and the mean of percent deviation predicted doses from the actual dose. In addition, we compared the performance of the algorithms using percentage of predicted dose falling within ±20% of clinically observed doses and dividing the patients into a low-dose group (≤3mg/day), an intermediate-dose group (3-7mg/day), and high-dose group (≥7mg/day).**Results**: A newly developed algorithms including the variables of age, body weight, and *CYP2C9* and *VKORC1* enotype. Our algorithm accounted for 51% of variation in the warfarin stable dose. The predicted doses using algorithms derived from Anderson and this study showed the best correlation with actual warfarin doses. Our algorithm performed best in predicting dose within 20% of actual dose and intermediate-dose.**Conclusion:** Our warfarin dosing algorithm may be useful for Korean patients with stroke. Further studies to elucidate clinical utility of genotype-guided dosing and find the additional genetic association is necessary.

679T

The identification of hematologic adverse drug events using the electronic health record in the eMERGE PGx cohort, in relation to pharmacogenetic variation. *D. Crosslin[1,2], A. Gordon[1], D. Carrell[3], P. Robertson[2], A. Scrol[3], J. Ralston[3], A. Hartzler[3], K. Leppig[3], M. de Andrade[4], J. Grafton[3], D. Kim[1,2], A. Burt[1], K. Doheny[5], P. Crane[6], S. Stallings[7], M. Dorschner[8], D. Nickerson[2], E. Larson[3], J. Denny[7], M. Ritchie[9], M. Williams[10], M. Palmer[11], I. Kullo[12], T. Manolio[13], G. Jarvik[1,2], The electronic Medical Records and Genomics (eMERGE) Network.* 1) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 2) Department of Genome Sciences, University of Washington, Seattle, WA; 3) Group Health Research Institute, Center for Health Studies, Seattle, WA; 4) Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN; 5) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 6) Division of General Internal Medicine, University of Washington, Seattle, WA; 7) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 8) Department of Pathology, University of Washington, Seattle, WA; 9) Center for Systems Genomics, Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA; 10) Genomic Medicine Institute, Geisinger Health System, Danville, PA; 11) Department of Urology, University of Washington, Seattle, WA; 12) Division of Cardiovascular Diseases, Mayo Clinic, Rochester, MN; 13) Division of Genomic Medicine, National Human Genome Research Institute, Bethesda, MD.

Using the electronic health record (EHR), we are attempting to algorithmically identify adverse drug events (ADEs) in ~9000 eMERGE participants that have been sequenced with the targeted Pharmacogenomics Research Network sequence platform (PGRNseq). This platform was designed with the coding regions, UTRs, 2kb upstream, and 1kb downstream for 82 pharmacogenes. The goals are to assess a trend in ADEs in this cohort overall and stratified by ancestry/sex, and to assess the possible association with pharmacogene(s). Network-wide recruitment is ongoing (~6000), and we are developing the algorithm on a pilot set of ~900 mixed-ancestry participants from the Group Health, Seattle. The algorithm harnesses information from demographic, laboratory, and prescription data derived from the outpatient EHR. Using laboratory data, we assess patterns of change in hematological counts that could suggest an ADE, and assess patterns of change in platelets to identify thrombocytopenia due to multiple mechanisms. These events, coupled with prescription data patterns, could further support the identification of ADEs. Once these patterns are identified, we will assess the association of these events with genetic variation identified by PGRNseq, overall and stratified by ancestry/sex. As proof of concept, we assessed 900 persons for platelet count < 2 standard deviations of the median for that participant. We evaluated all prescriptions from the prior 7-28 days. The most frequent drug identified was Acetaminophen/Hydrocodone with 11 events for 5 participants. Using guidelines from PharmGKB, we focused on variation in *ABCG2*, *CYP2D6*, *G6PD*, *HLA-B*, and *UGT1A1*. A total of 8 variants were identified (7 missense, 1 splice-acceptor) and found in *ABCG2* (n=2), *CYP2D6* (n=5), and *UGT1A1* (n=1). Notably, *ABCG2* is involved in metabolism and transport of acetaminophen in the liver. Next steps are to refine the EHR algorithm with a sliding window for patterns of hematologic changes, and a custom prior prescription window per blood cell index. We also will expand to the entire eMERGE PGx cohort as enrollment approaches ~9000. Identification of AEs using the EHR and the pattern of enrichment of these events with pharmacogenes in the eMERGE cohort could inform on drug selection and dosing guidance overall, and by ancestry.

680W

Automated *CYP2D6* diplotype assignment using whole genome sequences. *A. Gaedigk[1,3,5], G.P. Twist[2], N. Miller[2], E.G. Farrow[2], L.K. Willig[2,3], D.L. Dinwiddie[6,7], J.E. Petrikin[2,3], S. Herd[2], M. Gibson[2], J. Cakici[2], A.K. Riffel[1], J.S. Leeder[1,2,3,5], D. Dinakarpandian[8], S.F. Kingsmore[2,3,4].* 1) Clinical Pharmacology, Toxicology & Therapeutic Innovation, Children's Mercy Kansas City, Kansas City, MO, USA; 2) Center for Pediatric Genomic Medicine, Children's Mercy Kansas City, Kansas City, MO, USA; 3) Department of Pediatrics, Children's Mercy Kansas City, Kansas City, MO, USA; 4) Department of Pathology, Children's Mercy Kansas City, Kansas City, MO, USA; 5) School of Medicine, University of Missouri-Kansas City, Kansas City, MO, USA; 6) Department of Pediatrics, University of New Mexico Health Science Center, Albuquerque, NM, USA; 7) Clinical Translational Science Center, University of New Mexico, Albuquerque, NM, USA; 8) Division of Computer Science and Electrical Engineering, School of Computing and Engineering, University of Missouri-Kansas City, Kansas City, MO, USA.

Background: Whole genome sequencing (WGS) has the potential to inform drug choice and dosing after identifying genetic variation in pharmacogenes in a patient's genome. The highly polymorphic *CYP2D6* gene encodes one of the most important and highly polymorphic drug metabolizing enzymes. Due to the complexity of the gene locus, genotype analysis is challenging and often only includes common variants. Here we describe Constellation, a probabilistic scoring algorithm, which enables automated ascertainment of *CYP2D6* phenotype from WGS data. **Methods:** The Study was approved by the Institutional Review Board of Children's Mercy–Kansas City and included 61 subjects (7 HapMap; 54 patients/parents). Constellation uses a probabilistic scoring system to determine the most likely diplotype of a WGS sample for each locus. Data inputs included (vcf) files, a gene directory with chromosomal position, and allele definition files (allele definitions according to *CYP2D6* nomenclature at cypalleles.ki.se/. Constellation was implemented in Java. WGS was performed on HiSeq 2500 instruments (depth ~120GB; 2x100nt reads; TruSeq DNA PCR-Free kits). Variants were called with GSNAP 2012-`7-12 and GATK 1.6. All samples were genotyped using long-range (XL) PCR, TaqMan assays and a quantitative CNV assay. A subset of 34 subjects were Sanger sequenced to confirm or resolve differences between genotyping and Constellation calls.**Results:** When compared with genotyping (consisting of SNV, CNV and XL-PCR analysis) and Sanger sequencing, Constellation correctly called *CYP2D6* diplotypes for 56 of 61 subjects. Phenotype prediction between genotype, Sanger and WGS/Constellation were consistent for 57 subjects. Constellation also accurately identified 5 samples carrying a gene deletion and 2 subjects with a gene duplication. When combined with rapid WGS methods, Constellation can provide provisional *CYP2D6* phenotype prediction within 30 hours.**Discussion:** Although this study focused on *CYP2D6*, Constellation is generally extensible to other ADMER genes, and can be performed at marginal incremental cost in the setting of diagnostic WGS. Constellation depends on the quality of data input, i.e. sequence data and haplotype definition sets. Advancing NGS technology and refining allele/haplotype definitions will further improve the accuracy of Constellation-based phenotype prediction. Prospective studies of the clinical and cost effectiveness of WGS-clinical pharmacogenomics are, however, warranted.

681T

Comprehensive study of NAT2 acetylation status in the Greenlandic population. *F. Geller[1], B. Soborg[1], A. Koch[1], E. Birch[1], K. Bjorn-Mortensen[1], M. Blaszkewicz[2], K. Golka[2], J. Hengstler[2], S.W. Michelsen[1], L. Carstensen[1], A.C. Nordholm[1], M.M.B. Johansen[1], M.L. Børresen[1], B. Feenstra[1], M. Melbye[1,3,4].* 1) Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 2) Leibniz Research Centre for Working Environment and Human Factors, Dortmund, Germany; 3) Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark; 4) Department of Medicine, Stanford School of Medicine, Stanford, CA, USA.

N-acetyltransferase 2 (NAT2) is a well-studied phase II xenobiotic metabolizing enzyme relevant in isoniazid drug metabolism. Greenland has a high incidence of tuberculosis and isoniazid is widely used both in prevention and treatment of tuberculosis. Polymorphisms in the coding region of *NAT2* explain 88 % of the variability in isoniazid clearance. Individual enzyme activity is divided into three main categories as slow, intermediate and rapid and the distribution of NAT2 acetylation status is highly variable across the world.A recent clinical trial investigated a *NAT2* genotype guided treatment of tuberculosis, suggesting that a personalized dosage of isoniazid can improve both the safety and the efficacy of treatment. Therefore, we decided to study NAT2 acetylation status in 1,556 individuals from Greenland.As a first step, we determined NAT2 status based on four established SNP panels and the tagging SNP rs1495741. There was good concordance between the NAT2 status inferred by the different SNP combinations. Overall, the observed fraction of slow acetylators of 17.5% was one of the lowest in the world. The Greenlandic population is admixed with two major components: an Inuit part reaching back to the first migration waves from North America and a Northern European part that was introduced over the last 300 years. We split the study group in three parts according to the percentage of Inuit ancestry and observed a frequency of 12.2 % slow acetylators among individuals with a high percentage (>70%) of Inuit ancestry, which is substantially lower than the 25.6 % observed in individuals with low (< 50%) Inuit ancestry.Additional rare variants associated with slow acetylation are known, and we cannot rule out that these variants and maybe even Inuit-specific variants are present in Greenland. Thus, we currently investigate the agreement between the inferred NAT2 acetylation status and actual enzyme activity by measuring caffeine metabolites in urine. If prediction of NAT2 status by the applied genotype panel is good, our findings could lead to a practicable pharmacogenetics-based tuberculosis therapy in Greenland.

682W

Rapid implementation of a system-wide, high-throughput, EHR-integrated pharmacogenetics clinical service through cloud-based software automation. *M. Landsverk[1], M. Italia[2], M. Cornwell[1], L. Fricke[1], A. Rajendra[2], D. Schoolmeester[1], C. Larson[1], M. Mellacheruvu[1], V. Williams[1], C. Chan[1], A. Heid[2], A. Hsu[2], J. Hirsch[2].* 1) Sanford Health, Sioux Falls, SD; 2) Syapse Inc, Palo Alto, CA.

Adoption of precision medicine has the potential to improve quality of care while reducing toxicity, unnecessary treatments, and cost. In particular, availability of pharmacogenetics (PGx) data about a patient before a drug is ordered can improve patient safety. Sanford Health is one of the largest health systems in the nation with 43 hospitals and nearly 250 clinics in 9 states and 3 countries. In 2014, we initiated a program to integrate genomic medicine into adult primary care. A top priority was to implement a system-wide PGx clinical service with 3 components: 1) Fast-turnaround in-house analysis of a panel of drug metabolism genes, 2) delivery of clinically actionable results to Sanford Health's electronic health record (EHR) system, and 3) integration of PGx results into clinical workflow via real-time EHR alerts at the time of drug ordering. To ensure accurate decision support based on clinically valid gene-drug relationships, the panel run by Sanford Health's clinical molecular genetics laboratory only contains genes with clear clinical dosing guidelines published by the Clinical Pharmacogenetics Implementation Consortium (CPIC), including CYP2C19, CYP2D6, CYP2C9, CYP3A5, VKORC1, TPMT, and DPYD. Syapse Precision Medicine Platform (PMP) software was used to automate receipt of test orders from the EHR, convert nucleotide-based genotyping results to corresponding star (*) alleles, and convert star alleles to enzyme metabolizer status. These results were then automatically delivered back to the EHR. To determine the efficacy of Syapse PMP for clinical PGx services, we compared manual entry, data review, and result reporting of 60 samples submitted to our laboratory for PGx panel testing to automated processing by Syapse PMP. The manual process had a 5% error rate and took over 20 person-hours. One sample had an incorrectly spelled last name, one had the incorrect gender, and another was reported as wild-type for CYP2C19 instead of *1/*2 and thus would have been misreported as an extensive metabolizer instead of intermediate. Using Syapse PMP, no manual sample entry was required, eliminating the need for paper requisition forms. Data review and reporting were reduced to 2 hours, with no processing errors. Automation with Syapse PMP enabled Sanford Health to implement a clinical PGx service in less than 5 months and trigger best-practice alerts with patient-specific decision support content to fire upon ordering of a corresponding drug in the EHR.

683T

Exome sequencing outperforms chip-based testing for clinically useful pharmacogenetic variant interrogation: Implications for pre-emptive pharmacogenetic evaluation. *D. Ng[1], L.N. Singh[1], C.S. Hong[1], J.J. Johnston[1], J.C. Mullikin[2,3], L.G. Biesecker[1,2], NISC Comparative Sequencing Program.* 1) Medical Genomics & Metabolic Genetics Branch, MGMGB, NHGRI, NIH, Bethesda, MD; 2) NIH Intramural Sequencing Program, NHGRI, NIH, Bethesda, MD; 3) Comparative Genomics and Cancer Genetics Branch, NHGRI, NIH, Bethesda, MD.

One of the great challenges of pharmacogenetics is that the testing delays treatment decisions. The gold standard for testing is chip-based analysis of specific variants associated with these traits. While pharmacogenetic testing is not commonly performed, exome sequencing is increasingly used in research and clinical care for a variety of indications. Our objective was to pilot a future scenario where exome or genome sequencing is routine and could be used for pre-emptive pharmacogenetic screening. We set out to assess the utility of exome and genome sequencing for pre-emptive pharmacogenetic screening by comparing exome and genome sequence accuracy and coverage of 203 clinically-relevant pharmacogenetic variant positions selected from the Pharmacogenomics Knowledgebase and Clinical Pharmacogenetics Implementation Consortium and identify copy number variants in cytochrome P450 2D6 *(CYP2D6)* that affect the metabolizer phenotype. We recruited participants from the Washington DC metropolitan region to study the clinical application of next gen sequencing for precision medicine. We performed exome sequencing and baseline clinical testing on 973 volunteers. We also performed genome sequencing and genotyping with a pharmacogenetic chip assay on five of these 973. As expected genome sequencing had the highest sensitivity (1,003/1,015) for the interrogated variants. Surprisingly, exomes had very good sensitivity (875/1,015) and outperformed chip-based testing (260/1,015). Exomes had high genotype concordance (99%) with chip-based genotyping. *CYP2D6* copy number variants were identified in 57/973 exomes using XHMM and were validated by qPCR in 19/21 high confidence calls. Exomes outperformed chip-based genotyping in capturing more important pharmacogenetic variant positions and *CYP2D6* copy number variants for preemptive pharmacogenetic screening. The agnostic design of exomes enables it to capture a large proportion of pharmacogenetic variants, outperforming the current standard chip-based assay. These data suggest that pre-emptive next-gen sequencing will allow pharmacogenetic evaluation to become routine, as this sequencing is more widely deployed in clinical care.

684W

PGRN-seq v.2: A second-generation capture-sequencing reagent for prospective targeted sequencing of clinically relevant pharmacogenetic loci. *S. Scherer[1], X. Qin[1], A. Gordon[2], R. Fulton[3], E. Wieben[4], E. Mardis[3], D. Nickerson[5], R. Gibbs[1], the PGRN.* 1) HGSC/Mol & Human Gen, Baylor Col Med, Houston, TX; 2) Institute for Personalized Medicine, Mt. Sinai School of Medicine, New York, NY; 3) McDonnell Genome Institute, Washington University, St. Louis, MO; 4) Dept. of Biochemistry, Mayo Clinic, Rochester, MN; 5) Genome Sciences, University of Washington, Seattle, WA.

Using feedback garnered from users of the first version of this reagent, the Pharmacogenomics Research Network's Deep Sequencing Resource (PGRN-DSR) worked together with PGRN network members, pharmacogenomic community experts and NimbleGen to design a second generation capture-sequencing probe set that builds on successes realized with PGRN-seq v.1. Primary modifications to the target list reflect advances in the field of pharmacogenomics, address loci left unaddressed in the original and emphasizes clinical relevance. To this end, targets now include all of the current Clinical Pharmacogenetics Implementation Consortium (CPIC) drug-gene pairs with published clinical guidelines as well as genes with high CPIC priority and PharmGKB clinical annotation scores. The new 76 gene panel targets the entire CYP2D6 locus, critical HLA tag SNPs and once again includes probes aimed at all of the ADME and DMET microarray sites while maintaining the original design's low cost thus enabling both common and rare variant reporting. Initial data produced from a broad panel of Coriell trios exhibits outstanding performance. We will present further studies outlining performance and validation metrics for this new addition to the pharmacogenomic translation toolkit.

685T

Pharmacogenetics of antipsychotic dosing. *V. De Luca, A. Perreira Silva, G.H. Franca de Moraes, N. Hettige.* Dept Neurogenetics, Univ Toronto, Toronto, ON, Canada.

In the present contribution, we expand the work that we have done in the past using the CPZ dose equivalent standardization. The primary mode of treatment for schizophrenia is through the administration of antipsychotic (AP) medications. A major concern for physicians aiming to treat patients is to improve treatment outcome and reduce side effects associated with AP medication. Many studies have investigated the genetic polymorphisms of AP drug metabolizing enzymes, focusing mostly on drug response while few have investigated a genetic influence on AP dosage. For this study, we selected a panel of candidate genes, focusing on AP receptors and genes implicated in the neurobiology of schizophrenia to increase the gene coverage from Hettige et al (2014). We recruited 266 patients from the Centre for Addiction and Mental Health (CAMH) between the ages of 18 and 75. Participants' current AP and dosage were collected at the time of the interview and standardized according to Chlorpromazine equivalents (CPZe) as seen in Gardner et al (2010). We selected 185 SNP markers from 38 candidate genes involved in the neurobiology and pharmacological treatment of schizophrenia. The 185 SNP panel was genotyped using a customized Illumina Bead Chip. We corrected for ethnicity using the Principal Component Analysis to include only those of European ancestry. The top SNP was the rs13182501 in the HTR4 gene (p=0.003). While we demonstrate potential significant targets, after multiple testing correction none of the SNPs were associated with CPZe standardization. This suggests that different candidate genes involved in AP mechanism(s) should be investigated for this phenotype. Nevertheless, AP dosage is a valuable phenotype in psychiatric pharmacogenetics because it is clinically relevant and can be normalized using standard criteria.

686W

A Missense Mutation in *SVEP1* Influences On-Clopidogrel Platelet Aggregation in a Sex-Specific Manner. *J. Backman, R. Horenstein, B. Mitchell, J. O'Connell, A. Shuldiner, L. Yerges-Armstrong, J. Lewis.* Division of Endocrinology, Diabetes & Nutrition, University of Maryland at Baltimore, Baltimore, MD.

Clopidogrel therapy, often with aspirin, is standard of care for reduction of recurrent cardiovascular events in patients with acute coronary syndromes. Previous investigations have found genetic variation to be a significant determinant of on-clopidogrel platelet reactivity and incidence of clinical outcomes. Furthermore, while sex-specific differences in on-treatment platelet reactivity have been documented, identification of polymorphisms that contribute to gender-specific differences in clopidogrel efficacy has been limited.We perform sex-stratified analyses of exome chip (Illumina Inc.) data in 545 participants of the Pharmacogenomics of Anti-Platelet Intervention (PAPI) study to identify novel genetic variants that influence on-clopidogrel *ex vivo* platelet aggregation in a sex-specific manner. Extensive phenotype data was collected in all participants pre- and post-clopidogrel exposure (300 mg loading dose + 75 mg/d for 7 days). Platelet function was assessed in platelet-rich plasma by optical aggregometry using a PAP8E aggregometer (Bio/Data Corp.) after stimulation with adenosine diphosphate (20 μM). Association analyses were performed under a variance component framework that models the effect of genotype as an additive effect on the quantitative trait, while simultaneously adjusting for the effects of age, BMI, baseline platelet aggregation, and participant relatedness.Exome-wide analysis identified a variant in the Sushi, Von Willebrand Factor Type A, EGF, and Pentraxin Domains-Containing 1 (*SVEP1*) gene to be significantly associated with increased ADP-stimulated platelet aggregation post-clopidogrel exposure in females (rs10980419, P = 2.04 x 10-8, N=278), resulting in an approximate doubling of platelet aggregation between homozygote groups (maximal platelet aggregation = 33.3 for CC homozygotes, 44.1 for CT heterozygotes, and 68.0 for TT homozygotes). In contrast, no significant association was observed between rs10980419 and platelet aggregation in men (P = 0.23, N=267). A significant SNP x gender interaction was observed (P = 0.008).In this investigation, we identified a genetic variant in *SVEP1* that influences on-clopidogrel platelet aggregation in a sex-specific manner. *SVEP1* expression is known to be regulated by 17b-estradiol, strengthening our confidence in these results. Additional studies are required to replicate our finding as well as to determine the effects of this variant on recurrent cardiovascular events in women.

687T

Effect of SLCO2B1 Gene Polymorphism on the Lipid Lowering Effect of Rosuvastatin in Hypercholesterolemic Patients. *T.-E. Kim[1,2], D.-S. Shin[2,3], N. Gu[2,4], J. Choi[5], B. Jung[5], K.-S. Yu[2], I.-J. Jang[2], J.-Y. Cho[2].* 1) Department of Clinical Pharmacology, Konkuk University Medical Center, Seoul, South Korea; 2) Department of Clinical Pharmacology and Therapeutics, Seoul National University College of Medicine and Hospital, Seoul, Republic of Korea; 3) Clinical trials center, Gil hospital, Incheon, Republic of Korea; 4) Department of Clinical Pharmacology and Therapeutics, Dongguk University College of Medicine and Ilsan Hospital, Goyang, Gyeonggi-do, Republic of Korea; 5) Molecular Recognition Research Center, Korea Institute of Science and Technology, Seoul, Republic of Korea.

Rosuvastatin is an HMG-CoA reductase inhibitor widely used for the treatment of hypercholesterolemia. It is well known that the polymorphism of solute carrier organic anion (SLCO) 1B1 gene may affect the pharmacokinetics and pharmacodynamics of HMG-CoA reductase inhibitors. However, the effect of genetic polymorphism of SLCO2B1 has not been reported. The aim of this study was to investigate the effect of SLCO2B1 polymorphism on the lipid lowering effect of rosuvastatin in hypercholesterolemic patients.A total of 21 patients whose low density lipoprotein (LDL) level was over 130 mg/dL were participated in this study. For the subjects, three single nucleotide polymorphisms (SNPs) of SLCO2B1, c.935G>A, c.1457C>T and c.601G>A were evaluated. For 8 weeks subjects were administered 20 mg of rosuvastatin once a day. Lipid lowering effect of rosuvastatin was evaluated by the measurements of total cholesterol, LDL, high density lipoprotein (HDL), triglyceride (TG), apoprotein B (apo B) and apoprotein A1 (apo A1). Blood samples were taken before dosing and after finishing dosing of 8 weeks. The comparisons between genotypes were performed by Mann-Whitney U test. The genotyping were performed in 19 subjects. Among 19 subjects, 14 were CC and 5 were CT for c.1457C>T. As for c.935G>A, GG, GA and AA types were 7, 9 and 3, respectively. In the case of c.601C>T, all the subjects showed wild-type except one subjects of GA type. There was no significant difference in lipid lowering effect between genotypes with c.1457C>T and c.601G>A polymorphisms. However, in the case of c.935G>A, there were statistically significant difference between wild-type and variant-type in the lowering effect of LDL and ApoB. The LDL was decreased after the 8-week dosing by 62.8 ± 6.8% (mean ± SD) in wild-type (c.935GG) and 50.3 ± 11.5% in variant-type (c.935GA or AA) (p=0.035). The ApoB was decreased by 52.1 ± 7.1% in wild-type and 42.8 ± 9.1% in variant-type (p=0.039).In this study, it was demonstrated that SLCO2B1 polymorphism of c.935G>A can influence on the lipid lowering effect of rosuvastatin in hypercholesterolemic patients.

688W

Genotyping of New ADME Variants in a Phase 2 Axitinib Trial to Understand Exposure Variations. *J.C.M. Marshall[1], Y.P. Pithavala[1], B.J.S. Smith[2], L.W. Wood[1], S.L. Li[1], A.W. Williams[3], P.E. English[1].* 1) Pfizer, Groton, CT; 2) Gilead, San Mateo, CA; 3) Genentech, San Francisco, CA.

Axitinib is a selective 2nd generation inhibitor of vascular endothelial growth factor receptors (VEGFR) 1, 2, and 3 that is approved for the 2nd line treatment of patients with advanced renal cell carcinoma. This study undertook a retrospective analysis of a clinical trial set with AUC and CL data to understand if newly described genotypes could be used to predict variability in axitinib plasma exposure. Data from a Phase 2, randomized, double-blind study (n= 213 patients) with axitinib in treatment-naïve renal cell carcinoma (RCC) were used for this analysis. Patients (n=112) who were eligible for dose titration, based on pre-specified criteria were randomly assigned (1:1) to receive in combination with 5-mg BID either masked titration with axitinib or placebo. Patients who did not meet the criteria continued at 5 mg BID axitinib. The study met its primary endpoint of statistically significant higher objective response rate with active axitinib titration (54% ORR) versus placebo titration (34% ORR). A previously described population pharmacokinetic model was used to obtain post hoc oral clearance (CL/F) and steady-state AUC24 in each patient at the 5 mg BID starting dose. DNA from 172 patients was extracted from blood samples taken at enrollment. Genotyping for CYP3A4*22, CYP3A5*3,*6,*7 and POR*28 was carried out using TaqMan assays run on a QuantStudio 12K Flex system. The following genotyping frequencies were recorded for each gene of interest: CYP3A5 predicted poor metabolizer (PM) 77.9%, intermediate metabolizer (IM) 19.8%, extensive metabolizer (EM) 2.3%; CYP3A4 IM 7%, EM 93%, no individuals were PM (*22/*22). For POR*28, 51% patients were wild type (C/C), 41% were heterozygote (C/T) and 8% were homozygote variant (T/T). Multivariate statistical analysis was carried out to assess genetic effect for each gene polymorphism on AUC and CL after adjusting for important baseline characteristics.AUC analysis for POR*28 approached statistical significance (p = 0.0747 for higher AUC in TT vs. CC) no genotype was statistically significant for AUC or CL. While POR*28 shows potential, there is insufficient evidence from genotyping evaluated in this population to completely explain patient to patient differences in axitinib plasma exposure. Analysis of this population therefore indicated that genotype-based dose adjustment is not warranted for RCC patients receiving axitinib.

689T

Profiling of miRNA expression in Immune thrombocytopenia patients before and after Qishunbaolier (QSBLE) treatment. *B. Borjigin[1], E. Eerdunduleng[1], W. Huo[1], C. Gong[1], H. Gaowa[1], L. Long Mei[1], M. Li[1], X. Zhang[1], S. Wang[1], Y. Liu[2], H. Bai[1,3].* 1) Affiliated Hospital of Inner Mongolia University for the Nationalities, tongliao, China; 2) Department of Developmental Biology, Washington University in St. Louis School of Medicine, St. Louis, MO 63110, USA; 3) School of Life Science, Inner Mongolia University for the Nationalities, Tongliao, Inner Mongolia 028000, China.

Immune thrombocytopenia (ITP), also known as idiopathic thrombocytopenic purpura, is an autoimmune disease characterized by low platelet count and increased bleeding tendency. Currently, glucocorticoid and splenectomy are the main therapies for ITP but with obvious side effects including tendency of relapse and risk of internal bleeding. Searching for a safer alternative treatment from traditional medicine represents an active research direction. Mongolian medicine Qishunbaolier (QSBLE) (also called Xuebaowan or Shengxuekang) has been successfully developed for treatment of ITP, Aplastic Anemia, Allergic Purpura, myelodysplastic syndromes, and other bleeding disorders since the 1960s. QSBLE is a compound preparation made of gardenia, bezoar, Cornu Bubali, bear gall, saffron, Lithospermum, and Rubia. We followed closely 32 cases of ITP patients administrating QSBLE and found that QSBLE can not only significantly increase the number, quality, and life span of platelets, but also improve the immunity and hematopoietic function of ITP patients. We did not observe any side effect that normally associated with glucorticoids treatment including nausea, vomiting, increased heart rate et al.To better understand the ITP pathogenesis and therapeutic basis of QSBLE, we profiled miRNA expression in the blood sample of ITP patients and identified 44 miRNAs that are differentially expressed in ITP patients before and after QSBLE treatment. Interestingly, 25 of them are expressed in control subjects and are downregulated in ITP patients, whereas treatment with QSBLE restores their expressions to the level of control subjects. This result suggests that abnormal expression of these 25 miRNAs might be connected to the pathogenesis of ITP. Furthermore, 14 of those 44 miRNAs are predicted to target at least once on 31 ITP associated genes, indicating the possible mechanism of QSBLE on ITP therapy. This work is supported by the grants for the Key Project in the Chinese National Science & Technology Pillar Program during the Twelfth Five-year Plan Period (2012BAI27B02).

690W

Rare variants in ALOX5 may be associated with the risk of hypersensitivity reaction to penicillin . *R. Pellegrino, B. Almoguera, L. Vazquez, D. Abrams, T. Watanabe, L. Tiang, J. Garifallou, F. Mafra, J. Conolly, F. Mentch, P. Sleiman, H. Hakonarson.* CAG, Childrens Hospital of Philadelph, Philadelphia, PA.

Penicillin allergy is the most commonly reported medication allergy and it limits these antibiotics' use. There is evidence that genetic factors play a significant role in the risk of penicillin allergy and several efforts have been applied to identify genetic candidate genes such as *IL4*, *IL6*, *IL10* or *TNF*. However, results to date have not been replicated and genetic risk factors remain largely unknown. Here we investigate the role of rare variation in 82 genes related to the disposition of a variety of drug compounds in the risk of penicillin allergy. Cases with penicillin allergy were defined by specific **ICD9 codes V14.0, E930.0, and 995.27**, as documented in their electronic health records and allergic symptoms being consistent with hypersensitivity reaction, such as anaphylaxis, hives, angioedema, or respiratory symptoms. Control population was defined as subjects exposed to penicillins and no documentation of allergic symptoms. Sequencing of the complete coding and nearby regulatory regions of the 82 genes was performed using a custom-sequencing panel developed by the Pharmacogenomics Research Network (PGRN). For rare variant gene-level association analysis, we used three complementary algorithms: the Sequence Kernel Association Test, the Variant Threshold test and the Combined Multivariate and Collapsing test. Single variant analysis was also conducted using a Logistic Score Test. 125 subjects with penicillin allergy and 58 controls of European ancestry were identified for the study. We found *ALOX5*, a gene located in chromosome 10 approaching statistical significance (p=6.50E-04). The *ALOX5* gene is a member of the lipoxygenase gene family with a role in the synthesis of leukotrienes from arachidonic acid, which are important mediators of inflammatory and allergic conditions. These results show great promise in the identification of the genetic underpinnings of penicillin allergy and we are currently undergoing replication in an additional cohort.

691T

Next generation sequencing technologies improve pharmacogenetic-guided drug dosing. *I. Cohn[1,8], T. Paton[6], C.R. Marshall[2,6,7], R. Basran[4,7], P.N Ray[2,4,6,7], N. Monfared[2], P. Sinajon[3,7], R.D. Cohn[2,3,4,5,6], S. Ito[1,5,8].* 1) Division of Clinical Pharmacology and Toxicology, The Hospital for Sick Children, Department of Pediatrics, University of Toronto, Toronto, Ontario, Canada; 2) Center for Genetic Medicine,The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Department of Pediatrics, University of Toronto, Toronto, Ontario, Canada; 4) Department of Molecular Genetics, University of Toronto, Ontario, Canada; 5) Department of Pediatrics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 6) The Center for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 8) Program in Physiology and Experimental Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada.

Pharmacogenetics provides opportunities to optimize drug treatment and prevent adverse drug events. Currently, targeted genotyping strategies are used clinically to screen markers in genes with well-characterized drug-gene interactions. Next generation sequencing (NGS) technologies carry the promise to identify known and novel pharmacogenetic variants that are not covered by existing clinical panels. Here, we determined the accuracy of variant calls from NGS platforms and investigated additional variants of potential pharmacotherapeutic benefit. We performed a systematic comparison of 74 pharmacogenetic variant loci from whole genome sequencing (WGS, Complete Genomics) and mass spectrometry (Agena Biosciences) for 98 pediatric patients of varying ethnicities. Additionally, 11 patient samples of this cohort underwent whole-exome sequencing (WES, Illumina HiSeq2500).The 74 loci were sequenced to an average read depth of 20X or greater across the 98 samples that underwent WGS. However, there was considerable variability in coverage between cases, with 32 loci exhibiting low read depth (5X or less) or poor variant-quality reads, mostly within the CYP2D6 gene. However, these variant loci were not always required for metabolizer status assignment, as they are used to further subtype functional classifications. In contrast, WES exhibited superior coverage in the majority of the 74 loci (3 loci are in non-coding regions). Surprisingly, discordant genotypes were identified for 7 variants in the WGS data but only 1 variant in the WES data when compared to the mass spectrometry platform. At present, detection of copy number variants in the CYP2D6 gene cannot be easily interpreted from NGS and require separate testing by qPCR. In the WGS data, we identified 3 coding variants with potential clinical implications that change the metabolizer status or dosage recommendation in CYP2C9, CYP2D6 and DPYD.As WGS/WES becomes more routine for identification of primary diagnoses there is opportunity to gain relevant information about pharmacogenetic markers. Our data demonstrate that clinically important pharmacogenetic variants covered by targeted genotyping are accurately called by NGS platforms. Additionally, functional variants of potential clinical pharmacogenetic significance could be missed by using targeted genotyping panels, indicating that NGS technologies need to be considered for a more genome-wide approach to pharmacogenetic testing of individual patients.

692W

Comprehensive exploration of the high-risk rare variants for the cold medicine–related Stevens-Johnson syndrome/ toxic epidermal necrolysis (CM-SJS/TEN) with severe ocular complications. *Y. Hitomi[1], S. Khor[1], M. Ueta[2,3], H. Sawai[1], K. Zawlatt[1], C. Sotozono[3], S. Kinoshita[2], K. Tokunaga[1].* 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Department of Frontier Medical Science and Technology for Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan; 3) Department of Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan.

Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) are acute inflammatory vesiculobullous reactions of the skin and mucous membranes, and these reactions are reported to be caused by inciting drugs, viral infections, or malignant tumor. Although the occurrence of SJS/TEN is rare at about 1-6 cases per million, mortality rates are higher than other drug rush (SJS: 3%, TEN: 27%). In addition, Quality-of-life of the most survivors is often low, as severe ocular surface complications are often developed as the aftereffect.*HLA-A* and *IKZF1* have been recently reported as susceptible genes for Cold Medicine–Related SJS/TEN (CM-SJS/TEN) in Asian populations by genome-wide association study (GWAS). However, other genetic factors of CM-SJS/TEN including rare variants and structural variants remains to be discovered.In order to identify the functional variants for CM-SJS/TEN with severe ocular complications, whole-exome sequencing was performed in 128 Japanese CM-SJS/TEN patients with severe ocular complications using next generation sequencer (NGS) - ion Proton (Thermo-Fisher Scientific). After data cleaning, the following average scores per individual were acquired: 7.23±1.09 billion bases; 42.5±6.0 million reads; 169±7.59 of read length; 94.8±0.8 % of on target reads; 89.6±3.0 % of uniformity; x115.5±17.7 of coverage. Finally, 50,722±736 variants were called by Torrent suite software with appropriate parameters using NCBI hg19 as the reference.In order to evaluate the accuracy of variant calling, we compared the genotypes of approximately 5,000 variants which were overlapped on the Axiom chip (Affymetrix) from our previous GWAS using same CM-SJS/TEN sample set (Ueta M *et al., J Allergy Clin Immunol* 2015). The concordance rate of genotypes between GWAS and exome sequencing was over 99%.The identification of high-risk rare variants which have extreme effect on the pathogenesis of CM-SJS/TEN are carried out by comparing variants identified with approximately 1,000 Japanese individuals whole-genome sequencing database (Tohoku-Medical Megabank). In addition, pathway analysis will be carried out by SNP-set Kernel Association Test (SKAT) software. In this presentation, we will report the preliminary result of the identification of high-risk rare variants. This study may illustrates novel diagnostic and therapeutic methods for CM-SJS/TEN.(Equally contributed: Hitomi Y, Khor SS, Ueta M).

693T

Investigation of the role of rare variation in eighty-two pharmacogenes in the risk of vincristine-induced peripheral neuropathy in children. *B. Almoguera, L. Vazquez, D. Abrams, T. Watanabe, L. Tiang, R. Pellegrino, J. Garifallou, F. Mafra, J. Connolly, F. Mentch, P. Sleiman, H. Hakonarson.* Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

The dose-limiting toxic effect of vincristine, a drug prescribed to over 50% of children with cancer, is peripheral neuropathy, characterized by neuropathic pain and sensory and motor dysfunction. There is evidence that genetic factors play a significant role in the risk of neuropathy, however the precise risk factors remain elusive. In this study we investigated the role of 82 genes related to the disposition of a variety of drugs in the risk of developing vincristine-induced neuropathy in children. Phenotyping was performed by direct mining of the electronic health records (prescription of vincristine + ICD9 code for neuropathy, 355x) and identified 96 individuals with vincristine neuropathy and 52 controls that were exposed and confirmed negative by chart review. Sequencing of the complete coding and nearby regulatory regions of the 82 genes was performed using a custom-sequencing panel developed by the Pharmacogenomics Research Network (PGRN). For rare variant gene-level association analysis, we used three complementary algorithms: the Sequence Kernel Association Test, the Variant Threshold test and the Combined Multivariate and Collapsing test. Single variant analysis was also conducted using a Logistic Score Test. Neuropathy was classified and analyzed as autonomous, sensory, and motor symptoms, and also as a continuous trait. Preliminary results on 43 cases and 23 controls evidenced that *CYP3A4*, which is directly involved in vincristine metabolism, was nominally significantly associated with the risk of neuropathy (p=1.10e-03). Also, *SLC6A6*, a taurine transporter, was associated with the risk of sensory neuropathy (p=7.91e-04) Absence of *SLC6A6* in mice has been demonstrated to lead to several neurologic deficits and sensory losses. These results show great promise in the identification of the genetic underpinnings of vincristine-induced neuropathy and are currently undergoing replication in an additional cohort of 53 cases and 29 controls.

694W

Deep Sequencing of Genes Associated with Glucocorticoid Response in Asthma. *Q.L. Duan[1], M.H. Cho[1], J.A. Su[1], K.G. Tantisira[1,2], S.T. Weiss[1,3].* 1) Channing Division of Network Medicine, Brigham & Women's Hospital and Harvard Medical School, Boston, MA; 2) Pulmonary Division, Brigham & Women's Hospital, Boston, MA; 3) Partners Center for Personalized Genetic Medicine, Partners Health Care, Boston, MA.

Glucocorticoids (GCs) have potent, anti-inflammatory effects, and are commonly used for treatment of diseases including asthma, which occurs in over 300 million individuals worldwide. However, there is substantial inter-individual variability in GC response, with 25% of patients non-responsive. Multiple genetic loci have been associated with GC response but account for only part of the variability. We hypothesize that deep sequencing of loci from genome-wide association studies (GWAS) will identify novel genetic variations that could constitute part of the missing heritability of GC response. To test this hypothesis, we identified top associated genes from a meta-GWAS of four asthma clinical trials (N=723) and whole exome sequencing (WES) in an asthma cohort of extreme GC responders (N=188). In the meta-GWAS, linear regression was performed in PLINK using an additive model, with covariate adjustments (height, age and gender) and GC response measured as a percent change in lung function (baseline pre-bronchodilator forced expiratory volume in one second) following 4-8 weeks of therapy. WES data used the NimbleGen SeqCap EZ human exome v2.0 capture array and sequenced on the Illumina Hi-Seq 2000. Sequence alignment used BWA, and variants were called with Samtools. Our primary analysis included 75,655 coding (non-synonymous, splice, or stop) variants based on annotations from SnpEff, and used a per-gene burden test in SKAT, adjusted for covariates (gender, age, and height, and baseline FEV1). In total, we identified 46 genes from the meta-GWAS ($P < 0.001$), that were nominally associated ($P < 0.05$) with extreme GC response in our gene-based test. Our results include *KCNQ5,* which encodes a potassium channel in airway smooth muscle cells from human bronchioles and regulate airway constriction as well as members of a multidrug transporter family (*ABCA4* and *ABCC1*). Targeted sequencing of 30 such genes, prioritized by P values and biological plausibility, is underway in three asthma clinical trials (N=984) for validation analysis. In summary, we identified potentially deleterious variants from WES within loci previously associated with GC response in GWAS of asthmatics.Please address all inquiries to qingling.duan@channing.harvard.edu.

695T

Influence of common and rare genetic variation on warfarin dose among African Americans and European Americans using the exome-array. *N. Liu[1], A. Patki[1], M. Irvin[4], D. Zhi[1], T. Brown[3], C. Hill[5], M. Beasley[1], D. Nickerson[6], N. Limdi[2].* 1) Dept Biostatistics, Univ Alabama at Birmingham, Birmingham, AL; 2) Department of Neurology, University of Alabama at Birmingham, AL; 3) Division of Cardiovascular Diseases, University of Alabama at Birmingham, AL; 4) Department of Epidemiology, University of Alabama at Birmingham, AL; 5) Department of Pathology and Laboratory Medicine, Emory University, Atlanta, GA; 6) Department of Genome Sciences, University of Washington, Seattle, WA.

Personalized management of warfarin, the most widely used oral anticoagulant, is challenging. The large inter-patient variability in warfarin dose requirements and narrow therapeutic index make warfarin a "top offender" accounting for 33.3% of all adverse-drug-related hospitalizations in the US, with hemorrhage representing 63.3% of admissions. Achieving therapeutic anticoagulation rapidly and predictably is critical for safe and effective therapy. To this end multiple investigations have identified the influence of clinical, demographic and genetic factors on warfarin dose. To date, candidate gene and genome-wide associations studies (GWAS) have confirmed that the majority of the genetic influence is accounted by single nucleotide polymorphisms (SNPs) in two genes: *CYP-2C9*and *VKORC1*. Cytochrome P450 4F2 (*CYP4F2*) was recently shown to account for a small but significant proportion of variability in warfarin dose among European Americans.We hypothesize that novel coding SNPs (common and rare) may harbor additional clues that can explain variaiblity in warfarin dose. Toward this end, we assessed 247,870 SNPs using the Illumina Exome Array in 1,240 warfarin users (701 European Americans, 539 African Americans).Results show *VKORC1* (rs9923231) was the single most important genetic influence on warfarin dose in both European-Americans ($p<9.28E-45$) and African-Americans ($p<1.80E-8$). This association explains 22.8 % of the variance in dose among European-Americans 6.5% in African-Americans.Among European-Americans the single most important *CYP2C9* SNP was rs4086116 ($p= 1.24E-19$) as reported previously, *CYP2C9*3* (exm844046; $p=1.01E-13$) demonstrated significant influence on warfarin dose as did *CYP2C9*2* (exm-rs1799853; $p=2.34E-7$). Although significant ($p=0.03$), inclusion of rs4086116 explains little additional (0.6%) variance in warfarin dose, similar to that explained by *CYP4F2* (0.6%; $p=0.004$).Among African-Americans no *CYP2C9* SNP demonstrated statistically significant association with dose. The SNP rs12777823 identified in the GWAS of African-American was marginally significantly associated with dose ($p=1.93E-5$ with FDR adjusted q-value = 0.184) explaining an additional 4% of the variance in warfarin dose. *CYP4F2* did not demonstrate significant influence on dose.In addition, we identified novel significant association of single markers. We are validating our findings in an independent cohort.

696W

Rescuing the un- sequenceable: using Exome Sequencing and HRM to reliably and cost-effectively genotype non-trivial loci. *M.D. Napier[1], R.P. Loewe[2].* 1) Agilent Technologies, Santa Clara, CA; 2) Genewake GmbH, Neuried, Bavaria, Germany.

To move beyond the modest drug responses seen in oncology and immunology today, cutting edge researchers must generate better answers by changing the mode of questioning in clinical research. We should shift gears away **from** genomic biomarker associations with drug response, **toward** haplotype associations with drug response phenotypes. To tackle this challenge we must overcome two research obstacles: (1) obtain the proper tools to scan the exome comprehensively and (2) enrich cancer exome studies with targeted qPCR for fine loci mapping. Special emphasis in our study was given to: the tyrosine kinase family especially to EGFR and ERBB2 and the cytochrome p450 family for their pharmacogenomics utility. Together these two techniques will rescue loci that were previously unavailable or under-represented to the genetic researcher. The methods utilized include: Sure Select Human All Exon v6for exome capture followed by Next Generation Sequencing(NGS), qPCR as well as high resolution melting (HRM). Our study's objective was to find improved patient stratification methods for research. Our combined methodology gives a viable and easy option for a standard laboratory to venture deeper into the cancer and immunology genome. This workflow enables fine mapping of loci using NGS complemented with qPCR for accelerating clinical research beyond the event horizon of what is currently known.

697T

Novel genotyping algorithms for multiallelic and copy number ADME variants. *J.P. Schmidt, J. Gollub, D. Oliver, R. Varma, J. Brodsky, C. Bruckner, B. Eynon, S. Malakshah, M. Mittmann, A.H. Roter, B. Wong, M.H. Shapero, T. Webster.* Informatics, Affymetrix, Santa Clara, CA.

There are many well-known and well-characterized, clinically relevant mutations in genes associated with drug absorption, distribution, metabolism, and elimination (ADME). Most of these variants are biallelic single nucleotide polymorphisms (SNPs) or short insertion/deletion (indel) variants, and thus amenable to standard genotyping assays and calling algorithms. However, copy number variations (CNVs) in critical genes, as well as multiallelic SNPs or indels, can be more challenging to genotype. We have developed novel probe design methods and calling algorithms to address these variants with Axiom® Genotyping Solution.Genotyping multiallelic variants requires the design of a set of probes, each specific to a single allele. This is straightforward for SNPs, but more complex when indel alleles may be present. Under the assumption of diploidy, genotype calling proceeds in three steps, using the signals from those probes: 1) background estimation for each allele; 2) all-by-all biallelic clustering of appropriate samples for initial genotype assignment; 3) final assignment to a genotype "cluster" using an N-dimensional Gaussian mixture model with posterior distributions informed by the initial assignments in (2) and conjugate priors that may be either generic or trained on previous data. Using this unique approach we are able to address the underlying variations for important alleles such as CYP2C8*7 and *8, or CYP2D6*8 and *14.CNV calling in the ADME context presents challenges that are in some ways both easier and more difficult than those in exploratory analysis. There are only a small number of defined regions of interest, so probe design and analysis parameters can be optimized for them, and precise determination of breakpoints is not essential. However, the regions are smaller than normal for CNV calling, and in some cases the presence of highly homologous pseudogenes or extremes of GC content reduce specificity. We have developed new methods for the design and selection of maximally informative probes, correction of signals using a variety of covariate factors, and signal summarization and thresholds for calling copy number states from zero (homozygous deletion) to three or more (amplification) in five genes of great relevance to ADME: CYP2A6, CYP2D6, GSTM1, GSTT1, and UGT2B17.

698W

Variable read-through by small molecule drugs in Leber Congenital Amaurosis (LCA16) suggests need for personalized approach to read-through strategies. *D.M. Pillers[1, 2], P. Shahi[1, 2], V. Bakhutashvili[3], S. Brar[1, 2], R. Gatti[4,5], B.R. Pattnaik[1, 2, 6].* 1) Pediatrics, University of Wisconsin-Madison, Madison, WI; 2) McPherson Eye Research Institute, University of Wisconsin-Madison, Madison, WI; 3) School of Medicine and Public Health, University of Wisconsin-Madison, Madison, WI; 4) Pathology & Laboratory Medicine, UCLA, Los Angeles, CA; 5) Human Genetics, UCLA, Los Angeles, CA; 6) Ophthalmology and Vision Sciences, University of Wisconsin-Madison, Madison, WI.

Introduction: Leber congenital amaurosis (LCA [MIM 204000]) is asevere vision disorder that presents in infancy with progression to blindness. We studied LCA16 [MIM614186] which is due to defects in the Kir7.1 ion channel encoded by *KCNJ13* [MIM 603208].We sought to determine the response to small molecule read-through drugs by measuring their physiologic response as a marker for improvement in function. **Methods:** Chinese Hamster Ovary (CHO-K1) cells were transfected with N-terminal GFP-fused wild-typeor W53X (c.158 G>A)or R166X (c.496C > T) mutant plasmids. The cells were treated with 5 μM and 10 μM of the read-through compounds Gentamicin, PTC-124, RTC-13,or RTC-14 after eight hours of transfection, and the cells were incubated with these drugs for 36 hours. Whole-cell patch clamp electrophysiology was performed on the transfected cells. Function of the Kir7.1 channel was measured in the presence of Cs+ to block function, or the highly permeant Rb+ to enhance current. The students T-test was used and significance was determined at the $P<0.05$ level.**Results:** Both W53X and R166X transfection resulted in non-measurable Kir7.1 current as compared to the wild-type Kir7.1 channel, and the cells were depolarized. Upon treatment with Gentamicin, PTC-124 and RTC-13,we detected no measurable difference in either the current amplitude or the resting membrane potential (Vm). In contrast, RTC-14 showed partial rescue of both the Kir7.1 current amplitude and Vm. R166X expressing cells responded to RTC-14 treatment by a 15 mV hyperpolarizing shift in Vm (-25.92 ± 2.7 to -40.16 ± 4.05 mV; $P<0.02$) and an increase in current amplitude measured at -150 mV from -98.5 ± 24.7 to -122.8 ± 28.15 pA, $P = 0.52$. W53X, however, showed a 33 mV hyperpolarization shift in Vm (-29.37 ± 3.55 to -61.94 ± 2.9 mV; $P<0.001$). We also noticed a four-fold augmentation in both inward (-80.4 ± 11.2 to -316.16 ± 60.8 pA at -150 mV; $P<0.01$) and outward (-7.4 ± 2.4 to 33.3 ± 11 pA at -40 mV; $P<0.02$) current amplitude. **Conclusions:** We have shown that a measurable outcome such as the electrophysiology of the Kir7.1 ion channel at the cellular level can serve as a useful marker in the visual system to detect a response to translational read-through drugs. We have also shown that different mutations within the same gene may have very different responses to the same read-through drugs. This suggests that small molecule read-through therapies should beexplored using a personalized approach.

699T

***in silico* analyses of complex systems for the development of novel therapeutics in rare and orphan diseases.** *K. Nguyen[1,4], D. Li[1], L. Tian[1], R. Pellegrino da Silva[1], C. Kao[1], R. Pandey[1], M. Bakay[1], P. Sleiman[1,2,3], H. Hakonarson[1,2,3].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Center for Dynamic Data Analytics, Rutgers University, Piscataway, NJ.

We introduce a structural approach that incorporates modeling, simulation, and drug design to our translational, precision, and network medicine pipeline for the purpose of discovering novel therapeutics in rare and orphan diseases. The importance of an accurate model, whether acquired through experimental or computational methods, is vital to this approach. Once validated, the structure, dynamics, and energetics are calculated to determine if the overall conformation has converged enough using molecular simulation and high performance and parallel computing. It has been reportedthat modeling alone is not accurate enough and does not adequately represent the conformational space of the system in question – especially when it is the prerequisite for virtual screening and structure-activity relationships. Once the simulated structure has been validated through established metrics, determining targets for ligands/drugs for therapeutic purposes, elucidating ligand-receptor interactions, and linking those results to function are the final steps before analytical and instrumental validation. Of course, these steps may take several iterations, but are necessary to determine a successful set of candidates. Here, we present an overview of two systems using the resources described in our pipeline – ephrin type-B receptor 4 (EPHB4) and C-X-C chemokine receptor type 1 (CXCR1, also known as interleukin 8 receptor – alpha, IL8RA) - both of which represent novel disease associations with supporting functional data.

700W

Variation in transcriptional response to Vitamin D and LPS in monocytes. *S.N. Kariuki, J.D. Blischak, D. Witonsky, A. DiRienzo.* Human Genetics, University of Chicago, Chicago, IL.

Vitamin D plays an important immunomodulatory role, regulating genes in the monocyte microbicidal pathway involving toll-like receptors (TLRs). Activation of TLRs by bacterial ligands such as lipopolysaccharide (LPS) results in production of 1,25-dihydroxyvitamin D3 (1,25D), which induces transcription of genes. Studies have shown correlations between low vitamin D levels and risk of various diseases, but the mechanisms underlying these observations are poorly understood. We aim to characterize inter-individual and inter-ethnic variation in monocyte transcriptome-wide response to 1,25D and LPS*in vitro*.Primary monocytes obtained from 10 African-American (AA) and 10 European-American (EA) healthy donors were treated with 1,25D, LPS, and ethanol for 24 hours. Transcript levels of RNA were measured on the Illumina HumanHT12 microarray. Genome-wide differential expression (DE) was analyzed using linear mixed-effect models and a joint Bayesian analysis to identify patterns of shared and differential response across treatments.At a false discovery rate (FDR) < 0.01, we identified 2888 and 4461 genes that were significantly DE in response to 1,25D and LPS respectively, while4720 genes were significantly DE in response to1,25D plus LPS. The joint Bayesian analysis identified four broad patterns: non-DE genes, and DE genes responsive to 1,25D, LPS, and 1,25D plus LPS. Gene ontology analysis of these patterns showed enrichment of metabolic processes and immune response pathways among the 1,25D responsive genes, while cell proliferation and fatty acid oxidation pathways were enriched in LPS responsive genes. We identified 12 genes with differences in response to 1,25D plus LPSacross the two ethnic groups at an FDR < 0.10, including*PPAP2B,*which encodes a member of the phosphatidic acid phosphatase (PAP) family, and*AKNA*which encodes a transcription factor that activates expression of the CD40 receptor and ligand.In summary, we identified genes that were significantly DE in response to 1,25D and LPS, and were significantly enriched for immune response and metabolic functions. Treatment of monocytes with both 1,25D and LPS resulted in greater transcriptional response, increasing our power to detect genes with significant inter-ethnic DE within our small sample set. We speculate that the inter-ethnic differences in response to vitamin D in the presence of LPS could contribute to inter-ethnic disparities in immune-related diseases.

701T

Association of gene polymorphism in detoxification enzymes and urinary 8-OHdG levels in traffic policemen exposed to vehicular exhaust in Telangana State. *P.Reddy Penagaluru[1], B. Siva Prasad[1], C. Prashanth[1], P. Vidyullatha[2], T. Gudimella Vani[2], P. Rekha V. Devi[3], P. Usha Rani[2], M. Hema Prasad[2].* 1) Genetics, Bhagwan Mahavir Medical Research Centre, Hyderabad, Telangana State, India; 2) Institute of Genetics and Hospital for Genetic Diseases, Osmania University, Begumpet, Hyderabad; 3) Toxicology Unit, Biology Division, Indian Institute of Chemical Technology, Tarnaka, Habsiguda, Hyderabad.

Abstract Context: With rapid economic growth and massive development of transportation, the number of automobiles has greatly increased. Traffic police are the one of the vulnerable groups predominantly exposed to vehicular exhaust during traffic control. Objective The present study is aimed to study the relation between occupational exposure to vehicular exhaust and oxidative stress (OS) in traffic police. We investigated the levels of 8- hydroxydeoxyguanosine (8-OHdG), one of the most sensitive biomarkers for measuring OS and the association between polymorphisms in Cytochrome P450 (CYP) and Glutathione S-Transferase (GST) genes that are known to play a significant role in the activation and detoxification of xenobiotics. Materials and methods: 148 non smoking male traffic policemen and 135 control subjects were selected for this study. The 8-OHdG levels were analyzed by liquid chromatography with electrochemical detection method. Gene polymorphism was detected by multiplex PCR and RFLP method. Results: 8-OHdG levels were found to be increased in traffic police with increase in the years of service in traffic control ($p = 0.02$) when compare to the controls. The results showed a significant increase in urinary 8-OHdG levels in mutated CYP1A1m1 ($p < 0.007$) and null GSTM1 ($p < 0.01$) genotypes. However the genotype frequencies of CYP1A1 m2 and GSTT1 genes did not vary in both exposed and control groups. Conclusion: Our study suggests that exposure to vehicular exhaust over a period of time increases oxidative stress and subsequently induces oxidative DNA damage in traffic policemen. Preventive and therapeutic strategies may be considered for traffic policemen to minimize the adverse effects due to vehicular exposure.

702W

Molecular Modelling and Docking Analysis of *katG* and *rpoB* Genes in Clinical Isolates of Multi Drug Resistant Tuberculosis Patients from Sahariya Tribe of North Central India. *R. Prakash[1], R. GUPTA[1], V.M. KATOCH[2,3], P.K. TIWARI[1].* 1) CENTRE FOR GENOMICS, MOLECULAR & HUMAN GENETICS, JIWAJI UNIVERSITY, GWALIOR, MADHYA PRADESH, India; 2) DEPARTMENT OF HEALTH RESEARCH, MINISTRY OF HEALTH AND FAMILY WELFARE, GOVT. OF INDIA, NEW DELHI; 3) INDIAN COUNCIL OF MEDICAL RESEARCH, NEW DELHI INDIA.

Tuberculosis caused by *Mycobacterium tuberculosis*, requires a multi-drug therapy approach. Isoniazid and Rifampicin are two potent first line drugs. Alteration in drug binding due to mutation in *katG* and *rpoB* genes is an important factor for resistance to isoniazid and rifampicin. To investigate the drug-drug target interactions for these two clinically important drugs, mutations in *katG* and *rpoB* genes, of MDR-TB patients from Sahariya tribe were identified and molecular models were prepared. Multiple sequence analysis revealed three mutations (Arg463Leu, Asp529Thr and Asp529His) for *katG* gene and one most common mutation (Ser-531Leu) in *rpoB* gene. The docking analysis showed that isoniazid binds to wild type katG protein at one of its active site residue HIS108 with one hydrogen bond formation, whereas, the drug binds to different positions for two of its mutant forms, i.e., Arg463Leu and Arg463Leu & Asp529Thr, with formation of one hydrogen bond at different residues, LYS274 and HIS270, respectively. The drug binds to third mutant (Arg463Leu & Asp529His, a semi conservative substitution) at TYR229 and VAL230 residues forming 2 hydrogen bonds. Interaction of rifampicin with wild type rpoB protein revealed formation of 1 hydrogen bond at LYS799 residue along with the formation of 1 oxygen bond at THR829 residue, a hypothesized active site for rifampicin binding, since no crystal structure is available for rpoB protein till date. Whereas, its interaction with mutant protein (one of the most common mutation in 81bp RRDR) resulted formation of 1 hydrogen bond at ILE522 residue with bond distance of 2.23 Å, smaller than bond distance at THR829 residue 2.46 Å. Functioning of drug mainly depends on its binding on active site or residues. In mutant models, the inhibition of drug function may be due to change in binding residues or difference in formation of hydrogen bonds, which needs further investigations. Thus, the knowledge of structural mechanism of drug resistance could pave the way for design of new drug targets to tackle the situation of multi-drug resistant tuberculosis.

703T

Preliminary analysis of polymorphisms in the *ABCB1* and *ABCC1* genes associated to virological failure to antiretrovirals in a Mexican Mestizo population. *F.F. González-Galarza, I.J. Varela-Marrufo, D. Delgadillo-Guzmán, R.D. Arellano-Pérez Vertti, F.C. López-Márquez, J.R. Argüello.* Center for Biomedical Research, Autonomous University of Coahuila, Torreon, Coahuila, Mexico.

In recent years, there has been a significant increase in the number of studies that are being performed to understand the pharmacogenomics of highly active antiretroviral therapy (HAART). This treatment is commonly used in HIV infections by promoting the suppression of virus replication. The antiretroviral effect has been shown to vary among individuals of different ethnicities in which the existing polymorphism in the *ABCB1* and *ABCC1* genes has been suggested to provide evidence for the heterogeneity of the response. At present, many of these studies have been primarily carried out in individuals of Caucasian ancestry. Thus, there is a lack of analysis carried out in populations from other ethnic groups such as in the Mexican Mestizo population. In this study, we analyzed the prevalence of two SNPs (rs1045642 and rs212091) associated to virological failure to HAART in one hundred randomly-selected individuals from the Comarca Lagunera region that were SNP genotyped using the commercially-available Illumina HumanOmniExpress BeadChip. We found that the genotype frequencies for the rs1045642 were: AA-31%, GG-31% and AG-38%. In the case of the rs212091 the CT genotype was present in 20% and TT in 80% of the individuals whereas the CC genotype was absent. Additionally, 150 HIV patients have been included to evaluate the prevalence of these SNPs which may help researchers and clinicians in the personalized treatment of HIV infected patients in Mexico.

704W

X-linked genes with novel rare variants identified by WGS in ASD patients are involved in Neurodevelopment. *V. Chini[1], K. Shalaby[1], Y. Al-Sarraj[1], R. Taha[1], M. Kambouris[2,3], H. El-Shanti[1,4].* 1) Medical Genetics Center, Qatar Biomedical Research Institute, Doha, Qatar; 2) Pathology-Genetics, Sidra Medical and Research Center, Doha, Qatar; 3) Genetics, Yale University School of Medicine, New Haven, CT, USA; 4) Pediatrics, University of Iowa, Iowa City, IA, USA.

Two boys with Autism Spectrum Disorder (ASD) from two unrelated consanguineous families of Arabic origin were studied by Whole Genome Sequencing (WGS) together with their parents. The WGS data of the X chromosome were analyzed to identify possible predisposing X-linked variants. Comparative analysis of the WGS data for X-linked recessive inheritance identified the following three strong candidate gene variations, in order of priority; Family 1: *IL1RAPL2* c.206G>C/p.S69T; Family 2: *SHROOM4* c.3370C>G/p.Q1124E, and *SYTL5* c.1370G>A/p.R457Q. All variations validated by Sanger Sequencing, co-segregate with the disease phenotype within each family and are absent in known polymorphism databases, as well as in 1800 ethnically matched control chromosomes, genotyped by TaqMan assays and Real-Time PCR.In Family 1: *IL1RAPL2* is associated with non-syndromic X-linked ID and/or ASD and the protein is detected at low levels in fetal and adult brain (particularly in the frontal lobe, temporal lobe and cerebellum). In Family 2: *SHROOM4* plays a role in cytoskeletal architecture and it is considered an XLMR gene, as mutations have been linked to Stocco dos Santos Syndrome. Deleterious mutations might affect the morphology of the neural cells and eventually the neural development. The *SYTL5* encoded protein belongs to the synaptogamin-like protein family and seems to play a role in protein transportation in specific tissues, as it is expression is restricted to placenta and liver. Maybe during embryonic development its expression is required for downstream gene control that plays a role in neural development.Genes function and suggested mechanisms, as well as the absence of the mutations from the ethnically matched control population and publically available databases, indicate that these novel rare variants identified are strong candidates for predisposition to the development of ASD. Future studies on transcriptome analysis and gene expression will enable to confirm the indications and the mechanisms might justify their involvement to ASD.

705T

Inherited missense variation in Pfam protein domains contributes to autism risk. *S. Fu[1,2], M. Fromer[2,3], A. Goldberg[1,2,3], J. Buxbaum[1,2,3,4,5,6].* 1) Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 2) Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 3) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 4) Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 5) Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 6) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA.

Autism spectrum disorder (ASD) is a neurodevelopmental disorder affecting 1% of the population, with a large fraction of risk conveyed by genetic variation. The genetic architecture of ASD involves the interplay of common and rare variation affecting hundreds of genes. Gene function is defined by the structure of the resulting protein, with protein domains defining key functional and structural elements of the protein. Previous studies by the Autism Sequencing Consortium (ASC) and others have demonstrated that (1) *de novo* and (2) inherited loss of function (LoF) mutations and (3) *de novo* probably damaging missense (mis3) mutations all play a role in ASD risk. Here we hypothesized that inherited missense mutations could also contribute to autism risk by disrupting specific protein domains. Published rare variants identified with whole exome sequencing (WES) by the ASC and in the Simon Simplex Collection (SSC) in more than 17,000 individuals from seventeen sample sources were annotated with Pfam, which is a database of conserved protein domain families that is widely used to annotate and classify proteins. We analyzed both case-control and family data, incorporating the latter by using cases and pseudo-controls. *De novo* variants were analyzed separately.We grouped missense mutations into three categories based on polyphen2 annotation: mis1 (benign), 2 (possibly damaging), and 3 (probably damaging). Case-control burden was estimated using odds ratios from corresponding 2x2 tables, and significance was assessed using permutation that controls for population structure. We found that, when considering case-control and case-pseudo-control data, mis3 variation as a whole did not show excess burden (P=0.31;OR=0.96). In contrast, mis3 variation disrupting Pfam domains show excess burden in cases (P=0.02;OR=1.02). No corresponding signal was observed in mis1 and mis2 variation (P>0.3). For *de novo* variation, strong burden was found in mis3 (P=0.015). However subsetting down to the mis3 variants overlapping pfam did not show a strong signal.We next assessed whether specific Pfam domains could be implicated in these analyses. Although nothing survived correction for multiple testing across 4178 unique Pfam domains, the top hit was the nucleosome assembly protein family, consistent with previous findings of alterations in chromatin remodeling in risk factor for ASD. Further studies will group domains functionally and will incorporate additional ASD whole exome datasets.

706F

Whole genome sequencing of pedigrees with a high burden of autoimmune disease. *P. Bronson[1], J. Sitrin[1], T. Bhangale[1], L. Criswell[2], P. Gregersen[3], R. Graham[1], T. Behrens[1].* 1) Human Genetics, Genentech, Inc., South San Francisco, CA; 2) Dept. of Rheumatology, Univ. of California, San Francisco, CA; 3) Boas Center for Genomics & Human Genetics, Feinstein Institute for Human Genetics, Manhasset, NY.

Collectively, autoimmune diseases affect >23M Americans, and can be severely debilitating. Though the etiology of autoimmunity remains largely unknown, evidence supports a substantial genetic component, particularly with HLA. To date, ~400 genetic loci have been reported. However, common risk alleles account for only a fraction of heritability, and the role of rare causal variants is unknown. We undertook a combined sequencing and linkage screen to identify potential rare variants. We selected 75 pedigrees with a high burden of autoimmune disease (≥2 cases with rheumatoid arthritis (RA), inflammatory bowel disease (IBD), type 1 diabetes (T1D), multiple sclerosis (MS) or systemic lupus erythematosus (SLE)) out of 265 multiplex autoimmune disease pedigrees collected by the Multiple Autoimmune Disease Genetics Consortium (MADGC). We sequenced one or two cases from each pedigree (73 WGS, 50 exome, total N = 123). Individuals in the 75 pedigrees with available DNA were genotyped on the Illumina 660K or 1M SNP chip (N = 669). After calling reads with GATK, we removed variants that were present in WGS data (mean 30x coverage) available for 583 European Americans, the 1000 Genomes Project, or the NHLBI Exome Sequencing Project. For families with two cases sequenced, variants that were only present in one case were removed. To prioritize rare variants, we first restricted our search to ~400 known autoimmune loci. We also undertook a nonparametric genome-wide linkage scan (N = 5,132 common SNPs) to identify regions of the genome that segregate with disease within families. We identified 850 linkage intervals, and are using these family-specific linkage intervals to prioritize variants for follow-up genotyping. A private exonic missense (R>K) mutation in the SH2B adaptor protein 3 (*SH2B3*, chr12:111,882,571**A*) was identified in both affecteds from a single pedigree. *SH2B3* is important for hematopoiesis and T cell receptor signaling. *SH2B3* is pleiotropic for autoimmunity, and is a significant GWAS locus for T1D, RA, Celiac disease, ankylosing spondylitis, primary biliary cirrhosis, juvenile idiopathic arthritis, vitiligo and primary sclerosing cholangitis. Further work is being conducted to genotype this variant in additional affected and unaffected members of this pedigree, and to identify additional rare variants.

707W

The exome sequencing identified the YARS2 mutation as a nuclear modifier for the phenotypic manifestation of Leber's hereditary optic neuropathy-associated mitochondrial DNA mutation. *M. Guan[1], P. Jiang[1], X. Jin[1], Y. Peng[1], M. Wang[1], X. Liu[2], X. Zhou[2], J. Qu[2], T. Huang[3].* 1) Genetics, Zhejiang University, Hangzhou, Zhejiang, China; 2) Ophthalmology, Wenzhou Medical University, Wenzhou, Zhejiang, China; 3) Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA.

Leber's hereditary optic neuropathy (LHON) is one of the most common inherited optic neuropathies, with maternal inheritance. The MT-ND4 m.11778G>A mutation is the most prevalent LHON-associated mitochondrial DNA mutation. The m.11778G>A mutation is a primary factor underlying the development of blindness but insufficient to produce a blindness phenotype. Nuclear modifier genes were proposed to modulate the phenotypic manifestation of the m.11778G>A mutation. However, none of modifier gene(s) has been identified. An exome sequencing of four members in a Chinese family carrying m.11778G>A mutation identified a missense mutation (highly conserved 191 glycine to valine) (p.191Gly>Val) at the catalytic domain of mitochondrial tyrosyl-tRNA synthetase, encoded by YARS2. The Sanger sequence analysis of YARS2 in 307 symptomatic matrilineal relatives and 271 asymptomatic matrilineal relatives of 167 Chinese families carrying the m.11778G>A mutation showed that all matrilineal relatives carrying both homozygous YARS2 p.191Gly>Val and m.11778G>A mutations exhibited optic neuropathy. It was proposed that the YARS2 p.191Gly>Val mutation alters its structure and function. Functional analysis showed that homozygous YARS2 p.191Gly>Val mutation altered the aminoacylation and steady-state levels of tRNATyr. The failure in tRNATyr metabolism impaired mitochondrial translation, especially for polypeptides with high content of tyrosine codon such as MT-ND4, MT-ND5, MT-ND6 and MT-COX2. The primary defect of m.11778G>A mutation was the reduced activity of complex I. However, the YARS2 p.191Gly>Val mutation worsened the respiratory phenotypes associated with m.11778G>A, especially reducing activities of complex I and IV. The respiratory deficiency then caused the reducing efficiency of the mitochondrial ATP synthesis and increasing the production of reactive oxygen species. Thus, biochemical defects caused by the YARS2 p.191Gly>Val mutation aggravate the mitochondrial dysfunction associated with the m.11778G>A mutation, exceeding the threshold for expressing the blindness phenotype. These findings indicate that the mutated YARS2, acting as a nuclear modifier, modulates the phenotypic manifestation of the LHON-associated mtDNA mutations.

708T

Exome analysis of rare and common variants within the NOD receptor pathway. *G. Andreoletti[1], E.G. Seaby[1], T. Coelho[1,2], N.A. Afzal[2], A. Batra[2], R. Haggarty[3], R.M. Beattie[2], S. Ennis[1].* 1) University of Southampton, Southampton, Hampshire, United Kingdom; 2) Southampton Children's Hospital, University Hospital Southampton NHS Foundation Trust, Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, UK; 3) NIHR Nutrition Biomedical Research Centre, Southampton Centre for Biomedical Research, University Hospital Southampton NHS Foundation Trust (Mailpoint 218), Southampton General Hospital, Southampton, SO16 6YD, UK.

Objective: This study looks at the burden of coding mutation in PIBD patients within the genes comprising the NOD2 pathway using data derived from exome sequencing.**Methods**: 143 cases and 96 controls were recruited through tertiary referral clinics at University Hospital Southampton for whole-exome sequencing. For all exomed patients, genomic DNA was extracted using the salting out method. Data were aligned and tested for quality control using an in-house pipeline and customized scripts. Burden of mutation testing was conducted across all genes in the NOD2 pathway using the SKAT-O test.**Results**: We observed mutations in 39 of 40 genes comprising the NOD2 pathway. Despite the small sample size, SKAT-O test identified five genes as significantly associated with disease status (*NFKB1* p= 0.008; *BIRC3* p=0.016; *IL6* p=0.022; *BIRC2* p=0.030, and *NOD2* p=0.033). The evidence contributing to the association signal for these genes is primarily driven by rare variants that would not have been assessed in previous GWA studies.**Conclusion**: Within a modest cohort of children with IBD we successfully applied burden of mutation testing that includes common and rare variants. Three of the implicated genes (*NFKB1*, *NOD2* and *IL6*) had been previously identified using only common variant association but two previously unreported genes (*BIRC2* and *BIRC3*) were also detected. *BIRC2* and *BIRC3* are members of the inhibitor of apoptosis familywhich plays an important role in IBD pathogenesis.

709F

Identification of novel disease-causing genes expands molecular genetic diagnosis of Osteogenesis Imperfecta. *S. Fahiminiya[1], F. Rauch[2], J. Majewski[1].* 1) Department of Human Genetics, Faculty of Medicine, McGill University and Genome Quebec Innovation Center, Montreal, Quebec, Canada; 2) Shriners Hospital for Children and McGill University, Montreal, Quebec, Canada.

Background: Sequencing of protein coding regions of human genome (Whole Exome Sequencing; WES) has been successfully used to unravel the mystery behind several inherited human genetic diseases, in the past few years. We applied this approach to uncover novel genetic cause of Osteogenesis Imperfecta (OI). OI is a group of genetically heterogeneous disorders characterized mainly by variable degree of bone fragility and most often is caused by mutations in COL1A1 or COL1A2, the genes that code for collagen type I alpha 1 and alpha 2 chains. Other clinical manifestations include tooth abnormalities (dentinogenesis imperfecta), hearing loss and blue or gray sclera. Here, we present the results of our WES studies that led to the identification of several novel disease-causing genes in OI patients. We also show that analyzing only rare variants in coding regions is not always conclusive and the involvement of non-coding variants and copy number variations (CNV) should not be ignored during bioinformatics analysis of WES. **Method**: WES was performed on 69 patients diagnosed with different types of OI in our center at the Shriners Hospital for Children in Montreal, for whom the result of initial molecular screening of known or potential candidate genes was negative. **Results**: We successfully identified novel disease-causing mutations in coding regions of SPARC and WNT1 in patients with clinical diagnosis of OI type IV. Additionally, we showed that variants in non-coding regions and CNV carry important information and should be considered during bioinformatics analysis of WES data. For instance, we identified pathogenic variants in 5'UTR of IFITM5 (OI type V) resulting in the production of longer protein or in 3'UTR of BMP1, affecting the polyadenylation signal and impaired procollagen type I C-propeptide cleavage. In a patient with OI type VII, CNV in CRTAP was identified as the main cause of the diseases. In addition to new gene discovery, we also identified mutations in Known OI genes (COL1A1, COL1A2, FKBP10 and SERPINF1), which highlight the importance of using WES as an efficient and sensitive diagnosis tools in OI patients. Finally, promising candidate genes were found in a few other patients, however validation in unrelated patients is needed. **Conclusion**: Our study highlights that WES is a powerful molecular approach for discovery pathogenic and novel gene(s) and will help to establish more specific diagnostic panel and improve clinical diagnosis of OI patients.

710W

Utility of VAAST in targeted case-control sequencing of candidate genes in childhood onset Crohn disease. *J. Chen[1], Y. Yu[1], S. Watkins[2], H. Hu[1], L. Denson[3], S. Kugathasan[4], C. Huff[1], S. Guthery[5].* 1) Epidemiology, MD Anderson Cancer Center, Houston, TX; 2) Department of Human Genetics, Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT; 3) Gastroenterology, Hepatology, and Nutrition, Cincinnati Children's Hospital Medical Center, and The University of Cincinnati College of Medicine, Cincinnati, Ohio; 4) Department of Pediatrics, Emory University School of Medicine and Children's Health Care of Atlanta, Atlanta, Georgia; 5) Department of Pediatrics, University of Utah School of Medicine, Salt Lake City, UT.

Crohn disease is a heritable disease characterized by chronic inflammation of the gastrointestinal tract. Although over 100 common susceptibility variants have been identified from genome-wide association studies (GWAS), the role of rare genetic variation in Crohn disease susceptibility is less clear. To search for rare variation contributing to Crohn disease risk, we sequenced 205 childhood onset parent-offspring trios and 194 unaffected controls in nine genes previously implicated in GWAS: *NOD2*, *IL23R*, *ATG16L1*, *IRGM*, *CARD9*, *STAT1*, *IRF1*, *IL5*, and *SLC22A4*. Our single marker analysis replicated two common missense susceptibility variants, in *ATG16L1* (T300A, rs2241880, p = 0.0006) and *NOD2* (P268S, rs2066842, p = 0.0019). The risk allele effect sizes were consistent with effect sizes reported in adult onset populations, with per allele odd ratios [OR] of 1.63 (95% confidence interval [CI] = 1.2 – 2.2) and 1.60 (95% CI = 1.2 – 2.2) for rs2241880 and rs2066842, respectively. To evaluate evidence for rare variant association, we performed gene-based tests using VAAST 2.1 for variants with population minor allele frequencies (MAF) lower than 0.05. VAAST identified a statistically significant association for rare nonsynonymous variants in *NOD2* (p = 2.1 x 10-5). The association signal at *NOD2* was driven primarily by three variants: G908R (rs2066845, MAF = 0.01, OR = 4.7, 95% CI 1.6 to 13.9), R702W (rs2066844, MAF = 0.05, OR = 2.0, 95% CI 1.2 to 3.4), and A1006fs (rs2066847, MAF = 0.025 OR =3.86 95% CI 1.9 to 7.7). Our results replicate the association of rare nonsynonymous variation in *NOD2,* affirm the utility of the probabilistic approach implemented in VAAST in complex genetic traits, and provide a framework for assessing rare variation in other genes implicated in Crohn disease.

711T

LONP1 is A Novel Gene for Congenital Diaphragmatic Hernia. *L. Yu[1], L. Shang[1], J. Wynn[1], AD. Sawle[2], M. Tarnopolsky[3], M. Geraghty[3], A. Larson[3], A. Safdar[3], A. Nazli[3], GB. Mychaliska[4], B. Warner[5], AJ. Wagner[6], RA. Cusick[7], FY. Lim[8], D. Potoka[9], K. Azarow[10], T. Crombleholme[11], D. Chung[12], B. O'Roak[13], Y. Shen[14], WK. Chung[1].* 1) Division of Molecular Genetics, Department of Pediatrics, Columbia University Medical Center, New York, NY 10032 USA; 2) The Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York NY 10032 USA; 3) Department of Pediatrics, McMaster University Medical Center, Hamilton, Ontario, CA; 4) Section of Pediatric Surgery, Fetal Diagnosis and Treatment Center, University of Michigan Health System, Ann Arbor, MI, USA; 5) Division of Pediatric Surgery, Washington University School of Medicine, St. Louis, MO, USA; 6) Division of Pediatric Surgery, Children's Hospital of Wisconsin, Medical College of Wisconsin, Milwaukee, WI, USA; 7) Division of Pediatric Surgery, University of Nebraska Medical Center College of Medicine, Omaha, NE, USA; 8) Division of Pediatric General, Thoracic, and Fetal Surgery, Center for Molecular Fetal Therapy, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 9) Department of Instruction and Learning - Early Intervention Program, University of Pittsburgh, Pittsburgh, PA, USA; 10) Pediatric Surgery Division, Department of Surgery, Oregon Health & Science University, Portland, OR 97239, USA; 11) Colorado Fetal Care Center, Division of Pediatric General, Thoracic, and Fetal Surgery, Children's Hospital Colorado and the University of Colorado School of Medicine, Denver, CO, USA; 12) Department of Pediatric Surgery, Monroe Carell Jr. Children's Hospital at Vanderbilt, Nashville, TN, USA; 13) University of Oregon, USA; 14) Departments of System Biology and Biomedical Informatics, Columbia University Medical Center, New York, NY 10032, USA.

Congenital diaphragmatic hernia (CDH) is a life-threatening birth defect characterized by the incomplete formation of diaphragm. The incidence of CDH is approximately 1 in 3,000 births. CDH can occur as an isolated defect or in the presence of one or more anomalies including congenital heart, brain, renal, and genitourinary malformations in 40% of CDH. Chromosomal anomalies and mutations in several genes have been implicated in the etiology of CDH. We identified a *de novo* variant (p.T638M) in *LONP1* in a sporadic CDH patient using parental-child whole exome sequencing (WES). We then studied an additional 424 CDH patients using targeted next generation sequencing and Sanger confirmation to identify an additional 3 rare variants in *LONP1* that were either *de novo* or dominantly inherited. The p.T638M variant was found recurrently, once *de novo* and once inherited in a familial case and segregated with CDH and eventration of the right hemidiaphragm. An additional *de novo* missense variant p.R358W was identified in a sporadic CDH patient. Variant p.S866F segregated with disease in a pedigree with incomplete penetrance with CDH and mild eventration of the diaphragm. We also identified a splice variant c.639-1G>T in a familial case that was *de novo* in the proband's paternal grandmother. Analysis of mRNA shows that the splice variant results in aberrant splicing that produces a premature termination. Functional studies in fibroblasts from patients revealed defects in cytochrome *c* oxidase (COX) enzyme activity and COX subunit 2 protein aggregates with immunohisotchemistry. Our results provide the first genetic evidence for the role of *LONP1* in diaphragm development and suggest a novel mechanism for the development of CDH. Our results also suggest that there are subtle defects in diaphragm development that are not clinically apparent that can be associated with genetic causes of CDH.

712F

A Patient with Beaulieu-Boycott-Innes Syndrome, Illustrating the Utility of Whole Exome Sequencing. *J. Amos[1], S. Tang[2], K. Helbig[2], W. Alcaraz[2], R.R. Lebel[1].* 1) SUNY Upstate Medical University, Syracuse , NY; 2) Ambry Genetics Corporation, Aliso Viejo, CA.

A 19-year-old female, initially identified with developmental delays and anatomic abnormalities at infancy, presented for re-consideration. Her features include cleft palate, corpus callosum dysplasia, developmental delays, bilateral hearing deficit, hypothelorism, lordosis, retrognathia, and ventriculomegaly; this constellation of features was considered "non-syndromic," not fitting any currently identified syndromes. Previously, routine karyotype and subtelomere FISH had revealed a normal variant 2q deletion. Comparative genomic hybridization later revealed a 163 kb deletion at 10q21.3, which was shown to be of maternal origin but has uncertain clinical significance. A literature search had been performed (using London Dysmorphology Database) to identify candidate syndromes, without success. DNA was extracted from lymphocytes and subjected to whole exome sequencing in search of diagnostic insights; five mutations were found. Two significant novel mutations were found in the *THOC6* gene. One was a missense mutation causing a protein change T250P and the other was a nonsense mutation of R87 (stop codon). Other notable mutations were in the genes *ATR, FANCD2* and *NIPBL*; while they might have clinical importance, they are not thought to have any significance to the patient's phenotype since each is heterozygous and the conditions with which they are associated are autosomal recessive. Mutations of *THOC6* have been associated with Beaulieu-Boycott-Innes Syndrome (BBIS), an autosomal recessive condition (OMIM #613680); all four previously published cases were homozygous for Q46R inherited in an extended pedigree characterized by consanguinity. Comparing our patient's phenotype with previously reported patients, we note some overlap in features including intellectual disability, microcephaly, and retrognathia. Other features in our patient, such as brain anomalies found by MRI, submucous cleft palate, short stature, hypothyroidism, bilateral hearing deficit, strabismus and low-set ears, were not observed in previously reported patients. Mutations found in this patient are novel and likely causative of her adverse phenotype. Her divergence from previously observed phenotypes may be due to the differential effects of the specific mutations. We are in the process of obtaining follow-up information about the previously reported patients, seeking insight for this patient's optimum management.

713W

Searching for rare variants involved in fulminant hepatitis B virus infection using exome sequencing. *N. Chaturvedi[1,2], S. Asgari[1,2], P. Scepanovic[1,2], J. Fellay[1,2].* 1) Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland.

Background: Hepatitis B virus (HBV) infection is a global health problem causing around 1 million deaths every year. A rare and dramatic clinical presentation, observed in less than 1% of patients with acute HBV infection, is fulminant hepatic failure. We used exome sequencing in carefully selected patients to test the hypothesis that rare human genetic variants play a role in driving the course of HBV disease towards fulminant hepatic failure.Methods: A total of 21 liver transplant recipients who developed liver failure due to fulminant HBV infection were recruited in the Transplantation Units of Swiss and Australian University Hospitals. Exome capture and sequencing were performed using Agilent SureSelect technology and Illumina HiSeq2000 sequencer. We used *BWA* to align the short sequencing reads to the human reference genome, and *GATK* for variant calling and quality control. We used two sets of exomes as controls: an in-house set of 172 samples from patients co-infected with HIV and HBV (without any history of severe HBV infection), and the large collection of ~30,000 non-Finnish European exomes gathered by the Exome Aggregation Consortium (ExAC). We used principal component analysis to identify and exclude population outliers. Then, we ran a case-control association analysis of loss-of-function (LoF) and missense variants, collapsed by gene, using one-tailed and two-tailed Fisher's exact tests. We applied the mid-P correction to the two-tailed Fisher's exact test, and Bonferroni correction to control for multiple testing.Results and discussion: After quality control, we observed a total of 228,077 variants in the 21 cases, including 2,611 LoF and 60,822 missense variants. We compared the allele count of LoF variants, per gene, in fulminant HBV cases and in in-house and ExAC controls. While no significant difference was found in the comparison with in-house controls, we identified a total of 100 genes that were enriched in LoF variants in the HBV cases in comparison to the ExAC controls. Using this set of genes, we performed protein interaction analysis, disease pathway analysis and functional annotation clustering using DAVID, but did not observe any significant enrichment. We also obtained RNA-seq data from blood samples of the same set of fulminant HBV patients, and we are performing integrated genomic analysis to study the associational patterns between DNA variants and RNA expression levels.

714T

Sharing of sequence variants in multiple genes within and across multiplex nonsyndromic cleft lip/palate families supports a multifactorial model. *S.H. Blanton[1,2], P. Dillingham[2], S. Slifer[2], R.E. Plant[3], M.E. Serna[3], B.T. Chiquet[4], E.C. Swindell[3], J.T. Hecht[3].* 1) Dept of Human Genetics, Univ Miami, Miami, FL; 2) Hussman Institute for Human Genomics, Univ Miami, Miami, FL; 3) Dept of Pediatrics, UTHSC, Houston, TX; 4) Dept of Pediatric Dentistry, UTHSC, Houston, TX.

Nonsyndromic cleft lip/palate (NSCLP) is a common birth defect, with a birth prevalence of approximately 1/1000. While there is strong evidence for genetic underpinnings based on family and twin studies and GWAS in large samples, only a few specific genetic changes have been identified. Association studies have identified several candidate pathways involving HOX and WNT genes. We previously identified association to a novel gene, CRISPLD2, and have shown that it plays a role in orofacial development in zebrafish. To further explore its role, we knocked down CRISPLD2 in zebrafish and used RNA-seq to identify up and down regulated genes. In order to better capture the variants/genes involved in the etiology of NSCLP, we subjected a subset of 36 families from our large collection of multiplex NSCLP families to whole exome sequencing with the Agilent 50Mb Exome v5 + UTRs kit. After initial QC, there were thirty-four families with at least 2 affected individuals sequenced (range 2-6). We compiled a candidate gene list of 62 genes identified from association studies and our RNA-seq analysis. As an initial screen, we filtered variants by presence in our candidate genes, presence in all affected individuals within a family, and having a GERP score>4, or location in a predicted transcription factor binding site (TFBS), or a possibly or probably damaging functional prediction by Polyphen2. A total of 247 variants in 50 genes met these criteria. Multiple families had variants in the same gene(s). Individual families had variants in multiple genes, ranging from 15 to 34 genes; the number of segregating variants was not correlated with the number of affected individuals. 207 variants were exonic, 168 were rare or novel, 156 were TFBS, and 27 were possibly or probably damaging. Of the 57 nonsynonymous variants, 1 was in the 5'UTR, 53 were missense, 2 were splice donors and 1 was a gain of a stop codon. Variants in multiple genes are segregating in individual families as well as across families. This is consistent with the multifactorial trait threshold model proposed for NSCLP. We are currently developing multigenic haplotypes in our families.

715F

Whole genome sequencing in families to find new Alzheimer Disease genes: The analysis and current results from the Family Analysis Working Group of the Alzheimer's Disease Sequencing Project. *E. Blue[1,4], B. Kunkle[2,4], B. Vardarajan[3,4], J. Jaworski[2], Family Analysis Working Group of the Alzheimer's Disease Sequencing Project.* 1) Div Med Gen, Univ. of Washington, Seattle, WA; 2) Inst Hum Gen, Univ. of Miami, Miami, FL; 3) Neurology, Columbia Univ., New York, NY; 4) co-first author.

The Alzheimer's Disease Sequencing Project (ADSP) is a Presidential Initiative on Alzheimer's Disease (AD) with the goals of 1) identifying new genetic variants that influence risk for developing AD, 2) determining why some individuals with known risk factors for AD escape from developing the disease, and ultimately 3) identifying new pathways for the prevention of AD. Toward this goal, the Family-based Analysis work group of the ADSP is analyzing whole genome sequences generated as part of this initiative on 583 subjects from 111 families that contain multiple affected relatives with late-onset AD, and pre-existing SNP array data. These families were ascertained from multiple sites, and include 67 Caribbean Hispanic and 44 European-American and Dutch families. These data are publicly available through dbGaP.Families with multiple affected relatives can provide powerful information regarding the genetic underpinnings of AD. In particular, co-segregation of variants with the phenotype provides sharp focus to narrow the vast quantity of variation within whole genomes. Our previous linkage analyses of these families identified several regions of interest (ROI) with HLOD>4, including: 14q32.2 in European-Americans, 7p14.3 in Hispanics, and 11q12-14, a region previously associated with LOAD, in both ethnicities. Additionally, significant evidence for linkage was observed at 3q22 in European-American, Dutch, and Hispanic families. These results define our initial target ROIs for examination using the ADSP sequence results.Several complementary analyses of the sequencing data to find variants segregating with disease in these ROIs are underway, including: 1) linkage analyses using the full pedigree structures with and without accounting for local admixture, 2) prioritizing rare variants that co-segregate using IBD/linkage regions and functional annotation, 3) combined linkage and association analysis, 4) analyses of variants in known AD and dementia genes, and 5) gene-based analyses. Products of these analyses will be shared with the research community. These include 1) estimates of regional and global ancestry proportions in the Hispanic families, 2) inheritance vectors for identity-by-descent-based analyses computed from the complete pedigrees, and 3) variant imputation throughout the pedigrees. This presentation is meant to inform the research community of the resources available, and highlight our preliminary findings.

716W

Common and Rare Exonic Variants Associated with Type 2 Diabetes in Han Chinese. *G. Chen[1], Z. Zhang[2], A. Adeyemo[1], Y. Zhou[2], A. Doumatey[1], G. Liu[2], J. Zhou[1], D. Shriner[1], F. Tekola-Ayele[1], C. Wang[2], A. Bentley[1], C. Jiang[2], C. Rotimi[1].* 1) NHGRI/NIH, National InGuanjie Chenstitutes of Health, Bethesda, MD; 2) Suizhou Central Hospital, Suizhou, Hubei, China.

Type 2 diabetes (T2D) is a common health condition with a prevalence of over 10% in China and about a million new cases of T2D are projected to occur each year in the 21st century. Genome wide association studies (GWAS) performed in East Asians have identified several common variants associated with T2D. Although these findings have provided valuable insight into the pathogenesis of T2D, much of the heritability for this trait remains unexplained. It has been postulated that rare variants may explain some of this missing heritability. Here, we investigate the contribution of rare and common variants to susceptibility to T2D by analyzing exome array data in 1908 Han Chinese (917 cases) genotyped with Affymetrix Axiom® Exome Genotyping Arrays. A total of 57,704 autosomal SNPs within 18,455 genes passed quality control filters and were analyzed. Based on the joint common and rare variants analysis implemented in the Sequence Kernel Association Tests (SKAT), we identified significant associations between T2D and variants in *MUC5B* (p-value 1.01 × 10-14, 9 rare and 16 common variants). *MUC5B* encodes a member of the mucin family of proteins, which are highly glycosylated macromolecular components of mucus secretions. Emerging evidence indicates that diabetes is an independent risk factor for pulmonary dysfunction and that diabetes pulmonary fibrosis is a potential diabetic complication. Also, we replicated 8 (*NCNQ1, PEPD, ACHE, IGF2BP2, GPSM1, PCNXL2, ZPLD1, PSMD6*) of 46 genes previously associated with T2D in East Asian ancestry populations (GWAS catalog) using both rare and common variants with p value < 0.05. These findings provide potential insight into the role of rare and common variants in the pathogenesis of T2D in Chinese and provide an important evidence for understanding the mechanisms underlying diabetic pulmonary fibrosis.

717T

Genetics of normal cognitive and neurosensory variation in humans. *E.T. Cirulli[1], A.J. Bartholomew[1], D.B. Goldstein[2].* 1) Center for Applied Genomics and Precision Medicine, Duke Univ, Durham, NC; 2) Institute for Genomic Medicine, Columbia Univ, New York, NY.

While extensive efforts continue to be made to understand the genetics of human diseases, there has been much less serious effort dedicated to understanding the genetic bases of traits comprising everyday differences among healthy individuals, many of which are of great interest biologically and have the potential to further our understanding of diseases. We have administered a standard cognitive battery to more than 2,400 healthy volunteers and assessed more than 800 participants for their night visual acuity, face recognition ability, time perception, contagious yawn susceptibility, and auditory pitch discrimination; these are all traits that are relevant to disease, fairly easy to measure and involve neuronal processes. In addition to GWAS and exome chip genotyping, we have now sequenced the complete exomes of >100 of our participants. Our work provides a novel viewpoint into the biology underlying human diseases and the genetic architecture of human traits.

718F

Rare variation association testing in inflammatory bowel disease, using low coverage whole genome sequences. *K.M. de Lange, Y. Luo, C.A. Anderson, J.C. Barrett, UK IBD Genetics Consortium.* Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

The contribution of rare variants (MAF<1%) to complex disease has been difficult to study as they are poorly tagged by GWAS. Next generation sequencing studies can address this, but sample sizes have been modest, and mostly focused on the exome. Here we investigate the use of low coverage whole genome sequences to study rare variation in inflammatory bowel disease (IBD), an exemplar complex trait.Low coverage sequencing allows greater sample sizes (and increased power) for a fixed-cost, but yields less accurate individual genomes. Variation in read depth exists even in samples sequenced together and is exacerbated when combining existing datasets. Our IBD study has 3 groups with different mean depths: 1767 ulcerative colitis cases (UC; 2x), 2513 Crohn's disease cases (CD; 4x) and 3652 controls (6x).As read depth increases, sensitivity and specificity improves. If depth is correlated with phenotype two countervailing sources of error affect association tests: false positive calls (favors lower depth) and false negative calls (favors higher depth). When aggregating rare variants over a test region even tiny biases produce inflated test statistics. For common variation this bias can be reduced by joint-calling across large samples, and by using imputation improvement to borrow information from across individuals and correlated sites. However, our investigation suggests that for MAF<3% bias from read depth differences is not removed by joint-calling, as there is insufficient population-level information. Furthermore we discover that the bias in this range is worse after imputation improvement, which performs excessive smoothing on signals from sites with scarce information, making individual genotype probabilities poorly calibrated.We apply an approach based on Derkach et al's score test (2014), using expected genotypes given the data, and evaluating significance by bootstrapped permutation testing. Burden analyses are performed on rare variants in gene-coding regions and a signal is observed in the CD risk gene *NOD2* without inflation in other genes. The method is extended to non-coding regions (e.g active gene enhancers in immune-related cell types), where we find that the number of variants that may be included is limited by the corresponding increase in the aggregated bias. Analyses are ongoing, but it is clear that higher coverage sequencing of much larger sample sizes will be required to fully understand the role of rare variation in complex disease risk.

719W

Genetic epidemiology of ocular health and disease in the Jirel ethnic group of eastern Nepal: the Jiri Eye Study (JES). *M.P. Johnson[1], S.S. Thapa[2], K.L. Anderson[3], S. Laston[1], M.K. Shrestha[2], B. Towne[4], J. Subedi[5], J. Blangero[1], S. Williams-Blangero[1].* 1) South Texas Diabetes and Obesity Institute, School of Medicine, University of Texas Rio Grande Valley, Brownsville, TX; 2) Tilganga Institute of Ophthalmology, Kathmandu, Nepal; 3) Department of Ophthalmology, School of Medicine, University of Texas Health Science Center at San Antonio, TX; 4) Department of Community Health, Boonshoft School of Medicine, Wright State University, Kettering, OH; 5) Department of Sociology and Gerontology, College of Arts and Science, Miami University, Oxford, OH.

Approximately 90% (n=660 million) of the world's visually impaired reside in developing countries. Combating ocular-related health problems in these countries is critical if we are to reduce the global burden associated with visual impairment (VI). The primary objective of a new study established in eastern Nepal and focused on the Jirel ethnic group is to identify rare functional variants that influence VI. Nepal is a developing country where the prevalence of VI across regions ranges from 7% to 42%. Our family-based study design will recruit 2,000 members of the Jirel population to undergo an eye examination to document, in part, the prevalence of ocular diseases known to influence VI (e.g., cataract). We will use SOLAR to determine the underlying genetic architecture (heritability, pleiotropy) of all measured ocular-related traits and disease endpoints. An existing high-density SNP framework will be used to localize QTLs harboring ocular positional candidate genes. An exome sequencing strategy will be employed to objectively prioritize rare functional variants influencing ocular QTLs. An independent, population-based cohort (n=2,200) from the Bhaktapur District of Nepal will be used to replicate findings. The Jirels belong to a single extended pedigree containing >62,000 pair-wise relationships that are informative for genetic analysis. The Jirel pedigree has 80% power to detect an additive genetic heritability as low as 6.5%, a genetic correlation between two traits as low as 4.4%, and a genetic association explaining as little as 2.4% of the variation observed in a tested trait. The 316 founders of the Jirel pedigree could maximally transmit 10 to 112 copies of a private variant. The average probability that one of the Jirel founders disseminates 5 extra copies of a private variant ranges from 0.20 to 0.92. To date, 229 Jirel members have been recruited to the study. The mean (range) age, SBP and DBP are 45.7 (13-85) years, 129.9 (90-200) mmHg, and 81.2 (52-110) mmHg, respectively. Preliminary field diagnoses are: presbyopia, 28.8%; cataract, 18.3%; glaucoma, 7.9%; and hypertensive retinopathy, 5.7%. The Jirel population is ideal for identifying functional rare variants due to its genetic isolation and deep genealogical relationships. Results from our well-powered study design will explicitly implicate genes in causal biological pathways influencing VI trait variance and therefore, encourage testing of these genes in other global populations.

720T

***GRHL3* mutation screening identifies truncating mutations in patients with apparently nonsyndromic cleft palate.** *E. Mangold[1], A.C. Böhmer[1,2], P. Gültepe[1,2], H. Schuenke[1,2], J. Klamt[1,2], M.M. Nöthen[1,2], M. Knapp[3], K.U. Ludwig[1,2].* 1) Institute of Human Genetics, University of Bonn, Bonn, Germany; 2) Department of Genomics; University of Bonn, Bonn, Germany; 3) Institute of Medical Biometry and Informatics and Epidemiology; University of Bonn, Bonn, Germany.

Van der Woude syndrome (VWS) is a common syndromic form of cleft lip and palate. Mutations in the interferon regulatory factor 6 (*IRF6*) gene account for ~70% of cases. Recently, grainy head-like 3 (*GRHL3*) was identified as a novel VWS gene. Individuals with a *GRHL3* mutation are more likely to have cleft palate and less likely cleft lip or lip pits than individuals with an *IRF6* mutation.Of note, VWS patients have lip pits as sole symptom in addition to clefting. These can be subtle and difficult to recognize. Thus VWS might sometimes appear as nonsyndromic clefting. We therefore evaluated apparently nonsyndromic cleft patients for mutations in *GRHL3*. We sequenced the complete coding region of *GRHL3* in 576 patients with nonsyndromic cleft lip with or without cleft palate (nsCL/P) and 96 patients with nonsyndromic cleft palate only (nsCPO). Observed rare variants were investigated with respect to *de novo* occurrence. Nineteen rare single-base variants were found, four of them can be regarded as pathogenic: 1 frameshift mutation (present in an affected mother and her child), 3 splice site mutations (two of them *de novo* in the affected families, one inherited from an unaffected mother to her two affected children). All seven mutation carriers had an apparently nonsyndromic cleft palate only (nsCPO). We re-contacted/re-visited them, and again they denied any VWS typical lip pits.Notably, common variants in the regulatory region of *IRF6* have been shown to confer risk for nsCL/P, which has a genetically complex background. We therefore investigated whether also common variants in the *GRHL3* region are associated with nsCL/P. Based on genome-wide imputed data from our previously published meta-analyses of GWAS data we tested ~1300 SNPs in the *GRHL3* coding region +/- 200kb for an association in patients. None of the association *P*-values in that region withstood correction for multiple testing (lowest *P*-value observed: rs10903069, *P*=0.013), suggesting that single common variants at the *GRHL3* locus do not contribute to nsCL/P. To identify whether common variants within the *GRHL3* gene region (coding region, introns, UTR) show evidence for association in aggregate, we performed a gene-based analysis for nsCL/P using VEGAS. No evidence for gene-based association was identified.In summary, we did not find evidence for common *GRHL3* variants being causative for nsCL/P. However, we demonstrated that rare mutations in the coding region of *GRHL3* can mimic nsCPO. .

721F

Quality Control (QC) and Multi-Pipeline Genotype Consensus Calling Strategies for 578 whole genomes and 10,692 whole exomes in the Alzheimer's Disease Sequencing Project (ADSP). *A.C. Naj[1], H. Lin[2], B.N. Vardarajan[3], J. Dupuis[4], M. Schmidt[5], F. Sun[4], Y. Ma[6], A. Partch[7], W.J. Salerno[8], N. Gupta[9], S.H. Choi[4], D. Lancour[6], J.C. Bis[10], E.R. Martin[5], A.L. DeStefano[4], Alzheimer's Disease Sequencing Project (ADSP).* 1) Department of Biostatistics & Epidemiology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 2) Department of Medicine, Boston University School of Medicine, Boston, MA; 3) Department of Neurology, Columbia University Medical Center, New York, NY; 4) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 5) John P. Hussman Institute for Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 6) Department of Biomedical Genetics, Boston University School of Medicine, Boston, MA; 7) Department of Pathology, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 8) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 9) Medical and Population Genetics Program, Broad Institute, Cambridge, MA; 10) Department of Medicine, University of Washington, Seattle, WA.

To identify AD risk and protective variants the ADSP has sequenced whole genomes (WGS) of 578 samples from 111 multiplex families and whole exomes (WES) of 10,692 case/control samples at Baylor College of Medicine, Washington University-St. Louis, and the Broad Institute. WGS and WES from the 3 centers were centrally called using *GATK-HaplotypeCaller* at Broad and *Atlas V2* at Baylor. The ADSP QC Working Group applied a QC protocol to WGS VCFs from each pipeline and developed a novel consensus protocol to combine calls from both pipelines into a single high-quality genotype set.After initial sequencing-center-recommended filtering (variant-level filtering for *GATK*; genotype- and variant-level filtering for *Atlas*), additional variant-level metrics (e.g., call rate, average read depth) were computed across all subjects for each autosomal bi-allelic single nucleotide variant (SNV), and *a priori* filters were applied to exclude low-quality SNVs. Sample-level QC examined within-subject distributions of metrics (e.g., average read depth, Ti/Tv ratio), and outliers (by center or ethnic group) were noted, including correcting miscoded ethnicity for 5 pedigree members. The resulting QCed calls from each pipeline were then combined via a consensus protocol to integrate (a) genotypes concordant between pipelines and (b) high quality genotypes unique to one caller. For the consensus protocol, quality metrics from SNVs involved in Mendelian inconsistencies (MI) among parent-child pairs (n=67) were used to establish additional genotype-specific filters for *GATK* calls.Variant-level QC applied to WGS SNVs excluded ~12.0% and 14.5% of SNVs called by *GATK* and *Atlas*, respectively. After QC, ~91% of genotypes were concordant between pipelines, 4.23% of genotypes were exclusive to Atlas, 4.56% of genotypes were exclusive to *GATK*, and ~0.01% of genotypes were discordant between pipelines. Discordant genotypes were excluded. *GATK*-unique genotypes were kept if the genotype quality (GQ) score met specified thresholds. *Atlas*-unique genotypes were uniformly kept. Re-evaluation of MI in the final consensus set showed improvement in genotype consistency. The final WGS consensus data contains 27,896,774 bi-allelic autosomal SNVs. QC is ongoing for indels and WES.This protocol resulted in a single set of high-quality genotypes from 3 sequencing centers and 2 calling pipelines providing a robust pipeline applicable to other multi-center large-scale sequencing studies.

722W

Genome-wide analysis of stuttering susceptability. *L.E. Petty[1], S.J. Kraft[2], S. Shankaracharya[3], Y. Yu[3], J.M. Beilby[4], D.M. Munzy[5], S.N. Jhangiani[5], M.L. Grove[1], J.R. Lupski[6], R.A. Gibbs[5], E. Boerwinkle[1], C.D. Huff[3], J.E. Below[1].* 1) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 2) Communication Sciences and Disorders, Wayne State University, Detroit, MI; 3) Department of Epidemiology, MD Anderson Cancer Center, Houston, TX; 4) School of Psychology and Speech Pathology, Curtin University, Bentley, WA, Australia; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 6) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Stuttering is a speech and language disorder that is common in children and has a high estimated heritability, between 0.66 and 0.81. Prevalence in children is 5%, with higher regional reports, *e.g.* estimated prevalence in Australia is 11%. Existing research on the genetics of stuttering has been limited in both the number of studies and the size and context of those studies. To date, genes identified through candidate gene studies and linkage analyses have not replicated in the literature, suggesting the genetic architecture of stuttering remains largely undiscovered. We sampled 174 individuals comprising 95 families highly enriched for stuttering from Western Australia. From these families, all samples were chip genotyped on the Illumina HumanCore BeadChip, and two affected individuals from each of 25 families were whole exome sequenced in the Baylor Human Genome Sequencing Center through the Baylor Hopkins Center for Mendelian Genomics. Same chip data for additional population-based race-matched controls were drawn from the Childhood Autism Risks from Genetics and the Environment study, including genotypes for 48 HapMap CEU.Preliminary gene-based test analyses were performed using Pedigree-based Variant Annotation, Analysis, and Search Tool. Top results of chip and sequencing data include *CC2D2A* (p = 3.5 x 10-4), a gene related to formation of cilia, *DFFA* (p = 3.7 x 10-4), which has been linked to problems in apoptosis, *CYC1* (p = 2.4 x 10-3), which is involved in the mitochondrial respiratory chain, *DEFB128* (p = 4.1 x 10-4), and *ZNF467* (p = 3.5 x 10-3).To date, no GWAS have been published for stuttering susceptibility. Preliminary results of single variant GWAS controlled for relatedness identified a top signal, rs2516399 (p = 1 x 10-7), which was previously associated with eosinophil count and is intergenic between immune response genes *MICB* and *MCCD1*. Suggestive SNPs include rs10817858 (p = 2.4 x 10-5), rs3737787 (p = 5.9 x 10-5) located in the 3' untranslated region of *USF1*, which has prior associations with familial hyperlipidemia and Alzheimer disease, and rs948962 (p = 8.5 x 10-5) which is non-synonymous and exonic to *MYO7A*, a gene in which homozygous mutation causes Usher syndrome (deafness and diminished inner ear function).These top variants lend support to the idea that stuttering is associated with a pathway that causes structural brain changes. They also indicate new avenues of inquiry, including damage to hearing and the vestibular system.

723T

Genome-wide rare variant association study of extreme phenotypes of emphysema. *J.E. Radder, Y. Zhang, F.C. Sciurba, S.D. Shapiro.* Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA.

Chronic obstructive pulmonary disease (COPD) is characterized by irreversible airflow limitation due to repeated exposures to noxious inhalants, most commonly cigarette smoke. To date, loci identified in genome-wide association studies (GWAS) have only explained a fraction of the heritability of this disease. We hypothesized that a portion of the unexplained heritability of COPD is attributable to rare variants and that we could detect some of this variation using extreme phenotypes of a well-defined subtype of COPD. To test this, we conducted a whole genome sequencing study of extreme traits of emphysema as measured by density histogram quantitative computed tomography and tested for association of rare variants with this phenotype.We selected 70 subjects with severe emphysema and obstruction (f-950 = 0.00455 ± 0.00006, FEV1%predicted = 97.9% ± 2.6%) and 70 subjects with no indication of emphysema or obstruction (f-950 = 0.27600 ± 0.03080, FEV1%predicted = 25.4% ± 1.6%) of similar age (years, emphy = 61.1 ± 1.4, controls = 63.7 ± 1.2) and with similar smoking histories (pack years, emphy= 53.5 ± 7.4, controls=48.6 ± 5.9) from the SCCOR cohort at the University of Pittsburgh. These 140 subjects underwent whole genome sequencing and 137 were analyzed following quality control. Alignment was performed using bwa and variant calling according to GATK's best practices. Genome-wide depth of coverage averaged 27.6X across these samples. We identified 15,819,411 SNPs and indels, 3,532,222 of which were novel (compared to dbSNP138). We filtered all SNPs with minor allele frequency (MAF) > 0.01 in 1000 Genomes and 6500 exomes European American populations. Any SNP with no MAF recorded in one of these databases was excluded with an empirical MAF > 0.01, leaving 5,423,566 rare SNPs (2,221,186 not in dbSNP138). We performed principal component analysis to adjust for ancestry on a pruned set of linkage independent SNPs using EIGENSOFT. We tested for association using SKAT-O in 30kb regions across the genome with age, sex, pack years, and the first two principal components of the ancestry analysis as covariates.The top genomic region associated with emphysema in our study was chr3:81840000-81869999 (P = 2.5x10-6). We are further validating this finding and investigating the genomic elements in this region for their potential functional effects on emphysema. .

724F

Comparing genetic, proteomic and clinical profiles of discordant monozygotic twins. *N. Vadgama[1,2], M. Simpson[6], N. Niranjanan[7], D. Gaze[7], K. Pearce[3], M. Kristiansen[3], P. De Rijk[8], E. Rees[5], G. Kirov[5], A. Pittman[1], S. Morgan[1], D. Lamont[4], J. Hardy[1], J. Nasir[2].* 1) Instituite of Neurology, University College London, London, United Kingdom; 2) Division of Biomedical Sciences, St. George's University of London, London, UK; 3) Instituite of Child Health, University College London, London, UK; 4) College of Life Sciences, University of Dundee, Dundee, UK; 5) Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, UK; 6) Department of Genetics and Molecular Medicine, King's College London, London, UK; 7) St. George's Hospital and Medical School, London, UK; 8) Applied Molecular Genomics Group, University of Antwerp, Antwerp, Belgium.

Discordant monozygotic (MZ) twins provide a unique opportunity to study the genetics of complex disorders. We carried out whole-exome sequencing, copy number variation (CNV) and proteomic analysis on 13 pairs of MZ twins discordant for a range of complex disorders, including dystonia, inclusion body myositis, and lactose non-persistence. Sequence reads that passed the quality control threshold were filtered for repetitive DNA and screened against a variety of databases, including our in-house database with over 1000 exome sequences. We next manually screened all SNPs that were either different or showed >15% mosaicism.One twin pair discordant for amyotrophic lateral sclerosis (ALS) was screened against a panel of 25 ALS-linked genes. The panel was designed using Illumina TruSeq Custom Amplicon and implemented on an Illumina MiSeq platform. A mutation in *SOD1* (c.65A>G; p.E22G) was identified in both the patients, suggesting epigenetic and/or environmental involvement in familial ALS.Data from Illumina HumanCore BeadChip was analysed using Illumina Genomestudio and PennCNV software, involving an assessment of calls in each of the twin pairs in order to identify unique CNVs in either the affected and unaffected twin. Novel CNV differences were identified within MZ twin pairs discordant for attention deficit hyperactivity disorder, Tourette's syndrome, schizophrenia, and inclusion body myositis.Proteomic profiling of serum samples obtained from an MZ twin pair discordant for ischemic stroke was analysed through a label-free pipeline using the Progenesis application. Potential quantitative and qualitative biomarkers were identified, including afamin and desmoglein-1. We performed pathway analysis using Cytoscape and PANTHER on those proteins, which showed statistically significant differences between the twin pair.We have identified clear-cut quantitative and qualitative differences at the genetic and proteomic levels between MZ twins. This adds to our recently published study on a comparison of blood chemistries of the MZ twins discordant for stroke, showing marked elevation of γ-glutamyltransferase (GGT) and erythrocyte sedimentation rate (ESR) levels in the affected twin.These novel findings support our original hypothesis of postzygotic *de novo* mutations during development to explain discordance in MZ twins. Our study also presents a theoretical framework for mosaicism detection and indicates potential weaknesses of existing variant detection tools.

725W

Sequencing Analysis of Interferon Lambda loci In Individuals with Spontaneous Hepatitis C Virus Clearance and Persistence. *C. Vergara[1], C. Thio[1], M. Taub[2], R. Latanich[1], G. Kirk[1], S. Mehta[1], A. Cox[1], D. Thomas[1], P. Duggal[2].* 1) Johns Hopkins University, School of Medicine, Baltimore, Maryland, USA; 2) Johns Hopkins University, Bloomberg School of Public Health, Baltimore, Maryland, USA.

Hepatitis C virus (HCV) infections affect ~170 million people and is one of the leading causes of cirrhosis and liver cancer. The infection either spontaneously is cleared or persists; persistence is more common in persons of African descent. Polymorphisms near the genes for interferon lambda (IFNL) are significantly associated with HCV persistence and enriched in persons of African descent but the causal DNA sequence and biological mechanisms are unknown. We aimed to determine the distribution of functionally annotated variants (SNVs, insertions, deletions) in and around the IFNL genes in African Americans (AA) and European Americans (EA) with HCV clearance and persistence. Whole genome sequencing was performed in 19 AA (11 clearance) and 18 EA (7 clearance) at 30X coverage using the Illumina HiSeq platform. Variants were annotated and predicted using Snpeff. We interrogated a 1Mb region on chromosome 19 (chr19:39002893-40006156) that includes the IFNL genes and tabulated the distribution of annotated variants between clearance and persistence groups.142 variants in 31 genes were observed in AA and 122 variants in EA (Table 1). Only missense variants were annotated in the IFNL genes. The total number of variants was higher in the persistence group compared with the clearance group for EA (71 vs 51) and AA (74 vs 68). In EA, the Non-synonymous coding variants were more frequent in the persistence group compared with the clearance group (70 vs. 40) although this was not distinct in the AA (69 vs. 67) Most of these variants were predicted to have moderate impact. Two high impact variants were observed in the EA clearance group compared to 1 in the EA persistence group. In AA, 1 variant of high impact was identified in the clearance group, and 4 variants in the persistence group. A higher number of annotated variants (mainly non synonymous coding) were present in individuals with HCV persistence independent of ethnic group. Overall, we observed a diversity of variants with functional impact in individuals with both HCV clearance and persistence. Further analysis will help to elucidate the role of these variants in the resolution of HCV infection.

Table 1. Annotated variants in IFNL with HCV clearance and persistence.

Annotation	European Americans			African Americans		
	Clearance	Persistence	Total	Clearance	Persistence	Total
Non Synonymous Coding	49	70	119	67	69	136
Stop Gained	0	1	1	0	0	0
Splice Site Donor	1	0	1	1	3	4
Codon Deletion	0	0	0	0	1	1
Start Lost	1	0	1	0	1	1
Total	51	71	122	68	74	142

726T

Systematic screening of known associated genes confirms Congenital Hypogonadotropic Hypogonadism is characterized by oligogenicity. *D. Cassatella[1,3], J. Liang[2], J.S. Acierno[3], C. Xu[3], S. Santini[3], A. Dwyer[3], C. Chambion[3], G. Sykiotis[3], M. Hauschild[3], F. Phan-Hug[3], J.G. Xhang[3], B.J. Stevenson[4], N. Pitteloud[3].* 1) Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland; 2) BGI-Shenzen, Shenzen, PRC; 3) Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland; 4) Swiss Institute of Bioinformatics, Lausanne, Switzerland.

Congenital Hypogonadotropic Hypogonadism (CHH [MIM 308700]) is characterized by incomplete puberty and infertility due to absent GnRH secretion. In 50% of cases, CHH is accompanied by absent sense of smell (anosmia) and termed Kallmann syndrome (KS); otherwise, it is referred to as normosmic CHH (nCHH). CHH is a rare, genetically heterogeneous disorder, with 26 genes known to underlie the phenotype, either alone or in combination.CHH shows incomplete penetrance and variable expressivity, both within and between families. Previous studies by our group and others have demonstrated oligogenicity (>2 genes mutated) in CHH, however smaller subsets of genes were evaluated. This study aims to comprehensively evaluate oligogenicity in a larger set of genes, using whole exome sequencing in a cohort of CHH probands.We evaluated 22 known CHH genes in 106 CHH probands and 416 controls from the CoLaus cohort to identify rare, pathogenic non-synonymous and splice acceptor/donor variants. Overall, 45.3% of CHH probands harbored at least one rare pathogenic variant—a frequency higher than reported in previous studies (22-35%).The frequency of digenicty (including bi-allelic changes in one or two genes) in CHH probands (8.5%) was statistically higher relative to controls (0%), which only harbored monoallelic rare variants (RVs) in known CHH genes (p=4.4x10-7); previous studies demonstrated digenicity in the range of 2.8-4.6%.Finally, 4.7% of CHH probands exhibited oligogenicity (RVs in >2 CHH genes), which was absent in controls (p=3.2x10-4). Oligogenicity in previous studies was markedly less, at approximate 1% on average.On a single-gene level, we found two genes with higher mutational rates in our cohort than previously reported; *FGFR1* RVs are present in the 15.1% of our CHH population, while its incidence has been previously reported at 6-10%, and *CHD7* had an incidence of 10.4%, higher than shown in previous studies (5-8%).In summary, as more genes are evaluated in CHH patients, the more striking the frequency of oligogenicity becomes. The complete oligogenic profile for CHH patients is critical for understanding the pathogenicity of this disease and for improving genetic counseling, and next-generation sequencing provides a cost-effective means to accomplish this task.

727F

Exome sequencing identifies *SLC36A4* and *SLC4A8* as novel genes influencing the distribution of HDL-cholesterol in West Africans and African Americans. *A.R. Bentley[1], D. Shriner[1], F. Tekola Ayele[1], A.P. Doumatey[1], J.C. Mullikin[2], . NISC Comparative Sequencing Pg[2], C. Adebamowo[3], J. Oli[4], O. Fasanmade[5], T. Johnson[5], A. Amoah[6], K. Agyenim-Boateng[7], B.A. Eghan Jr.[7], A.A. Adeyemo[1], C.N. Rotimi[1].* 1) Center for Research on Genomics and Global Health, NHGRI/NIH, Bethesda, MD; 2) NIH Intramural Sequencing Center, Bethesda, MD; 3) Department of Epidemiology and Public Health, School of Medicine, University of Maryland, Baltimore, MD; 4) Department of Medicine, University of Nigeria Teaching Hospital, Enugu, Nigeria; 5) Department of Medicine, University of Lagos, Lagos, Nigeria; 6) Department of Medicine and Therapeutics, University of Ghana Medical School, Accra, Ghana; 7) Department of Medicine, University of Science and Technology, Kumasi, Ghana.

Much of the genetic epidemiology of high-density lipoprotein cholesterol (HDLC) in individuals of African ancestry has focused on replicating findings identified in those of European ancestry, an approach which may miss African ancestry-specific variation. Also, HDLC is influenced by both genetic and non-genetic factors, motivating study of individuals with similar ancestry living in different environments, such as West Africans (WA) and African Americans (AA). We conducted whole exome sequencing on individuals at the extremes (5th and 95th percentiles) of the HDLC distribution of samples of WA (n=98 [low], n=100 [high]) and AA (n=57 [low], n=60 [high]). Consistent with previous observations, the mean HDLC was lower in WA than AA (low: 19 vs. 27 mg/dl; high: 75 vs. 96 mg/dl). Common variants (MAF > 0.05) were analyzed in regression models of risk of being in the low group after adjustment for age, sex, and the first principal component. No variants reached statistical significance in the analysis of either WA or AA. We conducted gene-based analysis using SKAT with covariates as described above and statistical significance set at p=2.3 × 10-6 (α=0.05/21,711 genes). Among the WA, *SLC36A4* was associated with HDLC (p=1.6 × 10-7). Variants in this gene were not associated with HDLC among AA, but variants in a different solute carrier, *SLC4A8*, were (p=3.9 × 10-7). Adjusting for body mass index reduced the statistical significance of the *SLC4A8* association in AA (p=1.0 × 10-3), but had virtually no effect on the *SLC36A4* association in WA. To our knowledge, this is the first report of the contribution of either *SLC4A8* or *SLC36A4* to HDLC distribution. The different findings among WA and AA despite ~80% genome-wide shared ancestry are of interest. A potential explanation is that the variants interrogated were not identical in the two groups (e.g., only 77% of the *SLC36A4* and 52% of the *SLC4A8* variants were observed in both WA and AA samples). Alternatively, the differences may result from the European admixture in AA or genetic effects that manifest differently depending on the environment (e.g. the obesogenic environment of AA). This work highlights the importance of studying the genetic determinants of HDLC in African ancestry populations, as it may reveal novel mechanisms that contribute to the distribution of this trait.

728W

Intraocular pressure and exome sequencing variation in the Beaver Dam Eye Study. *F. Chen[1], AP. KLEIN[1,2,3], KE. LEE[4], R. KLEIN[4], BEK. KLEIN[4], P. DUGGAL[1].* 1) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 2) Department of Oncology, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, MD; 3) Department of Pathology, Johns Hopkins School of Medicine, Baltimore, MD; 4) Department of Ophthalmology and Visual Sciences, University of Wisconsin School of Medicine and Public Health, Madison, WI.

Intraocular pressure (IOP) is an important clinical parameter in the evaluation of ocular health. Elevated IOP is a major risk factor for primary open-angle glaucoma (POAG). This study aimed to identify novel gene candidates influencing IOP. A cohort of 139 individuals of European descent with extreme baseline IOP or refractive error measurements from the Beaver Dam Eye Study (BDES) underwent whole exome sequencing. Variants were called simultaneously across all samples by GATK Unified Genotyper and were filtered based on a number of quality control metrics, resulting in a total of 228,413 single nucleotide variants (SNVs) with a mean transition transversion ratio (Ti/Tv) of 2.62 and 11,500 small insertions/deletions (INDELs). In our preliminary analysis for IOP, we classified 15 unrelated individuals as “high IOP” with an IOP > 22 mmHg in at least one eye at baseline. We filtered all variants compared to the normal or low IOP individuals, and identified 3,467 SNVs and 48 small INDELs in exons that are unique to these high IOP individuals. We generated a list of genes with multiple exonic variants that were seen in more than one individual with elevated IOP excluding known false positive genes. The top two genes were *ADGRV1* (adhesion G protein-coupled receptor V1) on chromosome 5q13 (4 nonsynonymous and 2 synonymous SNVs) and *ZFHX3* (zinc finger homeobox 3) on chromosome 16q22.3 (2 non-synonymous and 4 synonymous SNVs). All identified SNVs of these two genes were singletons and 10 of 15 individuals with high IOP carried at least one SNV including four with glaucoma diagnoses. SNVs of these two genes in our study cohort each have a minor allele frequency (MAF) consistent with or higher than those in Europeans from the 1000 Genomes Project. The functional role of these two genes in the eye are unknown, however mutations in human *ADGRV1* are linked to Usher Syndrome type 2, and its homologous gene is associated with abnormal vision in mice. And *ZFHX3* is highly expressed in the retina. We will extend our analysis to other gene regions having multiple variants for low and high IOP as well as the known susceptibility loci for IOP and glaucoma from both GWAS and linkage studies.

729T

Whole-exome sequencing identifies mutations in *SQSTM1* and *VCP* genes in a series of 205 inclusion body myositis cases. *Q. Gang[1,2], C. Bettencourt[1], P.M. Machado[1,2], S. Brady[2,4], J.L. Holton[2], A.M. Pittman[1,3], D. Hughes[1], E. Healy[2], M. Parton[2], D. Hilton-Jones[4], P.B. Shieh[5], M. Needham[6], C. Liang[7], E. Zanoteli[8], L. Valente de Camargo[8], B. De Paepe[9], J. De Bleecker[9], A. Shaibani[10], M. Ripolone[11], R. Violano[11], M. Moggio[11], R.J. Barohn[12], M.M. Dimachkie[12], M. Mora[13], R. Mantegazza[13], S. Zanotti[13], A.B. Singleton[14], M.G. Hanna[1,2], H. Houlden[1,2], The Muscle Study Group and The International IBM Genetics Consortium.* 1) Department of Molecular Neuroscience, Institute of Neurology, University College London, London, United Kingdom; 2) MRC Centre for Neuromuscular Diseases, Institute of Neurology, University College London. London, United Kingdom; 3) Reta Lila Weston Institute, Institute of Neurology, University College London. London, United Kingdom; 4) Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, Oxford, United Kingdom; 5) Department of Neurology, University of California Los Angeles, United States; 6) Western Australian Neurosciences Research Institute (WANRI), University of Western Australia and Murdoch University, Fiona Stanley Hospital, Perth, Australia; 7) Royal North Shore Hospital, New South Wales, Australia; 8) Department of Neurology, Medical School of the University of São Paulo (FMUSP), São Paulo, Brazil; 9) Ghent University Hospital, Department of Neurology & Neuromuscular Reference Centre, Ghent, Belgium; 10) Nerve and Muscle Center of Texas, Houston, Texas, United States; 11) Neuromuscular Unit, IRCCS Foundation Ca' Granda Ospedale Maggiore Policlinico, Dino Ferrari Centre, University of Milan, Milan, Italy; 12) The University of Kansas Medical Centre, Kansas City KS, United States; 13) Neuromuscular Diseases and Neuroimmunology Unit, Fondazione IRCCS Isitituto Neurologico C. Besta, Milano, Italy; 14) Laboratory of Neurogenetics, National Institute on Aging, National Institute of Health, Bethesda, Maryland, United States.

Objective: Sporadic inclusion body myositis (sIBM [MIM 147421]) is the most common myopathy in individuals aged over 45 years. The etiology of sIBM is unknown, but clinico-pathologically overlapping inherited disorders indicate that genetic factors might be involved.**Methods:** We used whole-exome sequencing in a large group of 205 IBM patients. Muscle tissue was pathologically evaluated and whole-transcriptome expression profiles generated.**Results:** We identified eight rare missense mutations in the *SQSTM1* (MIM 601530) and *VCP* (MIM 601023) genes in 10 IBM patients (5%). Five of the mutations have been previously reported in patients with other degenerative diseases, typically associated with denervation, as in amyotrophic lateral sclerosis (ALS [MIM 105400]) and frontotemporal dementia (FTD [MIM 600274]) with Paget's disease of bone (PDB [MIM 602080]). These patients exhibited increased p62 immunohistochemical staining in muscle tissue. MHC-I markers were up-regulated in the muscle of patients as compared to controls, both at the protein and mRNA level.**Conclusion:** In the past the occurrence of denervation found on muscle electromyography (EMG) and the deposition of neurodegeneration associated proteins in sIBM patients was always considered an unexpected finding. The occurrence of mutations in *SQSTM1* and *VCP* in IBM, ALS, FTD and PDB reinforces the link between these disorders, pinpointing converging pathogenic pathways resulting in impaired autophagy-lysosome processing, causing dysregulation of protein homeostasis. The MHC expression could indicate the differentiating factor that directs the clinical phenotype towards IBM as opposed to the other neurodegenerative conditions. We suggest that *SQSTM1* and *VCP* need further investigation in IBM and collaborative efforts will undoubtedly lead to a greater understanding of the pathogenesis of this disorder.

730F

Identification of novel candidate genes in Primary Congenital Glaucoma by Whole Exome Sequencing. *M. Kabra[1], AK. Mandal[2], S. Senthil[2], I. Kaur[1], S. Chakrabarti[1].* 1) Brien Holden Eye Research Centre, L.V. Prasad Eye Institute, Hyderabad, India; 2) Jasti V Ramanamma Children's Eye Care Centre, L.V. Prasad Eye Institute, Hyderabad, India.

PurposePrimary congenital glaucoma (PCG) is an autosomal recessive disorder of the eye characterized by abnormal development of the trabecular meshwork and anterior chamber angle. Mutations in the *CYP1B1* gene have been implicated in 20-100% of all PCG cases worldwide. In the present Southern Indian cohort, *CYP1B1* accounted for 40.47% cases exhibiting 27.47% homozygous and 13.00% heterozygous mutations. Thus, a large majority of PCG cases were either partially or completely unlinked to *CYP1B1* suggesting the involvement of other candidate genes in the disease pathogenesis. The present study was designed to identify the other novel genes by whole exome sequencing (WES) in cases devoid of mutations in the known candidate genes (*CYP1B1, LTBP2, MYOC* and *FOXC1*).**Methods**Ten trios consisting of affected children born out of consanguineous marriages and their unaffected parents underwent WES on an Ion Proton platform (Life Tech, CA). The preparation of DNA libraries, enrichment and sequencing (mean depth of 100x) was done as per the manufacturer's guidelines along with the standard quality control measures and the filtered reads were aligned to hg19 reference genome. An algorithm was designed to delineate the homozygous variants followed by filtering the dbSNP and 1000 genomes variants. Non-synonymous deleterious variants assessed by SIFT, PolyPhen and Grantham were finally analyzed for disease segregation in the trios.**Results**WES data revealed 50,000-90,000 variants in each trio of which on1y 1500-6000 variants were homozygous in the probands. Subsequent analysis indicated 396 novel variants that were not available in the public database and only 28 variants in 27 genes were found pathogenic. These variants were not reported in the HapMap database. Further prioritization of these genes was based on their interactions (STRING software) and published literature.**Conclusions**Among the novel candidate genes, four harbored the known glaucomatous loci *(GLC1I, GLC1L*, *GLC1N* and *14q21-q22*). Five developmental genes included two transcription factors, one of which was earlier implicated in anterior segment anomalies in animal model and the other was involved in the development of the neural crest by interacting with *PAX6*. One gene variant was common across two trios, which is known to induce autophagy in RGCs following IOP elevation leading to neuro-degeneration. Further functional validation of these genes understanding their precise role in PCG is in progress.

731W

Understanding genetics of glucose and insulin dynamics in Mexican American individuals from Starr County, Texas. *M. Karhade[1], G. Jun[1], L.E. Petty[1], G. Bell[2], J.E. Below[1], C.L. Hanis[1].* 1) Human Genetics Center, The University of Texas Health Sciences Center, Houston, TX; 2) Department of Human Genetics and Department of Medicine, The University of Chicago , IL.

Oral glucose tolerance test (OGTT) is a procedure for diagnosis of impaired glucose tolerance (IGT) and type 2 diabetes (T2D). Genetic association studies typically assess fasting and 2-hour post-load glucose and insulin levels from OGTT, often ignoring in-between time points. To understand genetic components affecting glucose and insulin responses, we analyzed glucose, insulin, and C-peptide levels measured at 7 time points (0, 10, 20, 30, 60, 90 and 120 minutes) after oral glucose intake on 331 previously undiagnosed Mexican-American individuals from Starr County, Texas. For genetic analysis, we collected both whole exome sequencing data (Agilent Truseq capture reagents on Illumina HiSeq2000) and GWAS array data (Genome-Wide Human SNP Array 6.0 Affymetrix) for all 331 samples. In our preliminary analysis, we analyzed glucose, insulin, and C-peptide levels at baseline, and then extended our analysis to the differences in measurement levels between adjacent time points. In exome-wide single-variant association tests (MAC>5), no SNP reached genome-wide significance level of 5x10-8. Interestingly, top signals for glucose at rs17053612 (*DOCK5p*-value 4.6x10-7) and insulin at rs8178822 (*APOHp*-value 1.89x10-6) traits were both for changes between 10 and 20 minutes, which corresponds the to the effects of the first phase of insulin release.*APOH*&*DOCK5*have been previously implicated in T2D and in severe obesity respectively. The top signal rs112475035 (*ANKRD13Cp*-value 2.28x10-7) for C-peptide was for the change between 90-120 minutes. To assess contributions of rare variants, we performed gene-based tests using SKAT-O for variants with MAF<0.05. Given a gene-wise significance threshold of 2.5x10-6, one gene, *SPTBN5*was significantly associated with insulin changes between 0 and 10 minutes with*p*-value of 2.24x10-6.*SPTBN5*has been previously implicated in ischemic stroke. Glucose at baseline was found to be associated with the gene*OBSCN*with*p*-value 5.52x10-6. These intermediate time points provide novel insight into the biology of glucose tolerance.

732T

Whole-exome sequencing identifies novel candidate genes for early-onset Alzheimer disease. *B.W. Kunkle[1], B. Vardarajan[2], P.L. Whitehead[1], S. Rolati[1], E.R. Martin[1], J.R. Gilbert[1], R.P. Mayeux[1], J.L. Haines[3], M.A. Pericak-Vance[1], G.W. Beecham[1].* 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Taub Institute of Research on Alzheimer's Disease, Columbia University, New York, NY, USA; 3) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH, USA.

BACKGROUND. Mutations in*APP*,*PSEN1*and*PSEN2*lead to early-onset Alzheimer disease (EOAD). These mutations account for ~11% of EOAD overall, leaving the majority of genetic risk for the most severe form of Alzheimer disease unexplained. METHODS. To search for rare variants contributing to risk for EOAD we performed Whole-Exome Sequencing (WES) in 50 Caucasian EOAD cases screened negative for*APP*,*PSEN1*, and*PSEN2*. Variant filtering for functional, damaging rare variants (MAF<0.1%) was performed. Damage prediction was performed using a Combined Annotation Dependent Depletion (CADD) score. Rare, damaging variants shared by multiple cases (+2) were then selected for follow-up protein-protein interaction analysis with known EOAD genes (*APP, PSEN1, PSEN2, GRN*,*MAPT*) using the program STRINGdb. Additionally, we performed a separate text-mining screen of this set of variants using Phenolyzer and the terms 'Alzheimer disease', 'dementia', and 'neurodegeneration'. RESULTS. We identified 51 rare, damaging variants in 46 genes in two or more EOAD cases. Gene network analysis of these 46 genes with known EOAD genes identified three candidate genes: HSPG2 (interacts with GRN and APP), CLSTN1 (interacts with PSEN1 and APP), and DOCK3 (interacts with PSEN1 and PSEN2). Five cases have a variant in HSPG2, a gene in a LOAD susceptibility region and potentially involved in amyloidogenesis and tau aggregation in AD. Four cases have a variant in DOCK3, a gene shown to regulate amyloid-İ secretion, and associated with neurofibrillary tangles in AD brains. Several other cases have shared variants in CLSTN1. Disruption of calsyntenin-1-associated axonal transport of APP by mutations in CLSTN1, a known APP interactor, have been identified as a potential pathogenic mechanism of Alzheimer's. Moreover, CLSTN1's potential as a regulator of synapse formation and neuronal development suggests other mechanisms through which it could be involved in development of dementia. Top scoring genes for the Phenolyzer analysis included the genes HSPG2, STAT1, a BACE1 interactor, and CNTNAP2, whose expression is down-regulated in AD patients. CONCLUSIONS. WES of EOAD cases identified several genes with potential roles in AD pathogenesis.

733F

Exome sequencing for identification of potential causal variants for diffuse cutaneous systemic sclerosis. *A.C.Y. Mak[1], L.F. Tang[2], C. Cleveland[3,5], M.K. Connolly[4], T. Katsumoto[3,5], P. Wolters[5], P.Y. Kwok[1,2,4], L.A. Criswell[3].* 1) Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA; 2) Institute for Human Genetics, University of California, San Francisco, San Francisco, CA; 3) Rosalind Russell Medical Research Center for Arthritis, University of California, San Francisco, San Francisco, CA; 4) Department of Dermatology, University of California, San Francisco, San Francisco, CA; 5) Department of Medicine, University of California, San Francisco, San Francisco, CA.

Scleroderma is a genetically complex autoimmune disease with substantial phenotypic heterogeneity. Previous genome-wide association studies (GWAS) have identified a large number of gene regions associated with disease risk. However, GWAS directly capture only common genetic variants that are presumably in linkage disequilibrium with causal variants, some of which may be rare variants more frequently found in the disease population. Our goal was to identify potentially causal variants by comparing whole exome sequencing data between cases and controls. We focused on patients with severe disease, and specifically those with diffuse cutaneous systemic sclerosis (dcSSc), to limit disease heterogeneity.We produced whole exome sequencing data from 32 dcSSc patients on the Illumina HiSeq2500 platform with the Nimblegen SeqCap EZ v3.0 exome enrichment protocol. Paired-end 2 X 100 bp sequencing reads were generated with a mean coverage of 51X on the 64Mb of targeted exome regions. Control data came from 17 healthy control subjects whose data were produced using comparable methods, and from the 1000 Genomes and the NHLBI-ESP6500 Projects. Exome sequencing reads were processed according to the GATK Best Practices for DNAseq variant analysis. Variants were annotated and filtered with ANNOVAR and Variant Tools. We applied a gene mutation burden test to identify genes that were enriched with deleterious variants in dcSSc patients compared to controls.We identified 96 genes that were enriched with deleterious variants in the dcSSc patients. Among the 96 genes, 10 genes (*NOTCH4, BANK1, BLK, GRB10, IRF8, KCNA5, NMNAT2 and TNFSF4*) are in previously identified scleroderma susceptibility loci or pathways implicated in scleroderma pathogenesis. In addition, we identified novel genes and 3 new pathways associated with scleroderma, including pathways related to ABC-family proteins mediated transport, extracellular matrix organization and the CD320-dependent methylmalonic aciduria.Using exome sequencing and gene mutation burden analysis, we identified 96 genes that contain functionally deleterious variants that may contribute to the development of dcSSc. This study demonstrates the potential value of whole exome sequencing for the identification of causal variants that contribute to scleroderma risk and/or severity. The candidate genes we discovered are potential targets for in-depth functional study or therapeutics development.

734W

Whole exome sequencing of probands from 158 idiopathic pulmonary fibrosis families identifies a telomere related gene network carrying potentially damaging mutations. *R. Mitra[1], Q. Wang[1], J.D. Cogan[2], J.A. Kropski[3], D.B. Mitchell[3], H. Yu[1], J.E. Loyd[3], Z. Zhao[1], J.A. Phillips III[2,4], T.S. Blackwell[3,5,6,7].* 1) Department of Biomedical Informatics; 2) Division of Medical Genetics and Genomic Medicine, Department of Pediatrics; 3) Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine; 4) Department of Pathology, Microbiology and Immunology; 5) Department of Cancer Biology; 6) Department of Cell and Developmental Biology, Vanderbilt University School of Medicine, Nashville, TN; 7) Department of Veterans Affairs Medical Center, Nashville, TN.

Idiopathic pulmonary fibrosis (IPF) is a progressive and often fatal lung disease of unknown etiology. So far, familial forms of IPF have been linked to genes involved in regulation of telomere length (e.g. *TERT*, *TERC*, and *RTEL1*) and those encoding surfactant proteins. Interestingly, even in the absence of telomerase mutations, short telomeres are frequently found in IPF patients and many of them lack a family history. To investigate the mechanisms of developing IPF, we examined exome-sequencing data from 158 unrelated familial IPF subjects. Exomes from 4300 European subjects from the NHLBI GO Exome Sequencing Project (ESP) database were used as a control population. To identify the sequence variants contributing to IPF risk, we analyzed rare (MAF < 0.01 in ESP) or intermediate (MAF < 0.05 in ESP) frequency variants that had a significantly higher minor allele frequency (MAF) in IPF compared to control data ($P < 0.05$, Fisher's exact test). Based on these criteria, we identified 200 potentially damaging variants that mapped to 214 genes. We selected 9 of these predicted damaging variants for Sanger resequencing, which successfully validated each variant. Enrichment analysis of these 214 genes using Pathway Commons prioritized 'mTOR signaling pathway' as the top-ranked pathway mutated in IPF. The relevance of this pathway in IPF development has been explained by previous studies. Assuming that mutated proteins may alter the function of their interactors, we constructed an IPF-specific network, which started with the mutated gene *ATM* linking to mTOR signaling pathway and then added curated protein interacting pairs. A subset of the ATM interactors was found to be linked to 'Telomere maintenance', 'Regulation of Telomerase', and 'mTOR signaling pathway', which formed a network we termed as backbone network. We further extended the backbone network based on the literature evidence such as known DNA binding protein interactions and obtained a network of 34 genes. Intriguingly, in this systematically derived network, 14/34 (41%) genes carried damaging variants and were observed in 59% of the affected IPF subjects, indicating that genetic variation related to telomere biology may underlie a large proportion of IPF risk. Overall, our network approach is promising to detect rare or intermediate damaging variants from exome-sequencing data with their potential roles in IPF.

735T

Whole exome sequencing analysis of HIV-1 African American elite controllers. *M. Montasser[1], J. O'Connell[1], M. Sajadi[2,3,4], A. Parsa[1].* 1) Department of Medicine, Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD, USA; 2) Institute of Human Virology, University of Maryland School of Medicine, Baltimore, Maryland, USA; 3) Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA; 4) Medical Care Clinical Center, VA Maryland Health Care Center, Baltimore, Maryland, USA.

HIV-1 continues to plague the world with 35 million infected individuals, 25 million of them in sub-Saharan Africa. While highly active antiretroviral therapy is very effective in limiting HIV-1 progression, it is limited by drug resistance/toxicity, the high cost of lifetime treatment, and it's ineffectiveness in eliminating the latent HIV-1 reservoir. Currently, there is a consensus that host genetics, rather than viral factors, plays a major role in determining the response to HIV-1 infection. Among populations of African ancestry, the MHC region explains ~ 25% of the variability in response to HIV-1 infection, while the remaining variability is still unknown. HIV-1 elite controllers (EC) represent the 1% of the HIV-1 infected population that have the ability to naturally suppress their viral infection for many years without treatment, and therefor may be instrumental in identifying protective genetic factors. We performed exome sequencing in 22 EC and 2 pools of 20 HIV-1 progressors. After QC, filtering, statistical testing, and biological prioritization, we identified 15 SNPs located in 13 genes with interesting biological links to HIV or the immune system. These SNPs have a MAF of 2% or less in the 1000 genomes African population, but are increased to 10-15% in our EC. Five SNPs were missense, 4 silent, 5 intronic, and one non-genic. The genes are: *CPSF3, PRG4, PSMD1, SETDB2, TAF4B, TPR, MPDZ, C1orf150, KIR-REL2, IGSF10, PIAS1, POLR3A,* and *DEFB110*. In an attempt to replicate these findings, we performed TaqMan genotyping for 11 SNPs in 42 EC (including the original samples to confirm the exome data), and 42 HIV-1 progressors. While we were able to confirm the exome data in the same samples with TaqMan, we identified the minor allele in no more than 1-2 additional replication samples. These findings did not refute nor convincingly confirm our initial findings, but rather showed that identifying the genetic basis of response to HIV infection is very challenging. The difficulty in identifying coding variants contributing to EC may indicate that there are many very rare coding causal variants, thus, requiring very large sample sizes to sufficiently power discovery. It may also indicate that the causal variants are non-coding regulatory variants and cannot be detected by exome sequencing. Multiomic studies that can layer data from transcriptomic, epigenomic, metabolomic, and genomic data may be warranted to elucidate the genetic components of EC status.

736F

Identification of novel candidate genes for tuberculosis susceptibility by identifying disease-causing mutations in individuals with PIDs. *M. Müller[1], N. Schlechter[1], E.G. Hoal[1], M. Schoeman[1], M. Urban[1], M. Esser[2], B.S. Petersen[3], A. Franke[3], C.J. Kinnear[1].* 1) SA MRC Centre for TB Research, DST/NRF Centre of Excellence for Biomedical Tuberculosis Research, Division of Molecular Biology and Human Genetic, Faculty of Medicine and Health Science, Stellenbosch University, Cape Town, Western Cape, South Africa; 2) Department of Pathology, Stellenbosch University, Cape Town, South Africa; 3) Institute for Clinical Molecular Biology, Christian-Albrechts University Kiel, Kiel, Germany.

Despite the fact that approximately 33% of the world's population is infected with *M. tuberculosis,* the causative agent of tuberculosis (TB), only 10% of infected individuals will develop active disease. While the genotype of invading strain and environmental factors have been shown to be crucial in disease outcome, host genetic factors are just as important. Several investigations have been undertaken to identify genes involved in TB susceptibility and many TB susceptibility genes have been reported. Nevertheless, it is certain that more susceptibility genes exist, but identifying these genes in a complex disease such as TB is challenging. We believe that the answers may lie in the genomes of individuals suffering from a group of inherited primary immunodeficiency disorders (PIDs) for which multiple infections with *M. tuberculosis* is a common feature. We hypothesise that the genes that cause these PIDs could be plausible candidate genes for increased TB susceptibility in the general population. We therefore aimed to identify novel TB susceptibility genes by finding gene mutations in patients suffering from PIDs characterised by increased TB susceptibility.Six PID patients between the ages of eight and 17 years were recruited for this study. Once written informed consent was obtained, their exomes were sequenced using the Illumina HiSeq. Bioinformatics techniques were used to identify variations from the reference human genome for each of the patients.Bioinformatics analysis of the sequence data identified a large amount of potential candidate genes in these patients. We subsequently prioritised the genes based on OMIM and HGMD database entries which resulted in 11 genes. At least one variant in the TAP1 gene has already been identified for which functional studies will be done in order to investigate its involvement in the disease. .

737W

Exome sequencing reveals *GNE*, *DYSF* and *CAPN3* as the major contributors of muscular dystrophy with potential roles of other genes in mediating disease progression. *B. Nallamilli[1], K. Gandhi[1], S. khadilkar[2], A. Bhutada[2], M. Hegde[1].* 1) Human Genetics, Emory University, Atlanta, GA; 2) Bombay Hospital institute of Medical Sciences Mumbai, India.

Muscular dystrophies (MDs) constitute a group of genetically and clinically heterogeneous diseases of progressive muscle weakness. Without prior protein analysis on muscle biopsies, and most frequently even after initial protein analysis, determining the disease subtype and the causative gene is difficult. To date, over 54 genes have been shown to be associated with MDs. A majority of these genes are expressed at the sarcolemma to form an integrative network serving to provide structural support to individual muscle cell. The clinical and genetic heterogeneities of MDs make disease diagnosis complicated and expensive. However, current next generation sequencing based approaches have tremendously improved our molecular diagnostic capabilities, allowing timely management and participation of affected individuals in clinical trials that have become increasingly available. Comprehensive exome sequencing analysis facilitates identification of disease subtype and causative genes in one single test. In this project, we performed exome sequencing (ES) on 100 clinically well characterized MD patients to identify the causative gene and disease subtype. Among these, molecular diagnosis could be established in 70 patients with the majority having mutations in either of three genes, *GNE*, *DYSF* and *CAPN3*, suggesting that these are the major contributors to MDs. More interestingly, in at least 50 individuals, mutations were detected in more than one MD genes. For example, one patient clinically diagnosed with LGMD2B but with an unexpected severe progression had mutations in both *DYSF* and *MYH2* genes. Heterozygous loss of MYH2 coupled with complete loss of DYSF could be contributing to the rapid disease progression and severe phenotype observed in this individual. Similar multigenic combinations of mutations were detected in *DYSF, CAPN3, SYNE2, GNE* and *LAMA2* in other individuals suggesting a possible role of synergestic heterozygosity and digenic contribution to disease presentation and progression. Using ES we have recently identified new disease causative genes including *DDOST*, *NGLY1* and *PIGL* and are currently investigating several new candidate genes, with functional studies. Application of ES to MD diagnosis has improved our understanding of the clinical spectrum of different MDs. Overall this project allowed us to understand the phenotypic overlap between different MDs.

738T

Exome sequencing analysis in a Primary Open Angle Glaucoma Brazilian Family. *P. V. Svidnicki[1], J. P. C. Vasconcellos[2], H. D. T. Caldeira[3], M. B. Oliveira[1], H. F. Nunes[1], B. S. Carvalho[4], M. G. Borges[4], P. A. O. R. Aguiar Araújo[4], I. T. L. Cendes[4], M. B. Melo[1].* 1) Center of Molecular Biology and Genetic Engineering, UNICAMP, Campinas, SP, Brazil; 2) Department of Ophthalmology, Faculty of Medical Sciences, UNICAMP, Campinas, SP, Brazil; 3) Centro Oftalmológico de Janaúba, Janaúba, MG, Brazil; 4) Department of Medical Genetics, Faculty of Medical Sciences, UNICAMP, Campinas, SP, Brazil.

Background: Glaucoma is the leading cause of irreversible blindness worldwide, characterized by loss of retinal ganglion cells leading to visual field deficits. Multiple loci have been associated with primary open angle glaucoma (POAG), the most common type of the disease. However, the etiology and pathogenesis of this trait are poorly understood. **Methods**: We have studied a Brazilian POAG family (four members, the father and three siblings) with no mutation in the MYOC gene, through whole exome sequencing (WES) using Illumina Nextera Rapid Capture Expanded Exome kit. Sequencing data were generated on Illumina HiSeq2500 at mean coverage depth of 70X. Reads were aligned to the human reference genome (GRCh38) using Burrows-Wheeler Aligner. Variants were called using GATK and annotations were performed using VEP. Filtering strategies were: variants segregating only in the affected individuals (father and one of the siblings), excluding common and synonymous variants. After filtering, 363 variants remained. Subsequently, variant prediction by SIFT and Polyphen-2, gene expression in ocular tissue databases, and finally text mining in NCBI and Pubmed were evaluated. Mutation segregation in candidate genes according to the phenotype was further investigated by Sanger sequencing in 13 other family members (7 affected and 6 non-affected individuals). **Results**: Affected members tested negative for the most common known genes in glaucoma, for instance, OPTN, NTF4, TBK1, WDR36, CAV1/CAV2, CDKN2B, TMCO1, SIX1/SIX6. According to our filtering criteria, we selected five variants in the following genes: ANKS6, CPAMD8, LTBP2, NUMA1 and UTRN. However, none of these variants segregated with the disease. Five other variants in 5 genes have been chosen and are being evaluated according to their segregation in family members. **Conclusion**: We selected possible new candidate genes for POAG, since patients do not carry any mutation in the known causative genes for glaucoma. Further resequencing and experimental verification are still needed to confirm the causative gene in this family. Supported by Fundação de Amparo á Pesquisa do Estado de São Paulo - FAPESP. Process: 2013/17958-2.

739F

Exome sequencing in families affected with age-related macular degeneration identifies rare variants in Complement Factor H associated with extramacular drusen and lower serum factor H levels. *E. Wagner[1,2], M. Triebwasser[3], E. Schramm[3], M. Daly[4,5], S. Raychaudhuri[4,5], J. Atkinson[3], J. Seddon[1,2].* 1) Ophthalmic Epidemiology and Genetics Service, Tufts Medical Center, Boston, MA; 2) Department of Ophthalmology, Tufts University School of Medicine, Sackler School of Graduate Medical Sciences, Tufts University, Boston, MA; 3) Washington University School of Medicine, Department of Internal Medicine, Division of Rheumatology, St. Louis, MO; 4) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA; 5) Partners HealthCare Center for Personalized Genetic Medicine, Division of Genetics, Brigham and Women's Hospital, Boston, MA.

Age-related macular degeneration (AMD) is the leading cause of blindness in people over the age of 65. It is a common disease with multifactorial etiology including environmental, behavioral, and genetic components. In this study, we investigated the genetic factors contributing to AMD risk in families that are densely affected with advanced AMD, have an earlier age at diagnosis (less than 75 years of age) and have a low genetic risk score as calculated using the known genetic factors. We performed exome sequencing on members of families who met the above criteria and applied a series of filtering parameters to screen for rare, damaging variants in genes known to be associated with AMD from genome-wide association study results and which segregate with AMD in the families. Our results showed that in two independent families, a novel, loss-of-function variant in the AMD gene, complement factor H, segregated with the disease. Both variants are predicted to be deleterious by prediction programs; one confers an amino acid change while the other abolishes a splice site, and both mutations occur in a highly conserved region of the gene that interacts with C3b, an important component of the innate immune system. Neither rare variant was found in public genomic databases, nor were they found in any of the other subjects in our AMD registry database. To assess the functional impact of these variants, factor H protein (FH) serum levels were quantified using radial immunodiffusion with anti-human antibodies specific for FH. Family members carrying the rare variants had decreased serum FH levels compared to those without the variants. Additionally, a detailed assessment of the retinal phenotypes revealed presence of extramacular drusen among family members carrying the rare variants but not among those without the variants. These results contribute to the catalog of rare, highly penetrant variants in known AMD genes and aid in expanding our understanding of the genetic architecture of this disease.

740W

Exome sequencing of a family–based cohort with polycystic ovary syndrome highlights variants in known and novel genes. *S. G. Wilson[1,2,3], V. Oo[1, 4], S. Liew[1,4], P. J. Campbell[1], B. H. Mullin[2], A. L. Weeks[2], S. J. Brown[1], R. J. Mead[4], E. M. Lim[1,5,6], B. G. A. Stuckey[1,2].* 1) Endocrinology & Diabetes, Sir Charles Gairdner Hospital, Nedlands, WA, Australia; 2) School of Medicine and Pharmacology, University of Western Australia, Nedlands, WA, Australia; 3) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 4) School of Biological Sciences, Murdoch University, Murdoch, WA, Australia; 5) PathWest Laboratory Medicine WA, Nedlands, WA, Australia; 6) School of Pathology and Laboratory Medicine, University of Western Australia, Nedlands, WA, Australia.

Polycystic ovary syndrome (PCOS) is a heterogeneous hormonal disorder with a prevalence of 5-8% among women of reproductive age. The syndrome is characterised by elevated blood androgens, hirsutism, amenorrhoea and polycystic ovaries. This condition is also the leading cause of subfertility in women and increases the risk of metabolic complications. PCOS has a genetic basis and 11 loci have been identified to date from GWAS. Defining the underlying genetic mechanisms and functional variants for PCOS will be a key step in developing better understanding of the disease process and development of effective therapies. We recruited first degree relative pairs with physician diagnosed PCOS from a clinical-practice database, for an exome sequencing study. We prepared genomic DNA from the blood of 16 suitable study subjects and performed human exome capture and sequencing using Agilent SureSelect XT Human All Exon +UTR v5 (71Mb), 100 bp paired end sequencing (~5 Gb per sample), targeted 50x mapped coverage with SNV and Indel detection, on the Illumina HiSeq 2500 platform. Exome sequencing yielded an average 73. 4 Mb of DNA sequence per sample with median coverage achieved of 56. 2x across all targets (i. e. 21,522 genes across the genome). Analysis of sequence data of individuals in the study for functional changes in genes previously reported to be associated with PCOS phenotypes yielded results in the following genes: *THADA, INSR, LHCGR, DENND1A, HMGA2, SUOX.* Analysis of the data for variants in affected family members, to determine if private mutations observed were segregating with the disease and therefore presenting stronger evidence of a susceptibility locus for PCOS, identified 31 mutations of potential relevance across the families studied. Among the variants detected across the genome, missense variants in the following genes may be of particular interest: *WDFY3, MAP4K4*. GWAS for PCOS have identified genes of potential relevance, however the variants highlighted are often common SNPs and unlikely to be the causal variants. The results from this study demonstrate the ability of exome sequencing to detect potentially pathogenic variants in known genes and highlight novel genes with a possible role in PCOS, providing targets for additional study and functional characterisation. If functional relevance can be established, the findings may be of potential clinical value in molecular screening and in the creation of genetic risk scores for patients.

741T

Analysis of exome array markers with quantitative lung function measurements in the COPDGene study. *M. M. Parker[1], B. D. Hobbs[2,3], I. Ruczinski[4], R. A. Mathias[5], J. Crapo[6], E. K. Silverman[2,3], T. H. Beaty[1], M. H. Cho[2,3], COPD. Gene Investigators[1,2,3,4,5,6].* 1) Epidemiology, Johns Hopkins University, Baltimore, MD; 2) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA; 3) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, Massachusetts, USA; 4) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore MD, USA 21202; 5) Johns Hopkins Asthma and Allergy Center, The Johns Hopkins University, Baltimore, MD; 6) National Jewish Health, Denver, CO.

Background: Spriometric measures including the forced expiratory volume in one second (FEV1) and the forced vital capacity (FVC) reflect the severity of airway obstruction, and predict population morbidity and mortality. These measures are heritable and substantial advances have been made in our understanding of genetic risk factors underlying lung function using GWAS and large-scale meta analyses. However, these associated genetic risk factors account for only a small fraction of the estimated heritability of lung function. **Methods**:To determine if rare coding variants play a role in lung function, we genotyped 9,858 Non-Hispanic White (NHW) and African American (AA) participants from the COPDGene study with Illumina's HumanExome Beadchip. We performed: 1) single nucleotide variant (SNV) analysis for polymorphic markers, and 2) gene-based analysis of rare variants collapsed within a gene using SKAT. SNV tests included only functional variants with a minor allele frequency (MAF) >0. 5%. Gene based tests included only functional, rare (MAF<5%) variants. All analyses were adjusted for age, gender, pack-years of smoking and principal components to control for population stratification. Statistical significance was determined by Bonferroni correction for the number of tests performed. **Results**:**Single variant tests:**We identified four significant associations between SNVs and spirometry, all of which were nonsynonymous mutations. Three of these four associations had been previously identified in GWAS as risk loci for lung function: rs16969968 in *CHRNA5* (MAF=36. 6%) and rs2070600 in *AGER* (MAF=4. 2%). We identified one novel association between a nonsynonomous SNV and FEV1 percent predicted in AA subjects (rs34664882, p=2. 42 x10-7, MAF=1. 7%). The associated SNV is in the gene encoding the ankyrin 1 (*ANK1*) protein and the minor allele (A) is associated with significantly lower FEV1 percent predicted (b = -11. 59). **Gene based tests:**We identified three significant gene associations with FEV1/FVC or FEV1 percent predicted using SKAT. Associated genes include: *AGER* on chromosome 6 (associated with FEV1/FVC in the NHW group, p=2. 01x10-9); *ProSAPiP1* on chromosome 20 (associated with FEV1 percent predicted in NHWs, p = 2. 63x10-6); and *ANK1* on chromosome 8 (associated with FEV1 percent predicted in AAs, p = 5. 27x10-6). Association signals in these genes were largely driven by a single rare, nonsynonomous variant with a large effect.

742F

Exome chip meta-analysis identifies novel low-frequency variants contributing to central body fat distribution. *K. L. Young[1,2], A. E. Justice[1], T. Karaderi[3], H. M. Highland[1], M. Graff[1], V. Turcot[4], P. Auer[5], N. L. Heard-Costa[6,7], C. Schurmann[8], Y. Lu[8], D. Pasko[9], L. Southam[10], L. A. Cupples[11,6], C. -T. Liu[11], C. S. Fox[6], T. W. Winkler[12], N. Grarup[13], R. A. Scott[14], M. McCarthy[15], K. Mohlke[16], R. J. F. Loos[8], I. Borecki[17], C. M. Lindgren[18], K. E. North[1] on behalf of the BBMRI, GOT2D, CHARGE and GIANT Consortia.* 1) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 2) Carolina Population Center, University of North Carolina, Chapel Hill, NC; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, UK; 4) Montreal Heart Institute, University of Montreal, Canada; 5) Department of Biostatistics, University of Wisconsin-Milwaukee; 6) The Framingham Heart Study, National Heart, Lung, and Blood Institute; 7) Department of Neurology, Boston University School of Medicine; 8) The Genetics of Obesity and Related Metabolic Traits Program, The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai; 9) Genetics of Complex Traits, University of Exeter, UK,; 10) The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK,; 11) Department of Biostatistics, School of Public Health, Boston University; 12) Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, Regensburg, Germany; 13) The Novo Nordisk Foundation Center for Basic Metabolic Research, University of Copenhagen, Denmark; 14) MRC Epidemiology Unit, University of Cambridge, UK; 15) Oxford Centre for Diabetes Endocrinology and Metabolism, University of Oxford, UK; 16) Department of Genetics, University of North Carolina at Chapel Hill; 17) Department of Genetics Division of Statistical Genomics, Washington University School of Medicine; 18) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT.

Central adiposity is a leading risk factor for metabolic and cardiovascular disease, and evidence suggests genetic factors contribute to fat distribution, measured as waist-to-hip ratio adjusted for BMI (WHRa), and to differences in central adiposity prevalence. To date, 49 loci have been associated with WHRa, based on studies of common [minor allele frequency (MAF) ≥5%] single nucleotide variants (SNVs) mostly in populations of European descent. Our aim was to identify low frequency (LFV: MAF <5%) and rare (RV: MAF <1%) coding variants associated with WHRa using Exome-Chip data from 299,666 individuals of European, African, Asian, and Hispanic/Latino ancestry. We performed fixed effects meta-analyses of study-specific WHRa associations stratified by sex and ancestry and then combined across strata for both SNV and gene-based results. We used a strict definition of non-synonymous and missense variants annotated as damaging by 5 algorithms to perform gene-based analyses using the sequence kernel association test (SKAT). Analyses included up to 236,047 SNVs (218,195 with MAF<5%), and 9,175 genes with at least one SNV that met our inclusion criteria. Five LFVs reached chip-wide significance (P<2. 5E-7) in our all ancestry sex-combined analyses, including one novel non-synonymous RV in *RAPGEF3* [MAF<0. 01, b (SE) = -0. 088 (0. 013), P=2. 1E-11]. Gene-based analyses identified *RAPGEF3* (P=1. 42E-11) and *ACVR1C* (P=1. 29E-7) as significantly associated with WHRa after correction for multiple tests (P<2. 5E-6). *RAPGEF3* also shows a significant association (b (SE) = 0. 05 (0. 01), p=4. 68E-12) in all ancestry, sex combined gene-based analysis of BMI. Both *RAPGEF3* and *ACVR1C* are expressed in subcutaneous and visceral adipose tissue, and have been implicated in insulin regulation. *RAPGEF3* plays a role in the GLP1 pathway, which controls insulin secretion in response to blood glucose concentration. In animal models, mice lacking the protein ALK7, coded by *ACVR1C,* show evidence of diminished production of adipokines in white fat cells and reduced energy expenditure by brown fat, both of which can negatively impact glucose homeostasis and insulin sensitivity. Our results highlight the importance of large-scale genomic studies for identifying LFV and RV influencing central fat distribution. Understanding these genetic effects may provide insights into the progression of central adiposity and highlight potential population-specific variants that increase susceptibility.

743W

Whole Exome Sequencing in Families with Chronic Lymphocytic Leukemia Detects a Mutation in *ITGB2 (Integrin Beta 2)* associated with Disease Susceptibility. *L. R. Goldin[1], M. Rotunno[1], M. L. McMaster[1], S. Herman[2], M. Yeager[3], B. Hicks[3], L. Burdette[3], A. M. Goldstein[1], J. Boland[3], S. Ravichandran[4], B. Luke[4], L. Fontaine[5], M. Tucker[6], A. Wiestner[2], G. Marti[2], S. J. Chanock[7], N. E. Caporaso[1], NCI DCEG Cancer Sequencing Working Group.* 1) Genetic Epidemiology Branch, DCEG, NCI, NIH, Bethesda, MD; 2) Hematology Branch, NHLBI, NIH, Bethesda, MD; 3) Cancer Genomics Research Lab, DCEG, NCI, NIH, Bethesda, MD; 4) Simulation, Analysis, and Mathematical Modeling Group, FNLCR, Leidos Biomedical Research, Inc, Frederick, MD; 5) Westat, Inc. , Rockville, MD; 6) Human Genetics Program, DCEG, NCI, NIH, Bethesda, MD; 7) Division of Cancer Epidemiology and Genetics, NCI, NIH, Bethesda, MD.

Chronic lymphocytic leukemia (CLL) shows strong familial aggregation and co-aggregates with other lymphoid malignancies. A number of genes with small to moderate effects have been identified by GWAS but no high-risk genes have been identified. We conducted whole exome sequencing in 69 patients/obligate carriers in 19 CLL families (discovery set), each with 3-5 patients or obligate carriers using a Nimblegen v. 3 exome capture library and paired end sequencing on the Illumina HiSeq2000. Alignment, variant discovery and calling of substitutions, insertions and deletions were performed using standard methods. Variants were analyzed, annotated and prioritized using ANNOVAR, Ingenuity Variant Analysis, and custom R scripts. Dominantly segregating, rare (≤ 1% frequency in European populations), coding or potentially functional variants were further prioritized based on predicted deleteriousness, conservation, and potential importance in lymphoid malignancy pathways. Possible technical sequencing artifacts were eliminated by filtering out variants found in >1% of samples from other non-hematologic cancer studies identically processed in our laboratory. Ten highest priority genes were selected for further targeted follow-up studies using a custom set of AmpliSeq primers on the Ion Torrent platform. For replication, we sequenced 28 additional families with 2 CLL patients plus 12 single patients from multiplex families. A missense mutation in the *ITGB2* gene was detected in one CLL family in the discovery set and was replicated in 3 additional smaller families plus 2 single patients from families with only one patient tested. The mutation has a frequency of 0. 007 in Europeans, is predicted to be damaging using computational tools, and is in a highly conserved region. A few other promising mutations were identified but appear to be “private”. The *ITGB2* gene codes for a B-cell surface marker, CD18, involved in leukocyte adhesion. Other studies in sporadic CLL patients have shown the expression of CD18 is reduced on CLL lymphocytes compared to normal cells and is also associated with disease progression. By flow cytometry, we found that the expression of CD18 in B-cells from patients with the mutation was significantly lower than in patients who did not carry the mutation ($p<0.02$). While the full functional consequences of this mutation are not yet known, our results suggest that it may account for susceptibility to CLL in 10% of high risk families.

744T

Assessment of rare loss of function variants in previously healthy children with pediatric sepsis. *A. Bittencourt Piccini[1], P. J. McLaren[1], I. Bartha[1], C. Berguer[2], J. Fellay[1], L. Schlapbach[3,4,5], Swiss Pediatric Sepsis Study.* 1) School of Life Science, EPFL, Lausanne, VD, Switzerland; 2) University Children’s Hospital Zurich, Zurich, Switzerland; 3) Paediatric Critical Care Research Group, Mater Research, University of Queensland; 4) Paediatric Intensive Care Unit, Lady Cilento Children’s Hospital, Brisbane Australia; 5) Department of Pediatrics, Inselspital, University of Bern Switzerland.

Background: Bacterial sepsis is a leading cause of childhood mortality globally. Sepsis due to community-acquired bacteria in children without clinical risk factors could represent the first manifestation of previously unrecognized primary immunodeficiency (PID). Through exome sequencing, we sought to identify rare human genetic variants that are responsible for an increased susceptibility to sepsis in previously healthy children. Methods: 89 previously healthy children admitted to hospital with sepsis due to community-acquired bacteremia in the absence of known risk factors were enrolled in the study. Inclusion criteria were: 1) age >1 month and <17 years 2) no immunosuppression, chronic disease, or central venal line infections 3) positive blood cultures for *Staphylococcus Aureus*, *Streptococcus Pneumoniae*, or Group A *Streptococcus* 4) systemic inflammatory response syndrome. Exonic regions were selected using the Agilent V5 51Mb enrichment kit and sequenced to high coverage on the Illumina HiSeq2000. Variant annotation was performed using SNPeff v4. 1. Analysis was restricted to rare variants (<1% in the healthy population) annotated as loss of function (LoF), assuming a recessive model of inheritance. Results: We observed 14 homozygous LoF variants (stop-gain, frameshift or splice site) in patients that were either not observed, or observed below 1% frequency in the ExAc reference data set and ~400 in-house controls. These included 3 single nucleotide variants and 11 insertion/deletion polymorphisms affecting 13 genes. Comparing these candidates to a curated list of 256 genes implicated in PIDs showed no overlap. Extending this analysis to include 3,085 genes predicted to be PID candidates through functional connectivity identified 2 genes, *HPR* and *MAP3K1*, where homozygous LoF genotypes were observed in a single individual each. In both genes, homozygous LoF genotypes are rare in the population, with one homozygote observed for *HPR* and none for *MAP3K1* in the ExAc reference data (N~60,000). Conclusions: In this study focusing on highly selected pediatric sepsis cases, no previously described PID was identified. Rare, homozygous LoF genotypes potentially related to PID were identified in two patients. The methodology may prove useful to rule out PIDs of relevance for SIRS survivors and their relatives. Functional testing of the candidates and analysis of enrichment for heterozygous variants and compound homozygotes are ongoing.

745F

Large-Scale Association Study of 4,898 Age-Related Macular Degeneration via Whole-Genome Sequencing. *A. Kwong[1], X. Zhan[1,6], L. G. Fritsche[1], J. Bragg-Gresham[1], K. E. Branham[2], M. Othman[2], L. Gieser[5], R. Ratnapriya[5], D. Stambolian[3], E. Y. Chew[4], A. Swaroop[5], G. R. Abecasis[1].* 1) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI. , United States; 2) Department of Ophthalmology and Visual Sciences, University of Michigan Kellogg Eye Center, Ann Arbor, MI, United States; 3) Department of Ophthalmology and Human Genetics, University of Pennsylvania Medical School, Philadelphia, PA, United States; 4) Division of Epidemiology and Clinical Applications, National Eye Institute/National Institutes of Health, Bethesda, MD, United States; 5) Neurobiology-Neurodegeneration and Repair Laboratory, National Eye Institute/National Institutes of Health, Bethesda, MD, United States; 6) Southwestern Medical Center, University of Texas, Dallas, TX, United States.

Purpose: Age-related macular degeneration (AMD) is one of the major causes of blindness for the elderly. Past studies have examined the role of common variants in the disease, but there has not yet been a systematic study of the role of rare variants in this disease. We obtained whole-genome sequencing samples and plan to perform association studies on a set of variants that includes both SNPs and indels. Methods: We collected genetic samples from over 5,000 volunteers with European ancestry, balanced between cases and controls and matched by age and sex, from the Kellogg Eye Center at the University of Michigan, the Age-Related Eye Disease Study from the National Eye Institute, and the University of Pennsylvania. We used GotCloud to call SNPs and indels, beagle4 to phase the merged callset, and EPACTS to test associations between variants and disease. Results: So far, we have processed whole genomes of 4,898 samples (2,451 cases and 2,447 controls), representing >89 terabytes (8. 9 x 1013 bytes) of sequence data, corresponding to a total genomic coverage of ~30,000x and an average coverage of ~6x per sample. We estimate to have 95. 6% power to detect variants with minor allele frequency > 1%. From our current data, wecan already see an increase of total variants discovered when compared to the 503 European samples of 1000 Genomes Project (phase 3, version 5): for example, in the region around the known AMD risk locus *CFH* on chromosome 1 (195. 7-197. 8 Mb), the 1000 Genomes project contained 31,922 total variants, 151 non-synonymous SNPs, 1,275 deletions, 56 splice sites, and 5 stop gains, while our 4,898 samples contained 100,034 total variants, 646 non-synonymous SNPs, 3,730 deletions, 148 splice sites, and 31 stop gains. In chromosome 1, our current set of 4,898 samples improves our ability to discover variants compared to a previous dataset of 3,278 samples: we discovered a total of 3. 43 million vs. 2. 86 million SNPs, 383k vs. 350k indels, and rediscovered 95. 6% vs. 94. 6% of variants from HapMap3 in the larger vs. smaller datasets, respectively. Conclusion: We provide a more complete look at the genetics of AMD through whole-genome sequencing. With the increased number of rare variants in our dataset, we will be able to discover rare associated variants that could not be discovered with genotyping array technology. Our data will help us better explore the effects of rare coding and non-coding variation, both in previously associated and novel loci.

746W

Targeted deep sequencing of 28 SLE risk loci reveal regulatory haplotypes that potentiate systemic autoimmunity in Caucasians. *P. Raj[1], E. Rai[1], R. Song[1], BE. Wakeland[1], K. Viswanathan[1], C. Arana[1], C. Liang[1], B. Zhang[1], FC. Jonhson[1], M. Mitrovic[2], GB. Wiley[3], JA. Kelly[3], BR. Lauwerys[4], NJ. Olson[5], C. Cotsapas[2], CK. Garcia[6], C. Wise[7], JB. Harley[8], SK. Nath[3], JA. James[3], CO. Jacobs[9], BP. Tsao[10], DR. Karp[11], QZ. Li[1], PM. Gaffney[3], EK. Wakeland[1].* 1) Department of Immunology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA; 2) Department of Neurology, Yale School of Medicine, New Haven, CT, USA; 3) Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104, USA; 4) Pôle de pathologies rhumatismales, Institut de Recherche Expérimentale et Clinique, Belgium; 5) Division of Rheumatology, Department of Medicine, Penn State Medical School, PA, USA; 6) Eugene McDermott Center for Human Growth & Development, Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, TX 75335 USA; 7) Texas Scottish Rite Hospital for Children 2222 Welborn Street, Dallas, Texas 75219; 8) Cincinnati Children's Hospital Medical Center and Cincinnati VA Medical Center, Cincinnati, OH, USA; 9) Department of Medicine, University of Southern California, Los Angeles, CA 90089, USA; 10) Department of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA; 11) Rheumatic Diseases Division, Department of Medicine, University of Texas Southwestern Medical Center, TX 75335, USA.

Systemic lupus erythematosus (SLE) is an autoimmune disease caused by loss of humoral immune tolerance leading to the production of autoantibodies to a spectrum of self-antigens. Genetic predisposition is key for SLE susceptibility, however little is known about the nature or functional properties of causal genetic variants. We used targeted population sequencing to comprehensively characterize genetic variability at 28 risk loci for SLE in a panel of 1349 Caucasian SLE cases (773) and controls (576). The HLA-D region contains the strongest risk loci identified for SLE, with multiple alleles of both HLA-DR and -DQ showing strong associations. Sequence analysis of the 380 Kb segment spanning the BTNL2-DR-DQB2 region identified 15,261 common (MAF >0. 05) genetic variants. Analyses of these sequence-defined HLA-D variations identified three independent risk-associated signals reaching genome wide significance. Subsequent analyses demonstrated that these disease-associated variations are imbedded in a series of stable haplotypes formed by multiple, ENCODE and eQTL-defined functional variations impacting the transcription of more than 20 genes that encode components of the antigen processing and presentation (APP) pathways of HLA class I and class II genes. Median neighbor joining analyses identified three HLA-D region regulatory haplotypes forming a risk clade strongly associated with SLE, all of which contained eQTL variants that increased the transcription of HLA-DR, DQ, DP, and other elements of the APP pathway in multiple myeloid and lymphoid cell lineages. This risk clade contains all of the classical HLA-D class II alleles previously associated with SLE, indicating that the systemic upregulation of the APP pathway is a consistent feature of all SLE-associated HLA-D alleles. Similar analyses of non-HLA SLE risk loci identified regulatory haplotypes that were often associated with transcriptional changes in multiple genes within specific pathways. Our analyses demonstrate that such regulatory haplotypes have increased disease-associated odds ratios in comparison to the disease odds for maximal GWAS tagging SNPs in these loci. These findings are consistent with the hypothesis that the functional variations that underlie many common disease alleles form regulatory haplotypes that modulate the transcription of multiple cis and trans genes in immune system pathways and that their functional phenotypes are potent and complex.

747T

Polymorphism of gene *PNPLA3* in Brazilian patients infected with HCV with severe steatosis. *P. Moura[1], M. T. B. Wanderley[1], R. F. Carmo[2], V. C. S. Sousa[1], R. E. Lima[1], E. R. F Siqueira[3], S. C. R. Carvalho[3], L. M. M. B. Pereira[3], M. S. M. Cavalcanti[1], B. M. Carvalho[1], L. R. S. Vasconcelos[3,4].* 1) Biological sciences institute -University of Pernambuco, Recife, Brazil; 2) Federal University of São Francisco Valley- UNIVASF, Brazil; 3) Liver and transplant institute of Pernambuco - IFP, Brazil; 4) Aggeu Magalhães Research Center - FIOCRUZ-PE , Brazil.

Background & Aims: Patatin-like phospholipase domain-containing protein 3(PNPLA3)located on the long arm of Chr. 22 code for a triacylglycerollipasethat mediatestriacylglycerol hydrolysis inadipocytes. GWAS showed the *PNPLA3* I148M (rs738409) variation was associated to steatohepatitis. Patients with chronic infection by hepatitis C virus (HCV) frequently present steatohepatitis, which may contributes to liver disease progression. Our study evaluated the influence of *PNPLA3* gene polymorphism in Brazilian HCV infected patients and steatosis outcome. *Methods:* Four hundred twelve patients infected by HCV with negative serological test for HBV and HIV were enrolled in this study; they were stratified by intensity of steatosis in tree groups based on the liver histopathological exam: G1-no steatosis, G2-mild, G3-moderate/intense and G4- mild/moderate/intense. Genotyping for thers738409 *PNPLA3* gene was performed by real-time PCR using TaqMan Genotyping Assay *Results:* Biochemical analysis showed increase of hepatic enzymes (ALT, GGT and ALP) in groups with steatosis; higher blood glucose and Body Mass Index in the group 3 compared with group 1 (p=0. 004 and p=0. 003, respectively). We showed an association of the G allele with steatosis (G4) *vs.* no steatosis (G1) (p=0. 02, OR 1. 75 CI 1. 08-2. 87) as well as when compared G1 *vs.* G3 (p=0. 002, OR 2. 29 CI 1. 31-4. 01). In recessive model there was association of CG and GG genotypes with increase of moderate/intense steatosis (p=0. 034 OR 2. 23 CI 1. 06-4. 72) and (p=0. 018 OR 6. 88 CI 1. 29-48. 6) respectively. Additionally, for exclude potential confounding variables we perform a multivariate analysis and include in the model the statistically significantly variables in the univariate analysis. We found that genotypes CG/GG was significantly associated for moderate/intense steatosis prediction independently of male gender, BMI, glucose levels and HCV genotype 1, p=0. 010 OR 1. 869 (1. 15-3. 01). *Conclusions:* Thus, the *PNPLA* polymorphism rs738409 seems to influences steatosis severity in Brazilian patients infected with HCV and could be a useful marker for steatosis progression in this population.

748F

Whole Exome Sequence Meta-analysis of 13 White Blood Cell, Red Blood Cell, and Platelet Traits. *L. M. Polfus[1], U. M. Schick[2], P. L. Auer[3], A. P. Reiner[4], S. K. Ganesh[5], the CHARGE and NHLBI GO Exome Sequencing Project Hematology Working Groups.* 1) Human Genetics Center, University of Texas at Houston, Houston, TX. USA; 2) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY. USA; 3) The Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA. USA; 4) Department of Biostatistics, University of Wisconsin-Milwaukee, Milwaukee, WI. USA; 5) Department of Internal Medicine and Department of Human Genetics, University of Michigan, Ann Arbor, MI. USA.

Blood cell counts and erythrocyte indices are clinically important indicators of a variety of disease states. The ability to detect and include novel rare variants in gene-based tests is best accomplished via sequencing in well-phenotyped samples. In order to identify novel genes influencing hematologic traits, we analyzed whole exome sequencing data in 14,770 participants of 6 population-based cohort studies for 13 traits in European descent (EU) and African Americans (AA), analyzed together and separately, in fixed-effects meta-analyses. All traits underwent rank based inverse normal transformation of the trait residuals to account for outliers influence on rare variants. We analyzed variants individually and aggregated functional and loss-of-function (LOF) variants with MAF<5% within genes using SKAT and T5 burden tests. Discovery phase associations identified 5 novel gene-based results and 3 novel single variants meeting exome-wide Bonferroni corrected thresholds (2. 6x10-6 and 1. 6x10-7). All 5 novel gene-based results were associated with red blood cell traits. In race-combined analyses, rare functional variants in *MYOM2* were associated with mean corpuscular hemoglobin concentration (SKAT PEU+AA=2. 2x10-6). In EUs, rare variants in *MRPL43* were associated with hemoglobin (T5 PEU=1. 2x10-6), and those in *PLAGL1* were associated with mean corpuscular hemoglobin (T5 PEU=1. 5x10-6). LOF Variants in *MMACHC* were associated with hemoglobin (SKAT PEU=1. 3x10-6). *MMACHC* regulates vitamin B12 metabolism and known variants cause a Mendelian syndrome including anemia. In AAs, rare functional variants in *ACTN4* were associated with mean corpuscular volume (SKAT PAA=1. 5x10-6). *ACTN4* variants were previously associated with focal segmental glomerulosclerosis of the kidney. Among the single variant tests, we found low-frequency variant rs9656446 (MAF=0. 03, PEU+AA=1. 5x10-7) in *AGBL3* to be associated with basophil count. A common variant in *CPS1* (rs1047891, MAF=0. 33, PEU+AA=5. 7x10-8) was associated with platelet count. We also identified an association of platelet count with a low-frequency (MAF=0. 009) synonymous variant (rs150813342, PEU=4. 7x10-8) in *GFI1B,* a gene known to cause gray platelet syndrome. *In silico* algorithms predict rs150813342 to affect exon 5 splicing. Replication efforts are currently ongoing. In this exome sequencing study, we identified novel low frequency loci, which may be followed in subsequent functional studies to describe attenuated disease states.

749W

Association of several genes and prostate cancer susceptibility using the Exome-Chip genotyping array. *A. Amin Al Olama[1], S. Benlloch[1], P. D. Pharoah[2], J. Brown[3], K. R. Muir[4,5], A. Lophatananon[5], D. Leongamornlert[6], E. Saunders[6], T. Dadaev[6], K. Govindasami[6], Z. Kote-Jarai[6], D. F. Easton[1], R. A. Eeles[6,7], United Kingdom Genetic Prostate Cancer Study Investigators.* 1) Centre for Cancer Genetic Epidemiology, Department of Public Health & Primary Care, University of Cambridge, Cambridge, UK; 2) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK; 3) Centre for Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 4) Institute of Population Health, University of Manchester, Manchester, UK; 5) Warwick Medical School, University of Warwick, Coventry, UK; 6) Institute of Cancer Research, London, UK; 7) Royal Marsden National Health Service (NHS) Foundation Trust, London and Sutton, UK.

Prostate cancer is the most common non-skin male cancer in the western world and one of the leading causes of cancer related death. Genome-wide association studies have identified over 100 common, low penetrance susceptibility variants. So far only a small number of susceptibility rare variants have been identified. To identify rare coding alterations associated with prostate cancer, we genotyped 8,797 samples from four studies (UKGPCS, SEARCH, SIBS, and UKO) in the UK. After quality control exclusions, we analysed 153,155 SNPs in 3,937 cases and 4,423 controls of European ancestry. We performed single SNP analysis using the Wald, score and likelihood ratio tests. Two gene-based analysis were carried out; a cohort allelic sum test (CAST) and the sequence kernel association test (SKAT). For CAST method we took all variants in a given gene and summed up the minor alleles in cases and controls. Then we used a logistic regression to discover if the difference in these sums is greater than would be expected by chance. SKAT aggregates score test statistics of SNPs in a gene and computes gene level *P*-values. All analyses were adjusted for two principal components. We analysed 16,744 genes and carried out analyses twice; once including all 153,155 SNPs and once including SNPs with MAF ≤ 1%. Single SNP analysis identified 25 SNPs at $P < 5x10\text{-}8$, 5 out of 25 SNPs had MAF less than 1%. For gene-based analysis we used a *P*-value threshold of 2. 8x10-6 after Bonferroni correction for the number of gene tested. We identified 16 genes (*NLGN4X, CD24, USP9Y, PCDH11Y, ZNF673, SSX5, PLCB2, MLNR, OPN1LW, OPN1LW, FRMPD2, TAF1, SLC38A8, MAST4, OR6C3, GAMT*) associated with prostate cancer using both burden tests. These results need to be replicated in larger studies as the study size was limited.

750T

Rare variants in the functional domains of *Complement Factor H* are associated with age-related macular degeneration. *E. D. O Roberson[1,2], M. Triebwasser[1], Y. Yu[3], E. C. Schramm[1], E. Wagner[3,4], S. Raychaudhuri[5,6,7,8], J. M. Seddon[3,4], J. P. Atkinson[1].* 1) Washington University School of Medicine, Department of Internal Medicine, Division of Rheumatology, St. Louis, MO, USA; 2) Washington University School of Medicine, Department of Genetics, St. Louis, MO, USA; 3) Ophthalmic Epidemiology and Genetics Service, New England Eye Center, Tufts Medical Center, Boston, MA, USA; 4) Department of Ophthalmology, Tufts University School of Medicine, Sackler School of Graduate Medical Sciences, Tufts University, Boston, MA, USA; 5) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA; 6) Partners HealthCare Center for Personalized Genetic Medicine, Division of Genetics, Brigham and Women's Hospital, Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Boston, MA, USA; 7) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK; 8) Department of Medicine, Karolinska Institutet, Solna, Sweden.

Age-related macular degeneration (AMD) has a substantial genetic risk component, and several common variants, including Y402H in complement Factor H (*CFH*), have been identified. More recently, rare variants have been discovered to play an independent role in AMD risk. In a previous study, we found association between advanced AMD (A-AMD) risk and the presence of rare variants in the complement component 3 (*C3*), complement factor I (*CFI*), and complement component 9 (*C9*) genes. We hypothesized that rare variants in *CFH* would be a logical contributors to AMD risk. In this study we evaluated whether rare variants in *CFH* contribute to AMD risk independent of two known, common risk alleles. All cases had previously been examined to confirm their A-AMD disease status, and controls had been examined to exclude a diagnosis of AMD. We captured coding regions of approximately 700 candidate genes with hybridization based capture, and sequenced the enriched DNA by Illumina HiSeq. We used the Genome Analysis ToolKit Unified Genotyper tool to genotype the resulting sequencing data after alignment with bwa. We tested 1,665 cases and 752 controls for the effects of rare variants in *CFH* using logistic regression, where the common variants were coded as an additive allele dose and presence of a rare variant was coded as a 0/1 indicator variable. Significance of enrichments was calculated by randomly permuting case/control status and determining how often an equal or larger odds-ratio (OR) was observed. We identified 65 missense, nonsense or splice-site alterations with a control minor allele frequency (MAF) of ≤1%. Rare variants with a MAF ≤1% (OR=1. 5, p=4. 0X10-2) or ≤0. 5% (OR=1. 7, p=2. 4x10-2) and singletons (OR=2. 4, p=2. 6x10-2) were all significantly enriched in A-AMD cases compared to controls. Rare variants in the three major functional domains of *CFH* were also enriched in cases (OR=3. 4, p=9. 8x10-4). Additionally, serum factor H levels in 10 cases that carried rare loss-of-function variants that resulted in a nonsense allele, splice-site alteration or loss of a conserved cysteine were less than the 10th percentile of factor H levels in 60 controls. Like factor I, low serum levels of factor H will be a strong predictor of risk of developing A-AMD. Together, these data further implicate failure of complement cofactor activity in the retina as a central component of AMD pathophysiology, and highlight the importance of both rare and common variants in AMD risk.

751F

Stop gain mutation in the *ZNF528* gene co-segregates with familial early onset spinal-osteoporosis. *S. Skarp[1,2,3], M. Löija[1,2,3], L. Ruddock[2], O. Mäkitie[4], M. Männikkö[1,2,3].* 1) Center for Life Course Epidemiology and Systems Medicine, Faculty of Medicine, University of Oulu, Finland; 2) Faculty of Biochemistry and Molecular Medicine, and Biocenter Oulu, University of Oulu, Finland; 3) Oulu Center for Cell - Matrix Research, University of Oulu, Finland; 4) Division of Pediatric Endocrinology and Metabolic Bone Diseases, Central Hospital and Folkhälsan Institute of Genetics, University of Helsinki, Helsinki, Finland.

Osteoporosis (OP) is the most common bone disorder affecting more than 200 million people around the world. It affects most often the elderly and especially postmenopausal women. As the human population ages cases with OP are increasing making OP a growing socioeconomic challenge. OP is characterized by lowered bone mineral density (BMD) and deterioration of the bone's microarchitecture leading to porous and weakened bones. This in turn leads to increased fracture risk. Osteoporotic fractures (OF) are a severe clinical outcome of osteoporosis. Both OF and BMD are heritable traits with BMD has heritability for BMD ranging from 50-80% and for OF 50-70%. The genetic background of OP is still largely unknown. Studying genes associated with rare OP forms improves the understanding of genetic mechanisms and pathogenesis of this complex disorder. We carried out whole exome sequencing in a Finnish family of five with three male patients suffering from early onset OP of the spine with compression fractures. The exomes of healthy family members and an in-house exome set (N=71) were used to identify variants co-segregating with the disease. We focused on rare (MAF < 0. 01) and private variants with harmful *in silico* estimations. We observed 1236 variants shared by the affected individuals. Out of these variants 239 were exonic or splice site variants. Ten of these were chosen for genotyping based on functional annotation in the five family members and three additional healthy relatives. A stop gain mutation in gene coding zinc finger protein 528 (*ZNF528*) co-segregated with the disease. *ZNF528* encodes a zinc finger protein which is a DNA binding transcription factor. It is expressed in various tissues and cells including osteoblasts. The novel stop gain mutation disrupts the 8th C2H2-type zinc finger of fifteen found in the protein and hence is likely to affect the function of ZNF528. The roles of the gene and the mutation are being studied further in functional studies. This gene is a novel candidate for osteoporosis and bone diseases.

752W

The Helmsley Inflammatory Bowel Disease Exome Sequencing Program (HIESP) – Whole Exome Sequencing of 20,000 Samples. *C. R. Stevens[1], M. A. Rivas[1], R. J. Xavier[1,2,3], M. J. Daly[1,2].* 1) Medical and Population Genetics, The Broad Institute, Cambridge, MA; 2) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 3) Gastrointestinal Unit and Center for the Study of Inflammatory Bowel Disease and Center for Computational and Integrative Biology, Massachusetts General Hospital, Harvard Medical School, Boston, MA.

Crohn's disease (CD) and ulcerative colitis (UC) are debilitating, inflammatory diseases of the gastrointestinal tract, collectively known as the inflammatory bowel diseases (IBD). Among complex diseases, genetics has been particularly successful in the identification of genes for IBD. With support from the Helmsley Charitable Trust, we have launched a transformative program of exome sequencing in IBD for which we continue to engage collaborative partners. The overarching aim of the program, outlined here, is to define the full allelic spectrum of protein-altering variation in genes associated to IBD, assess their role in both CD and UC risk, clinical course and response to therapy, and to determine whether loss-of-function variants confer risk or protection in each IBD gene in order to articulate the most opportune therapeutic targets. Recent technical innovations in DNA sequencing and analysis enable exome sequencing to take place at an unprecedented low cost and high accuracy and have facilitated the launch of this program, which aims to evaluate at least 20,000 exomes. The long-term aim of this program is a complete assessment of the role of rare coding variation to IBD risk and protection. In order to achieve these ultimate goals, we are setting up a collaborative infrastructure for exome sequencing studies to provide sequencing and analysis support for many such studies. As with GWAS, achieving the long term objective of the complete articulation of the contribution of rare variation will ultimately require the analysis of tens of thousands of samples and a worldwide collaborative effort. Here we embark on a series of projects that take advantage of established strategies to increase the power of genetic studies to discover high-impact variants, or which address specific questions of clinical importance over and above IBD risk. In the first year, we specifically have launched exome sequencing of cohorts from the following areas: 1) IBD in Unique Populations, 2) Very Early Onset IBD and Infantile GI, 3) Adverse Drug Response in IBD, and 4) Genetics with Deep Clinical Phenotypes. These categories, focused on specific populations or clinical hypothesis, aim to maximize the impact of HIESP - presenting an unprecedented opportunity to examine the role of rare variants and its contribution to inflammatory bowel disease predisposition.

753T

Identification of Polymorphisms Associated with Extrapulmonary Dissemination of Coccidioidomycosis (Valley Fever). *M. Yourshaw[1], P. Krogstad[1], R. Johnson[2], G. Ameri[2], H. Abukamleh[2], S. Meyer[2], D. Aguirre[2], R. Grewal[2], A. Heidari[2].* 1) Pediatrics, UCLA, Los Angeles, CA; 2) Kern Medical Center, Bakersfield, California.

INTRODUCTION: Valley Fever (coccidioidomycosis) is a fungal infection caused by the inhalation of airborne spores found in the soil. It infects some 150,000 persons each year, most in the central valley of California, and has increased 8-fold in California in the past decade. The disease is found where dry, hot, windy conditions favor dispersion of spores, and has received national attention as climate change increases its range. In some patients the fungus spreads to the skin, bones, and brain, where it may be fatal. There is no cure and treatment can be as debilitating as chemotherapy. The disseminated form of the disease especially affects African- and Filipino-Americans, and thus likely has a genetic component. **METHODS:**We extracted DNA from peripheral blood or archived tissue from a group of 22 ethnically diverse adults with extrapulmonary dissemination of coccidioidomycosis and performed whole exome sequence analysis. We filtered the resulting data for variants in genes known or suspected to be in pathways relating to responses to fungal infections. We sequenced candidate genes in additional samples. **RESULTS:** We identified an enrichment of variants predicted to be damaging in genes associated with resistance to fungal infection, including CHIT1, CLEC7A, and IL17RA. **CONCLUSIONS:** In a pilot study of 22 cases of severe disseminated cocci we found genetic variants in immune system genes that appear to be enriched in susceptible populations.

754F

Mutation in the chromatin-remodeling factor *BAZ1A* is associated with intellectual disability. *A. Zaghlool, J. Halvardson, J. Zhao, M. Etemadikhah, A. Kalushkova, K. Konska, H. Jernberg-Wiklund, A-C. Thuresson, L. Feuk.* Immunology, Genatic & Patholohy, Science For Life Laboratory, Uppsala University, Uppsala, Sweden.

Exome sequencing of trio samples has led to the identification of mutations in several genes involved in chromatin remodeling in syndromic forms of intellectual disability. To date, the majority of these mutations have not been functionally investigated to validate and understand the pathological contribution of these mutations to the disease symptoms. Here, we used exome sequencing to identify a single non-synonymous de novo mutation in *BAZ1A*, encoding the ATP-utilizing chromatin assembly and remodeling factor 1 (ACF1), in a patient with severe intellectual disability, autism, profound speech impairment, seizures, hypotonia, hypoplastic finger-nails and distal phalanges. ACF1 is a member of the ISWI chromatin-remodeling complexes ACF and CHRAC. It has been previously reported to have a function in transcriptional repression of Vitamin D3 receptor (VDR) regulated genes by blocking the accessibility of the transcription factors to VDR in the absence of vitamin D3. Based on RNA sequencing of the family trio, we find that the mutation in *BAZ1A* affects the expression of many genes, mainly involved vitamin D3 metabolism, Wnt signaling, axon guidance and nervous system development. Notably, *CYP24A1,* involved in vitamin D3 metabolism, and *SYNGAP1,* involved in axonal formation and synaptic function and *SMARCA4,* previously linked to Coffin-Siris syndrome, are differentially expressed in the patient. The differential expression of these genes correlates with the clinical diagnosis of the patient. By chromatin immunoprecipitation analysis, we show that the differential expression of *CYP24A1* is caused by reduced binding affinity of the mutated ACF1. Furthermore, RNA interference experiments further supports the role of *BAZ1A* in the regulation of the differentially expressed genes. We therefore propose that *BAZ1A* represents yet another chromatin remodeling gene involved in causing an intellectual disability syndrome. Currently, we are performing RNA sequencing of brain samples from *BAZ1A* knockout mice to further understand the regulatory role of ACF1 in the brain.

755W

Whole exome sequencing in a family with early onset primary angle-closure glaucoma. *B. B. Souza[1,2], J. P. C. Vasconcellos[3], B. S. Carvalho[4], M. G. Borges[4], I. T. Lopes Cendes[4], P. A. O. R. Aguiar Araújo[4], M. B. Melo[1].* 1) Center of Molecular Biology and Genetic Engineering, University of Campinas - Unicamp; Campinas SP, Brazil; 2) Institute of Biology, University of Campinas - Unicamp, Campinas, SP, Brazil; 3) Department of Ophthalmology, Faculty of Medical Sciences, University of Campinas - Unicamp, Campinas SP, Brazil; 4) Department of Medical Genetics, Faculty of Medical Sciences, Unicamp, Campinas, SP, Brazil.

Background: Glaucoma is a degenerative disease of the optic nerve considered the leading cause of irreversible blindness worldwide, which affects about 64. 3 million people between 40 – 80 years old. Primary angle-closure glaucoma (PACG) is a subtype of glaucoma caused by obstruction of the iridocorneal angle with increase in intraocular pressure leading to slow progressive excavation of the optic disc and corresponding visual field loss. Demographic factors like age, ancestry and female sex are still recognized as the main risk factors for PACG, as well as positive family history. PACG is a complex heterogeneous disease with molecular mechanisms poorly understood. Until now, only three genetic markers where associated with susceptibility to PACG, rs11024102 in *PLEKHA7* on chromosome 11, rs3753841 in *COL11A1* on chromosome 1, and rs1015213 located between *PCMTD1* and *ST18* genes on chromosome 8q, all these variants were identified from GWAS studies. **Objective:** To identify genes associated with PACG in an early onset disease family. **Methods:** Whole exome sequencing was performed in a family of four members, being mother and daughter affected by PACG and father and son unaffected. Exome was targeted with Nextera Rapid Capture Expanded Exome kit (Illumina™) and sequenced in a high-performance Hiseq Illumina 2500 DNA Analyzer (Illumina™) to obtain more than 70X average coverage per sample. A bioinformatics analysis was performed using the GATK software package. Sequences were aligned using BWA algorithm. Variant calling and functional prediction were performed through VariantAnnotator and VEP tools. Non-synonymous, frameshift, splicing, and indel variants were prioritized according to novelty, quality score, and degree of pathogenicity. **Results:** We found a total of 376. 742 variants in this family. After bioinformatics processing, we identified 317 functional variants that are shared only by patients and absent in the unaffected individuals. Among them, we observed four genes as potential candidates for PACG in the studied family. **Conclusions:** Putative roles related to cell adhesion, spreading, growth and apoptotic processes make the four candidate genes identified relevant for PACG pathophysiology. **Supported by:** Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPQ, Brazil, process: 157539/2013-0.

756T

Detection of somatic genetic variation at the *LMNA* locus in non-progeroid individuals. *R. R. Palla[1], N. Viceconte[1], A. Witasp[2], E. Wallén Arzt[1], P. Stenvinkel[2], M. Eriksson[1].* 1) Department of Biosciences and Nutrition, Karolinska Institutet, Huddinge, Sweden; 2) Department of Clinical Science, Intervention and Technology, Division of Renal Medicine, Karolinska Institutet, Stockholm, Sweden.

Several recent studies have investigated structural genetic variation in the context of somatic mosaicism, but so far only few studies have analyzed somatic genetic variation at the single-nucleotide level. Hutchinson-Gilford progeria syndrome (HGPS) is a rare genetic disorder with features reminiscent of premature aging. Previous studies have shown that >90% of affected patients have a mutation in the *LMNA* gene (c. 1824C>T, p. G608G). This mutation and several other mutations have been shown to result in increased activation of a cryptic splice donor site in exon 11, resulting in a truncated lamin A protein, named progerin. Progerin has also been found at low levels in human cells from non-progeroid diseases and its levels were shown to increase during aging. There is no explanation to the expression of progerin in non-progeroid tissue or its contribution to tissue pathology and aging. In this study we aimed to test if occasional progerin expression in human tissues is the result of somatic mutations. We used droplet digital PCR (ddPCR) to analyze single nucleotide variants in a highly mosaic tissue sample. Eight ddPCR assays were developed to test variants in the *LMNA* gene. These variants have previously been identified in children with Progeria and Werner syndrome or recorded in SNP databases. Prediction of alternative splicing was performed using the Alamut Visual mutation analysis software that estimates each mutation effect on splicing. The presence of low-degree somatic single nucleotide variants was analyzed in peripheral blood mononuclear cell (PBMC) DNA samples from 26 patients with end stage kidney disease (ESRD; a group of patients with a prematurely aged phenotype) and controls. The ddPCR allowed detection of low frequency variants in the PBMC DNA, suggesting the presence of low-degree somatic mosaicism. However, the mutation frequency detected in the PBMC from ESRD patients was below, or at similar levels, to that found in unaffected controls. Our results suggest that a more extensive analysis, including a map of progerin expression in several tissues, together with low-degree somatic variation analysis using ddPCR on DNA from the same tissue, is needed to fully explain the significance of these findings. Nevertheless, detection of low frequency genetic variants in non-progeroid patients, previously recorded in premature aging syndromes, might suggest a role for these somatic mutations also in normal aging.

757F

Somatic mutation calling in lung tissue of ex-smokers with different degrees of airway obstruction. *G. -A. Thun[1], S. Derdak[1], F. Castro-Giner[1], L. Ziegler-Heitbrock[2], I. Gut[1], EvA and AirProm Consortium.* 1) Centro Nacional de Análisis Genómico, Parc Cientific de Barcelona, Barcelona, Spain; 2) EvA Study Center, Helmholtz Center Munich, Gauting, Germany.

Chronic obstructive pulmonary disease is a smoking-related disorder characterized by airflow limitations and a high degree of inflammation and tissue remodeling in the lung. As the association with lung cancer can only be partly explained by the duration or amount of smoking, we hypothesized that a higher susceptibility to mutagenesis might be present in the affected lungs. For that purpose, twelve elderly participants of the EvA Study, including both never and former smokers, were selected and whole genome sequencing of lung brush and blood samples (average median coverage 30x) was carried out. The nine ex-smokers all had a history of heavy smoking, but very variable degrees of airflow obstruction. In the absence of a single best protocol for calling somatic mutations, we used several of the existing methods comparing pairwise data of lung brush tissue with blood as the reference. Applying default settings, we observed high heterogeneity with respect to the number of single somatic mutation (SSM) calls (means ranging from ≈600 to 8000). The concordance among the methods was modest (overlapping calls for SSMs between 5 and 48% of those produced by the more restrictive caller in all pairwise method comparisons). Moreover, taking intersects did not appear a fruitful strategy to confine the most promising candidates as we found substantial enrichment of variants present in dbSNP. The results of the selected methods agreed with respect to the frequency of the variant allele peaking close to 0. 25, which suggested some degree of polyclonality or considerable presence of white blood cells in the lung specimen. Independent of the method applied, the number of somatic mutations was neither associated with airflow obstruction nor with smoking history in our small data set. However, we found the proportion of tobacco-related C>A transversions to be associated with former smoking depending on the calling method. Even in the hitherto absence of call verification, the low concordance and the relative high overlap with dbSNP point to low positive predictive values of the applied methods. Careful method-specific post-filtering of the results seems indispensable. Furthermore, ultra-deep sequencing will be necessary to discover if the number of somatic mutations in lung epithelial tissue normalizes years after quitting smoking or if the high polyclonality in brush samples obscured the elevated numbers of somatic mutations in ex-smokers or individuals with airflow obstruction.

758W

Genetics of Obesity and Diabetes Related Quantitative Traits in Omani Arabs. *N. N. Al Kharusi[1], S. A. ALBARWANI[1], M. O. HASSAN[1], R. A. BAYOUMI[1], H. M. HIGHLAND[2,4], M. C. GINGRAS[3], D. M. MUZNY[3], R. A. GIBBS[3], E. A. BOERWINKLE[3,4], L. E. PETTY[4], J. E. BELOW[4], S. A. AL YAHYAEE[1].* 1) Biochemistry and Molecular Biology, Sultan Qaboos University, Muscat, Oman; 2) Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA; 3) Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA; 4) Human Genetics Center, University of Texas Health Science Center, Houston, TX 77225, USA.

Obesity (OB) and type 2 diabetes (T2D) are heterogeneous complex diseases driven by environmental and polygenic effects. Previously identified variants of OB and T2D and their traits account only for a fraction of their expected heritability. Screening for variants in an isolated founder population can reduce the phenotypic heterogeneity explained by environmental effects and focus on enriched segregating genes that explain the genetic risk of the disease. A single large consanguineous family of 227 Arabs from Nizwa village in Oman were genotyped on the Illumina Infinium chip. Heritability estimates of underlying characteristics of OB and T2D, such as body mass index (BMI), waist circumference (WC), fasting glucose (FG) and 2 hour glucose (2hrG) in the village range form 0. 17 (2hr glucose) to 0. 68 (BMI). Quality control of genotyped data was performed with PLINK1. 9 and individuals with excess heterozygosity (F statistic < -0. 1) or missingness of >1%, and SNPs with >10% missingness, and HWE p <0. 001 were removed, leaving 221 individuals with 239,464 genotyped SNPs for analysis. Single variant association analyses of quantitative trait were performed in EPACTS-EMMAX accounting for the pairwise relatedness structures. Both normal residual and inversed normalized residual of the traits were tested. Although no genes were genome wide significant (*p* value < 5 x 10-8), our top associations include variants in *PLD1* (2hrG, rs145015377 (G/A), *p* value 2. 1 x 10-5), *SERPINA6* (BMI, rs2228541 (C/A), *p* value 2. 5 x 10-5), *IRF4* (WC, rs9405666 (A/G), *p* value 1. 93 x 10-5). Gene *PLD1* encodes a phosphatidylcholine-specific phospholipase which was reported to interact with *PED*/*PEA-15* gene and influence insulin-stimulated glucose disposal. *SERPINA6* encodes an alpha-globulin protein, a major transport for glucorticoids and progestins in the blood, which influence body adiposity. Variants in *IRF4* were found to be associated with waist circumference; interferon regulatory factor 4 is required for lipolysis. It is up regulated during fasting, and its absence stimulates lipid synthesis. Although the low sample size in these preliminary analyses did not provide sufficient power to reach genome-wide significance at variants of modest effect, gene function suggests an enrichment of phenotypically relevant pathways in top signals. Formal analyses of functional enrichment and analyses leveraging the unique haplotypic architecture of this cohort will increase power.

759T

Identification of Rare Variants in Schizophrenia Families. *Z. Brkanac, L. F. F. Gelin, M. J. Archer, R. R Nesbitt, A. Patowary, J. Rehker.* Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA.

Schizophrenia is a rare and severe psychiatric disorder affecting approximately 1% of the adult population worldwide. Symptoms can include hallucinations, delusions and catatonic behaviors as well as negative symptoms including a lack of initiative, motivation, affect expression and poverty of thinking. Schizophrenia is a highly heritable and complex disorder with multiple genes contributing to its cause and spectrum of symptoms. In order to identify novel genes responsible for schizophrenia we focused on 16 multiplex families of different ethnicities (two Hispanic, five African-American and nine European-American) from NIMH collection. Using exome sequencing of affected subjects, we identified alleles that are shared among all cases in each family. In order to identify candidate genes, we focused on private (not present in dbSNP 132, 1000 genomes or ESP 6500) and rare (not present in dbSNP 132 and with frequencies lower than 1% in the same databases) protein changing SNPs (nonsense, splicing, frameshift and missense). The exome sequences were analyzed using the BWA-Picard-GATK workflow and annotated using ANNOVAR. Data manipulation and filtering was done using custom python scripts. The number of rare and private variants within families varied from 5 to 45. That data reflects the different family structures with expected genomic sharing among affected individuals ranging from 0. 08% to 1. 25%. Based on the analysis of rare and private SNPs, 256 genes were selected for association analyses. For the association analysis, we used the published Swedish schizophrenia exome data set containing 2,536 cases and 2,543 controls. Gene based and single variant association tests were performed for the 256 genes using PLINK/SEQ. In addition, we performed chi-square analysis for a number of rare and unique variants in cases vs. controls. Using this approach we identified 19 candidate genes that either had either significantly more variant counts in cases than in controls (15 genes, p<0. 01) or had one variant that appeared more often in cases than in controls (4 variants, p<0. 01). To confirm the association from the first sample, multiplex target sequencing is being performed for the 19 candidate genes in 960 independent schizophrenia cases and 960 controls from NIMH collection. The results of the combined analysis will be presented.

760F

Whole-genome sequencing reveals schizophrenia risk mechanisms in humans with 22q11. 2 deletion syndrome. *M. Zarrei[1], D. Merico[1], G. Costain[2,3], L. Ogura[2], B. Alipanahi[4], M. Gazzellone[1], N. Butcher[2], B. Thiruvahindrapuram[1], T. Nalpathamkalam[1], E. Chow[2,5], D. Andrade[6,7], B. Frey[4], C. Marshall[1], S. Scherer[1,8], A. Bassett[2,5,9,10,11].* 1) The Centre for Applied Genomics and Program in Genetics and Genome Biology, The Hospital for Sick Children, Ontario, Canada; 2) Clinical Genetics Research Program, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 3) Medical Genetics Residency Training Program, University of Toronto, Toronto, Ontario, Canada; 4) Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario Canada; 5) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 6) Division of Neurology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada; 7) Epilepsy Genetics Program, Toronto Western Hospital, University Health Network and University of Toronto, Toronto, Ontario, Canada; 8) McLaughlin Centre and Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 9) Department of Psychiatry, and Toronto General Research Institute, University Health Network, Toronto, Ontario, Canada; 10) Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 11) The Dalglish Family Hearts and Minds Clinic for 22q11. 2 Deletion Syndrome, Toronto General Hospital, University Health Network, Toronto, Ontario, Canada.

Chromosome 22q11. 2 microdeletions impart a high but incomplete risk for schizophrenia. Possible mechanisms include effects of *DGCR8* haploinsufficiency on genome-wide microRNA biogenesis. We obtained high quality whole-genome sequencing data for individuals with 22q11. 2 deletions and extreme phenotypes (schizophrenia or no psychotic disorder at age >50 years). We investigated rare variants in coding and non-coding sequence most likely to have deleterious consequences. The schizophrenia group had a greater burden of damaging coding variants in neurofunctional gene-sets, including genes involved in neuron projection (nominal p=0. 02, joint burden of three variant types). Variants in the intact 22q11. 2 region were not major contributors. Restricting to genes affected by a *DGCR8* mechanism tended to amplify between-group differences and effect sizes. Damaging variants in highly conserved long intergenic non-coding RNA genes were also enriched in the schizophrenia group (nominal p=0. 04). These initial findings support a multiple within-individual rare variant hypothesis for schizophrenia. Here, the 22q11. 2 deletion appears to act as a threshold-lowering first hit to genome-wide deleterious rare variants in coding and non-coding sequence. A functionally and mechanistically-based design using a genetic model of schizophrenia provides a promising approach for using whole-genome sequencing technology to further delineate the genetic architecture of this complex disorder.

761W

A whole exome study identifies novel candidate genes for vertebral bone marrow signal changes (Modic changes). *M. Kraatari[1,2,3], S. Skarp[1,2,4], J. Karppinen[1,3,5], M. Männikkö[1,2,4].* 1) Center for Life Course Epidemiology and Systems Medicine, Faculty of Medicine, University of Oulu, Finland; 2) Center for Cell - Matrix Research, Faculty of Biochemistry and Molecular Medicine, University of Oulu, Finland; 3) Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Finland; 4) Biocenter Oulu, University of Oulu, Finland; 5) Finnish Institute of Occupational Health, Health and Work Ability, and the Disability Prevention Center, Oulu, Finland.

Lumbar disc degeneration (LDD) is one of the contributing factors behind low back pain (LBP). The most prominent change in LDD is the loss of proteoglycans in the lumbar disc. Modic changes (MC) are a specific phenotype of LDD and are visualized as bone marrow signal intensity changes on magnetic resonance imaging (MRI). MC are a heritable trait with a strong association with LBP. The heritability of MC is estimated around 30%. Two families were studied to identify predisposing variants for MC. Nine individuals were chosen for whole exome sequencing. We focused on rare (MAF < 0. 01) and private variants with harmful *in silico* predictions and variants located in regulatory regions. To find variants co-segregating with MC the identified variants were genotyped from additional family members. One rare variant co-segregated with the MC in each family. In the Family I the observed variant was an insertion and deletion mutation in the *HSPG2* gene, resulting in a premature stop codon. *HSPG2* encodes a heparin sulfate proteoglycan called perlecan, which is a structural protein expressed in mammalian cartilage and basement membranes. Rare autosomal recessive disorders with osteochondrodysplasia are caused by mutations in the *HSPG2* gene. In the Family II a single nucleotide polymorphism in the *MAML1* gene was identified in all affected family members. *MAML1* has been reported to affect the activity of RUNX2, a transcription factor essential in the osteoblast differentiation. RUNX2 has been reported to be highly expressed in degenerated discs. MAML1 could affect the disc structure via RUNX2 activity. We identified two promising candidate genes for MC, *HSPG2* and *MAML1*. Our findings are novel in lumbar spine degenerative phenotypes.

762T

Low-frequency germline variants across 6p22. 2-6p21. 33 are associated with non-obstructive azoospermia in Han Chinese men. *Y. Jiang[1], Y. Wen[2].* 1) State Key Laboratory of Reproductive Medicine, Nanjing Medical University, Nanjing; 2) Department of Epidemiology and Biostatistics and Key Laboratory of Modern Toxicology of Ministry of Education, School of Public Health, Nanjing Medical University, Nanjing.

Genome-wide association studies (GWAS) have identified several common loci contributing to non-obstructive azoospermia (NOA); however, a substantial fraction of NOA heritability remains undefined, especially those low-frequency (defined here as having minor allele frequency (MAF) between 0. 5-5%) and rare (MAF below 0. 5%) variants. Here, we performed a 3-stage exome-wide association study in Han Chinese men to evaluate the role of low-frequency or rare germline variants in NOA development. The discovery stage included 962 NOA cases and 1,348 healthy male controls genotyped by exome chips and was followed by a 2-stage replication with additional 2,168 cases and 5,248 controls. We identified three low-frequency variants located at 6p22. 2 (rs2298090 in *HIST1H1E* encoding p. Lys152Arg: OR=0. 30, *P*=2. 40×10-16) and 6p21. 33 (rs200847762 in *FKBPL* encoding p. Pro137Leu: OR=0. 11, *P*=3. 77×10-16; rs11754464 in *MSH5*, OR=1. 78, *P*=3. 71×10-7) associated with NOA risk after Bonferroni correction. In summary, we report an instance of newly identified signals for NOA risk in genes previously undetected through GWAS on 6p22. 2-6p21. 33 in a Chinese population, and highlight the role of low-frequency variants with large effect in the process of spermatogenesis.

763F

Exome sequencing to detect rare variants associated with biliary atresia. *O. Migita[1,2], A. Matsui[2,3], H. Yamamoto[1], Y. Matsubara[2], K. Hata[2].* 1) St. Marianna University School of Medicine, Kawasaki, Kanagawa, Japan; 2) National Research Institute for Child Health and Development, Tokyo, Japan; 3) College of Nursing St. Luke's International University, Tokyo, Japan.

Biliary atresia (BA) is characterized by complete inability to excrete bile from the liver to the duodenum due to sclerosing inflammation of the extra- and intrahepatic bile ducts. It affects 0. 5 (Caucasian), 1. 0 (Japanese) and 2. 9 (French-Polynesian) in 10,000 live-born infants, respectively but is one of the most life-threatening hepatobiliary disorders in childhood. It may start in fetal, in perinatal or early in postnatal life. What causes this inflammation remains unclear in spite of extensive investigation. It is assumed that multiple genetic and environmental factors might affect individuals with BA. Family members, especially siblings, share genetic and environmental backgrounds. Twin studies remain a favorable means to find genetic factors. Identical twins, developed from one single cell, share the same genome and environment. To address genetic factors, we performed whole exome sequencing of BA patients and their family members including identical twins. We collected the DNA from six families with BA which had a non-affected twin sibling in their family. Using exome sequencing to find rare variants which could not or could rarely be found in control data, we made a list of non-synonymous variants in the families. Our exome sequencing identified over 100 candidate genes with biliary atresia. We could not find deleterious variants in previously reported candidate genes such as ADD3, ABCB4, JAG1 and NOTCH2. It is still unknown what affective nucleotide variants are included in the variants list. To our knowledge, this is the first exome sequencing study of families with BA which might suggest new candidate genes for BA.

764W

Rare variants in sporadic Hirschsprung disease patients. *C. Tang[1], M. So[2], S. Cherny[1,3], P. Sham[1,3], P. Tam[2], M. Garcia-Barcelo[2].* 1) Department of Psychiatry, The University of Hong Kong, Pokfulam, Hong Kong; 2) Department of Surgery, The University of Hong Kong, Pokfulam, Hong Kong; 3) Centre for Genomic Sciences, The University of Hong Kong, Pokfulam, Hong Kong.

Sporadic Hirschsprung disease (HSCR), accounting for ~80% of the patients, represents the most common form of the disorder and is believed to be genetically complex. Studies of rare mutations have discovered more than 10 genes associated with HSCR. The major HSCR-susceptibility gene, *RET*, have both rare coding sequence mutations and common regulatory variants contributing to the disease. The differential contributions of rare and common, coding and noncoding variants tend to vary in accordance with gender and length of aganglionosis. In view of this, we performed a whole genome sequencing study of 5 trios of sporadic patients with the rarer subtype (long segment HSCR) to identify rare mutations causal to HSCR. Our data show that de novo as well as inherited variants contribute to the development of ENS and thereby to HSCR. The discovery might give rise to a new insight into the disease pathophysiology.

.

765T

Gene variants associated with spherical equivalent in the Beaver Dam Eye Study. *K. E. Lee[1], F Chen[2], P Duggal[2], B. E. K. Klein[1], R Klein[1], A. P. Klein[2,3,4].* 1) Ophthalmology & Visual Sci, Univ of Wisconsin School of Medicine and Public Health, Madison, Wisconsin; 2) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland; 3) Department of Oncology, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, Maryland; 4) Department of Pathology, Johns Hopkins School of Medicine, Baltimore, Maryland.

Uncorrected refractive errors are the leading cause of visual impairment and blindness worldwide. Refractive errors are associated with a number of ocular disorders, further increasing risk for vision loss later in life. The etiology of refractive errors is complex with both environmental and genetic components reported in many studies worldwide. As the rates of myopia have been increasing worldwide, understanding the underlying causes are important. We conducted an exome array analysis in 1632 individuals from the Beaver Dam Eye Study to identify rare and low frequency variants that influence spherical equivalent. The Beaver Dam Eye Study is a population-based study of eye traits in older Caucasian adults (ages 43-86 years). Genotyping was conducted using the Illumina exome array. We analyzed 35,095 single nucleotide variants and 11,644 autosomal genes across the genome. After adjustment for age, gender and years of education, we identified two novel variants in the TCTE1 gene region at 6p21. 1 (rs2297336, MAF = 14. 2%, b = -0. 67, P = 3. 5 × 10-7; rs324146, MAF = 17. 2%, b = -0. 59, P = 1. 7 × 10-6), and with the *FSCB* gene at 14q21. 1 (P = 2. 1 × 10-8). We successfully replicated the association on rs634990 near *GJD2* at 15q14 (MAF = 47%, b = -0. 34, P = 2. 1 × 10-4), and discovered a novel association with spherical equivalent on rs1550094 (MAF = 31%, b = -0. 27, P = 8. 0 × 10-3), a locus in *PRSS56* at 2q37. 1 that was previously reported for myopia. Novel genetic variants and genes with multiple rare and low frequency variants may play a role in the control of spherical equivalent. Our results contribute to the increasing evidence of *GJD2* and *PRSS56* in the development of refractive errors. The implication of these two genes, as the common genetic factor for both spherical equivalent and myopia may provide new insights to the underlying mechanism leading to myopia and consequent vision loss.

766F

Testing the neuronal carnitine hypothesis for autism using whole exome sequencing in SSC families. *D. Zhang[1,3], J. Ge[1], N. Krumm[2], T. Turner[2], E. Eichler[2], A. L. Beaudet[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX. USA; 2) Dept. Genome Sciences, U. Washington, Seattle, WA, USA; 3) Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China.

Autism is a highly heritable neurodevelopmental disorder with an incidence of 1 in 68 infants in one more recent CDC report. Although there is strong evidence that rare CNV and point mutation variants, often *de novo*, contribute substantially to the causes of autism at the severe end of the spectrum, much less is known about the etiology at the very mild end of the phenotypic spectrum. We have hypothesized that neuronal/synaptic carnitine deficiency may contribute to the etiology of autism especially in males with a normal physical examination and normal structural brain imaging. In this hypothesis, common variants or carrier states in carnitine-related genes could be risk factors for autism. In this study, we investigated the mutation profiles in 29 candidate genes related to carnitine metabolism based on whole exome sequencing data of the Simons Simplex Collection (SSC) families as carried out and made available by the laboratories E. Eichler, M. State, and M. Wigler. We focused on transmission of all variants in these genes among probands, unaffected siblings, and their parents. Although most variants were equally transmitted in probands and unaffected siblings, some candidate variants showed higher frequencies in probands. For instance, the frequency of rs180830030 in *CPT2* in probands was almost 2 folds higher than that in unaffected siblings (p<0. 001). Furthermore, in 6 quad families transmitting this variant, the frequency was higher in probands (6/6) than in siblings (2/6) (p=0. 03, Fisher Exact Test). Next, we explored the variants at gene level. For some genes, pathogenic variant frequencies were higher in SSC families than those in the Exome Aggregation Consortium (ExAC) database. For example, *SLC22A5*, a gene where homozygous or compound heterozygous mutations cause systemic primary carnitine deficiency, had 11 variants, classified as pathogenic but also reported in ExAC database. Over 1% of individuals in SSC families carry pathogenic variants, but at most 0. 5% of individuals in ExAC database have pathogenic variants (p<0. 001). These findings could be compatible with a hypothesis that genotypes which deplete whole body carnitine or impair carnitine metabolism in mitochondria could be a risk factor for autism. We are examining the prediction of our hypothesis that these trends will be stronger in male probands with a normal physical exam and brain imaging and absence of a high penetrance autism mutation compared to all other probands.

767W

Distinct genetic variants in the vitamin D pathway contribute to risk of multiple sclerosis and vary by presence of *HLA-DRB1*15:01*: Results from the Kaiser Permanente MS Research Program. *A. Mok[1,2], X. Shao[2], B. Rhead[2], L. Shen[3], H. Quach[2], A. Bernstein[4], C. Schaefer[3,5], L. F. Barcellos[1,2,3,5].* 1) School of Public Health, Division of Epidemiology, UC Berkeley, Berkeley; 2) Genetic Epidemiology and Genomics Laboratory, UC Berkeley, Berkeley; 3) Kaiser Permanente Division of Research, Oakland, CA; 4) Palm Drive Hospital, Sebastopol, CA; 5) Research Program on Genes, Environment, and Health, Kaiser Permanente, Oakland, CA.

Studies have shown an association between vitamin D serum levels and multiple sclerosis (MS [MIM 126200]). However, the role of genes involved in vitamin D metabolism and of vitamin D response elements (VDREs) in MS has not been extensively examined. We set out to investigate whether genetic variants in 18 vitamin D-related genes and 43 VDREs are associated with MS in non-Hispanic White members of Kaiser Permanente Northern California (1,200 MS cases; 10,000 controls). Gene boundaries were defined as 2 kb upstream/downstream of each gene to capture regulatory regions. VDREs were defined as the vitamin D receptor complex binding sites detected by ChIP-seq in six cellular models (Tuoresmäki, PLoS One 2014). Genotypes were obtained through whole-genome profiling and imputation. Loci failing standard quality control measures and with an imputation info score <0. 8 were removed. Variants in linkage disequilibrium with SNPs identified in MS genome-wide association studies (GWAS) were removed, and SNPinfo was used to select tagging SNPs. In total, we retained 496 SNPs over 17 gene regions and 7 VDREs for analysis. Associations between SNP and case/control status were controlled for sex, year of birth, smoking, body mass index, education, ancestry, and weighted MS genetic risk score for 110 GWAS SNPs order to detect the effect of vitamin D genetics on MS susceptibility independent of well-established risk factors. Analyses were also stratified by *HLA-DRB1*15:01* allele presence/absence, the strongest genetic contributor to MS risk. Preliminary results showed significant disease associations for rs3782905 in *VDR* and rs2892802 in *LRP2* (OR=0. 70 and 0. 61, respectively; FDR q<0. 05) among **15:01* negative subjects, and rs12143452 in *ST6GALNAC3* (OR 0. 36; FDR q<0. 05) among **15:01* positive subjects. These findings further underscore the complex genetic architecture underlying MS susceptibility as effect modification of the association between variants within vitamin D genes and MS risk by the **15:01* allele was observed. A total of 23,409 VDREs have been identified to date, and further analyses are underway to characterize involvement of these VDREs and their downstream targets in MS, and to further understand the biological mechanisms underlying the relationship between vitamin D and MS. This is the first study to examine VDREs outside the *HLA-DRB1* promoter in MS, and the largest study with the most comprehensive multivariable model of vitamin D genetics in MS, to date.

768T

Fine-mapping major histocompatibility complex association in systemic lupus erythematosus identifies a novel association of *HLA-DPB1* in three Asian populations. *JE. Molineros[1], K. Kim[2], C. Sun[1], X. Zhou[3], KH. Chua[4], H. Zhang[3], SC. Bae[2], SK. Nath[1].* 1) Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul 133-792, Republic of Korea; 3) Renal Division, Peking University First Hospital, Peking University Institute of Nephrology, Key Laboratory of Renal Disease, Ministry of Health of China, and Key Laboratory of Chronic Kidney Disease Prevention and Treatment (Peking University), Ministry of Education, Beijing, China; 4) Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia.

Systemic lupus erythematosus (SLE) is a debilitating heritable autoimmune disease, characterized by a broad spectrum of clinical manifestations including pathogenic autoantibody production and multi-organ damage. *HLA-DRB1* in the major histocompatibility complex (MHC) region is a key genetic factor conferring risk of SLE but residual effects outside of *HLA-DRB1* has not been fully understood. To dissect the MHC association, we imputed classical alleles and amino-acid residues of at 8 HLA genes (*HLA-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1, -DRB1*) in 2,496 SLE cases and 7,657 controls from Korean (KR), Han-Chinese (HC) and Malaysian-Chinese (MC) origin using the SNP2HLA software and an ethnicity-matched HLA reference panel. The association was examined by logistic regression for bi-allelic markers and log likelihood ratio test for multi-allelic amino-acid positions, adjusting for principal component and ancestral origin. Consistent with our recent report, the strongest association was localized at *HLA-DRB1* amino-acid position 13 (PKR = 3. 53x10-31; PHC = 7. 86x10-6; PMC = 1. 56x10-9; Pall =1. 41x10-32) and its proxy position 11 across the extended MHC locus, followed by *HLA-DRB1* amino-acid position 26 in a conditional analysis adjusting for the primary-effect positions 11 and 13. We also confirmed the well-established association of *HLA-DRB1**15:01 (Pall=8. 97x10-23, OR=1. 71, 95% CI = [1. 66-1. 76]) along with its tightly correlated *HLA-DQB1**06:02 (Pall=4. 64x10-26, OR=1. 79, 95% CI =[1. 74-1. 85]) in Asian populations. In addition, the second conditional analysis adjusting for the three independent amino-acid positions of *HLA-DRB1* identified a novel association at *HLA-DPB1* amino-acid position 35 (PKR = 9. 85x10-12; PHC = 1. 98x10-1; PMC = 7. 99x10-2; Pall = 2. 50x10-14) in the epitope binding site. In summary, our study keeps emphasizing the strong contribution of MHC class-II molecules as we newly identified a pathogenic role of *HLA-DPB1* in Asians, raising the possibility of additional MHC class-II molecules such as *HLA-DPB1* with complementary or independent roles in the MHC class-II-mediated SLE pathogenesis.

769F

Genetics of sarcoidosis determined by polygenic risk score analysis derived from CD4:CD8 ratio T-lymphocytes genetic variants. *NV. Rivera[1], M. Ronninger[1], M. A. R Ferreira[2], N. G Martin[2], A. Eklund[1], L. Padyukov[3], J. Grunewald[1].* 1) Respiratory Unit, Department of Medicine Solna and, Karolinska Institutet and Karolisnka University Hospital, Stockholm, Stockholm, Sweden; 2) Queensland Institute of Medical Research, Brisbane, Australia; 3) Rheumatology Unit, Department of Medicine Solna and CMM, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden.

INTRODUCTIONThe bronchoalveolar lavage (BAL) cell CD4:CD8 ratio has been suggested to be important for diagnosing sarcoidosis. The accumulation of CD4+ T-lymphocytes in BAL fluid extrapolates on the plausibility of a pathogenic role of T-lymphocytes in the disease. Few genetic stuides of sarcoidosis have been performed, and thus a substantial portion its heritability remains unexplained. Herein, we constructed polygenic scores using GWAS results of CD4:CD8 as to evaluate the impact of associated-alleles and predict entities of sarcoidosis. METHODSBased on summary association results from a gene discovery set of 2,538 individuals from the general population whose CD4:CD8 was derived from CD4+/CD8+ T-lymphocytes1, we calculated polygenetic risk scores in a target sample of 518 sarcoidosis patients, of which 205 were diagnosed with Löfgren's syndrome (LS) and 313 were non- Löfgren's syndrome (non-LS). In the target set, bronchoscopy with BAL was performed in all patients with recent disease onset and a median (sd) age of 39 (12) years and BAL T-lymphocytes were measured. The polygenic scores were calculated in the target sample for different cluster sets (by P-value thresholdand by chromosome) using the score function in PLINK software. The significance of the polygenic score was tested in the likelihood ratio test by using it as a predictor in a logistic regression model adjusted by gender. The difference in the explained variance (Nagelkerke R2) between the null and alternative modelwas considered as the variance explained by the genetic score. RESULTSCross-prediction analysis for both LS and non-LS sarcoidosis phenotypes showed distinctive genetic architecture based on CD4:CD8 ratio associated-SNPs. For LS, polygenetic scores explained maximum variations of 7. 55% (*P* = 1. 93 x 10-24) at P-discovery and 7. 71% (P = 7. 5x10-25) by chromosome 6. For non-LS, maximum variations of 0. 42% (P = 2x10-2) at P-discovery and0. 66% (P = 9. 65x10-3) by chromosome 10 were observed. CONCLUSIONSOur study provides significant evidence that genes influencing CD4+ and CD8+ T-lymphocytes are involved in pathogenesis of sarcoidosis phenotypesREFERENCES1. Ferreira, M. A. , Mangino, M. , Brumme, C. J. , Zhao, Z. Z. , Medland, S. E. , Wright, M. J. , Nyholt, D. R. , Gordon, S. , Campbell, M. , McEvoy, B. P. , et al. (2010). Quantitative trait loci for CD4:CD8 lymphocyte ratio are associated with risk of type 1 diabetes and HIV-1 immune control. American journal of human genetics 86, 88-92.

770W

Focused HLA Analysis in Idiopathic Inflammatory Myopathy Identifies Significant Associations of Classical HLA Alleles with Autoantibody and Clinical Subgroups. *S. Rothwell[1], R. G. Cooper[2], I. Lundberg[3], F. W. Miller[4], P. K. Gregersen[5], J. Bowes[1], L. R. Wedderburn[6], N. J. McHugh[7], Z. E. Betteridge[7], J. Vencovsky[8], K. Danko[9], V. Limaye[10], A. Selva-O'Callaghan[11], L. M. Pachman[12], A. M. Reed[13], L. G Rider[4], O. Molberg[14], O. Benveniste[15], P. Mathiesen[16], T. Radstake[17], A. Doria[18], J. De Bleecker[19], M. G. Hanna[20], P. M. Machado[20], W. E. Ollier[21], L. Padyukov[3], T. P. O'Hanlon[4], A. T. Lee[5], H. Chinoy[1], J. A. Lamb[21], Myositis Genetics Consortium (MYOGEN).* 1) Centre for Genetics and Genomics, Arthritis Research UK, University of Manchester, Manchester, United Kingdom; 2) Department of Musculoskeletal Biology, Institute of Ageing and Chronic Disease, University of Liverpool, Liverpool, United Kingdom; 3) Rheumatology Unit, Department of Medicine, Karolinska University Hospital and Karolinska Institutet, Stockholm, Sweden; 4) Environmental Autoimmunity Group, Clinical Research Branch, NIEHS, NIH, Bethesda, MD; 5) The Robert S Boas Center for Genomics and Human Genetics, The Feinstein Institute, Manhasset, NY; 6) Arthritis Research UK Centre for Adolescent Rheumatology, University College London, London, United Kingdom; 7) Department of Pharmacy and Pharmacology, University of Bath, Bath, United Kingdom; 8) Institute of Rheumatology, Charles University, Prague, Czech Republic; 9) Internal Medicine, University of Debrecen, Debrecen, Hungary; 10) Rheumatology Unit, Royal Adelaide Hospital, Adelaide, Australia; 11) Internal Medicine Department, Vall d'Hebron General Hospital, Barcelona, Spain; 12) Ann & Robert H. Lurie Children's Hospital of Chicago, Northwestern University Feinberg School of Medicine, Chicago, IL; 13) Department of Pediatrics, Duke University, Durham, NC; 14) Department of Rheumatology, Oslo University Hospital, Oslo, Norway; 15) Pitié-Salpêtrière Hospital, Paris, France; 16) Paediatric Clinic, Copenhagen University Hospital, Copenhagen, Denmark; 17) Department of Rheumatology and Clinical Immunology, University Medical Center, Utrecht, the Netherlands; 18) Division of Rheumatology, University of Padova, Padova, Italy; 19) Department of Neurology, Ghent University, Ghent, Belgium; 20) MRC Centre for Neuromuscular Diseases, UCL Institute of Neurology, London; 21) Centre for Integrated Genomic Medical Research, University of Manchester, Manchester, UK.

Introduction: The most associated region in idiopathic inflammatory myopathy (IIM) is within the major histocompatibility complex (MHC). Due to the complex linkage disequilibrium/haplotype structure in this region, interpretation of causal associations and independent effects using SNPs may be inadequate. **Methods**: We used SNP2HLA to impute classical human leukocyte antigen (HLA) alleles and amino acids from Immunochip genotyping data. HLA imputation was carried out on Caucasian IIM samples from the MYOGEN consortium. Initial analysis was conducted on clinical myositis subgroups; polymyositis (PM, n=931), adult dermatomyositis (DM, n=879) and juvenile dermatomyositis (JDM, n=481). Analysis was then conducted on common serological subgroups (n>40 individuals). Classical 4-digit HLA alleles are preferentially reported, unless an amino acid association explains more risk than HLA alleles alone. **Results**: In IIM, the strongest associations were with alleles of the 8. 1 ancestral haplotype (8. 1 AH). In PM, there were independent effects of *HLA-DRB1**03:01 (p=3. 2x10-82) and *HLA-B**08:01 (p=3. 3x10-9). Outside of the 8. 1 AH, independent effects were seen with amino acid position 57 of *HLA-DQB1* (p=2. 1x10-14) and *HLA-DQB1**04:02 (p=2. 1x10-6) in DM, and with *HLA-C**02:02 (p=2. 5x10-6) in JDM. Antihistidyl–tRNA-synthetase (anti-Jo-1) autoantibodies (n=331) were associated with *HLA-B**08:01 (3. 1x10-89) and *HLA-DRB1**03:01 (1. 3x10-9). Anti-PMScl (n=127) autoantibodies were associated with *HLA-B**08:01 (p=4. 0x10-41). More modest associations with anti-TIF1 autoantibodies and multiple *HLA-DQB1* alleles may be explained by an association with position 57 of *HLA-DQB1* (p=7. 7x10-10). Associations with rarer autoantibodies were replicated, such as *HLA-DRB1**07:01 (p=1. 6x10-11) with anti-Mi-2 autoantibodies (n=99) and *HLA-DRB1**11 (p=4. 2x10-9) with anti-HMG-CoA-reductase (HMGCR) autoantibodies (n=41). **Conclusions**: The strongest associations were with alleles of the 8. 1 AH. Autoantibody subgroups may have unique HLA associations and there is evidence that specific amino acids within the HLA may be functionally important. Associations of clinical subgroups with the HLA may be explained by the strong associations with prevalent autoantibodies. Although limited statistical power may not allow us to tease out all effects within the MHC, this analysis has shown that stratifying patients by serology is important to expand our knowledge of IIM immunogenetics.

771T

A cross-ethnic survey of *CFB* and *SLC44A4,* novel Indian Ulcerative colitis GWAS hits, identifies allelic heterogeneity and underscores their potential role in disease biology. *B. K. Thelma[1], A. Gupta[1], A. Sood[2], G. Juyal[1], V. Midha[2], K. Yamazaki[3], A. Vich Vila[4], M. Kubo[3], R. K. Weersma[4].* 1) Department of Genetics, University of Delhi, South Campus, New Delhi; 2) Department of Gastroenterology, Dayanand Medical College and Hospital, Ludhiana, Punjab, India; 3) Laboratory for Genotyping Development Center for Integrative Medical Sciences, RIKEN 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan; 4) Department of Gastroenterology and Hepatology, University of Groningen and University Medical Centre Groningen, Groningen, The Netherlands.

The first ever genome wide association study (GWAS) of Ulcerative colitis (UC) from north India (NI) that we previously carried out showed strong association of complement factor B (*CFB*) and *SLC44A4* with the disease (Juyal et al. , Gut, 2014). Notably, these genes have been previously missed by several GWA studies in different ethnic backgrounds. Such differences could be attributed to allelic/genetic heterogeneity and distinct environmental components across ethnic groups. Considering the biological relevance of *CFB* and *SLC44A4* in UC, this study aimed to investigate allelic heterogeneity in these two genes in UC cohorts from three genetically diverse populations namely NI, Japanese and Dutch. Comparative linkage disequilibrium profiling across each of these two genes and test of allelic/haplotypic association using high density ImmunoChip case-control genotype data (available as part of the IIBDGC project) were performed in this study. Of the 28 SNPs in *CFB* on the ImmunoChip, single SNP analysis showed similar strength of association for rs4151657, the intronic SNP identified in our GWAS, in NI (p=1. 73x10-10) and Japanese (p=2. 02x10-12) UC cohorts but not in the Dutch. Further, a three marker haplotype namely rs17201431-rs2072634-rs4151657 was shared across NI and Japanese (p <10-8), but a different five marker haplotype namely rs4151651-rs4151652-rs17201431-rs512559-rs537160 was significantly associated (p = 2. 07x10-6) in the Dutch population. Of the 22 SNPs in *SLC44A4*, NI UC GWAS index SNP rs2736428 was found significantly associated in NI (p = 4. 94 x 10-10) and Japanese (p = 3. 37 x 10-9) UC cohorts, but not among the Dutch. This study suggests apparent allelic heterogeneity in *CFB* and genetic heterogeneity in *SLC44A4* across the three different ethnic groups. It also highlights that re-exploration of GWAS findings together with high density genotyping/sequencing and trans-ethnic fine mapping approaches may help identify shared as well as population specific risk variants, which holds promise to explain the missing disease heritability.

772F

Epigenetic marks are strongly enriched in IBD fine-mapped variants. *M. UmićevićMirkov[1], H. Huang[2, 3], M. Fang[4], L. Jostins[5], G. Trynka[1], M. Georges[4], M. Daly[2, 3], J. C. Barrett[1] on behalf of International the IBD Genetics Consortium.* 1) Wellcome Trust Sanger Institute, Hinxton, United Kingdom; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston MA; 3) Broad Institute of MIT and Harvard, Cambridge MA; 4) Unit of Animal Genomics, Groupe Interdisciplinaire de Génoprotéomique Appliquée (GIGA-R) and Faculty of Veterinary Medicine, University of Liège (B34), Liège, Belgium; 5) Wellcome Trust Centre for Human Genetics, University of Oxford, Headington, UK.

Genome wide association studies have been very successful in identifying genetic regions associated with inflammatory bowel disease (IBD) risk, with the most recent GWAS reporting 163 IBD-risk loci. Nevertheless, evidence that the causal variant has been conclusively identified exists for only a small number of these loci. Protein coding variants are the most straightforward to interpret in terms of functional impact, and are responsible for the vast majority of highly penetrant rare disease causing variants identified, yet previous estimates suggest that protein coding variants are causally responsible for less than 20% of GWAS effects. Recent studies have shown that GWAS associated variants are enriched for cell type specific epigenetic marks, offering possible insight into the function of non-coding variants. Here, we integrated predictions of causality for each of the variants in associated regions, generated by a set of Bayesian fine-mapping methods run on Immunochip data in 66,849 individuals, with H3K4me1, H3K4me3 and H3K27ac chromatin mark annotations. We overlaid individual SNP's posterior probabilities of causality from 116 independent associations on chromatin mark annotations in 120 adult and fetal tissues, assayed by the NIH Roadmap Epigenomics Mapping Consortium. We summed posterior probabilities across all chromatin mark peaks in each of the 120 cell types. Expected values and enrichment p-values were calculated by shifting chromatin marks randomly over the fine-mapped regions and calculating sum of posterior probabilities for each permutation. We observed significant enrichment of H3K4me1 marks (p-value<1E-4) in several immune cell types and for H3K27ac in three GI samples (sigmoid colon and colonic and rectal mucosa). Further, we defined a set of "core immune peaks" for H3K4me1 and "core intestinal peaks" for H3K27ac as the intersection of peaks in cells that showed the strongest enrichment. For each of these sets we also created a set of "control" peaks made up of intersecting peaks from fine-mapped region in non-immune and non-intestinal cell types. We observed highly significant enrichment of both "core immune" H3K4me1 and "core intestinal" H3K27ac peaks in comparison to their respective controls. Our results suggest that large portion of IBD risk variants may ultimately be explained by functional variation in cell type-specific regulatory elements.

773W

Dissection of *HLA* class II in Japanese primary biliary cirrhosis: epistasis of protective *HLA-DQ* allele and additional contribution of *HLA-DP* allele. *M. Yasunami[1], H. Nakamura[1,2], K. Tokunaga[3], M. Kawashima[3], N. Nishida[3,4], Y. Hitomi[3], M. Nakamura[2,5,6].* 1) Institute of Tropical Medicine, Nagasaki University, Nagasaki, Japan; 2) Clinical Research Center, National Hospital Organization Nagasaki Medical Center, Omura, Japan; 3) Department of Human Genetics, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan; 4) Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Japan; 5) Department of Hepatology, Nagasaki University Graduate School of Biomedical Sciences, Omura, Japan; 6) Headquarters of PBC Research in the NHO Study Group for Liver Disease in Japan (NHOSLJ), Omura, Japan.

Recent genome-wide association studies (GWAS) of primary biliary cirrhosis (PBC) have consistently identified significant association signals with the markers in the *HLA* region on the short arm of chromosome 6 at the highest levels of significance. In the present study, the contribution of chromosomes harboring certain HLA alleles in the GWAS SNP signals was evaluated by the determination of the alleles of seven classical *HLA* loci (*HLA-A*, *-B*, *-DRB1*, *-DQA1*, *-DQB1*, *-DPA1* and *-DPB1*) in 1200 Japanese patients with PBC and 1196 controls who underwent GWAS, for the better understanding of*HLA*-associated genetic risk of PBC. When the dominant model of either predominant or less frequent allele of 4103 SNPs in the *HLA* region was applied, 305 SNPs showed significant signals with p values less than 5 x 10-8, and a SNP (rs9268644) near the *HLA-DRA* locus gave the minimal p value (p = 5. 64 x 10-24). HLA analyses identified four independent *HLA-DR-DQ* alleles satisfying after Bonferroni's correction for multiple comparisons (p < 5. 75×10-4):*DQB1*06:04* (odds ratio: 0. 19, p = 1. 91×10-22), *DQB1*03:01* (odds ratio: 0. 50, p = 6. 76×10-10), *DRB1*08:03* (odds ratio: 1. 75, p = 1. 01×10-7) and *DQB1*04:01* (odds ratio: 1. 50, p = 9. 20×10-6). Association signals of SNPs in the HLA region were mostly attributable to four major *DR-DQ* alleles, which were demonstrated by the stepwise exclusion of *DR-DQ* factors. In the absence of while four *DR-DQ* factors, certain SNPs in the *DP* loci remained significant, which were mostly explained by the protective effect of *DPB1*02:01* (odds ratio: 0. 50, p = 3. 08×10-7 in the absence of four major *DR-DQ* alleles). Further a non-reciprocal "epistatic" interaction of protective *DQB1*06:04* to the risk conferred by *DRB1*08:03* was demonstrated by subpopulation stratification, implicating the presence of active suppression mechanism for pathogenic autoimmune reactions. In conclusion, genetic association of SNPs in *HLA* region with PBC was explained by the effect small number of *HLA* alleles in Japanese, with a special contribution of protective alleles such as *DQB1*06:04, DQB1*03:01* and *DPB1*02:01*.

774T

Identifying autosomal variants that differ in frequency between males and females and implications for sex-based differences in disease. *S. M. Raj[1], A. L. Williams[2], H. Hakonarson[3], A. G. Clark[1].* 1) Dept. Molecular Biology and Genetics, Cornell University, Ithaca, NY; 2) Dept. Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 3) Center for Applied Genomics, The Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA.

Many disorders, including common, complex diseases and severe autoimmune and metabolic diseases, have different severity in males and females. Some diseases, including systemic lupus erythematosus and multiple sclerosis, are heavily skewed toward affecting females, while others such as autism and attention deficit/hyperactivity disorder show markedly increased incidence in males. Moreover, the associated symptoms of various diseases often differ between males and females. Many non-disease related, sex-differential traits have been associated with polymorphisms on the X chromosome, but relatively few studies have critically examined sex-based autosomal differences on a genome-wide scale. We hypothesize that some of the differences in disease state and etiology between the sexes may result in sex-specific differential survival at autosomal loci that confer sex-specific effects on disease. If these effects are sufficiently strong, they would manifest as male-female differences in SNP and haplotype frequencies. Using genome-wide genotype data on a diverse collection of 58,000 individuals (primarily European Americans and African Americans) as well as data on 10,000 Europeans from the Wellcome Trust Case-Control Consortium (WTCCC2), we identified SNPs that significantly differed in frequency between males and females ($Z \geq |4|$), even after correction for multiple testing and population structure. We identified regions of the genome enriched for genes involved in immune function, and some evidence that the haplotypes on which these variants are located may have sex-specific fitness consequences. These results suggest that autosomal loci may also have substantial sex-specific phenotypic consequences.

775F

Targeted sequencing of familial idiopathic scoliosis regions of interest on chromosomes 1, 6, 9 and 16. *C. M. Justice[1], E. Baschal[2], C. I. Wethey[2], K. Swindle[2], N. H. Miller[2], A. F. Wilson[1], NIH Intramural Sequencing Center.* 1) Genometrics Section, Computational and Statistical Genomics Branch, NHGRI, NIH, Baltimore, MD; 2) Department of Orthopedics, University of Colorado Anschutz Medical Campus, Aurora, CO.

Idiopathic scoliosis (IS) is characterized by a lateral curvature of the spine greater than ten degrees documented by radiographic analysis, which presents in the late prepubertal period in otherwise phenotypically normal individuals. Approximately 2 to 3% of the pediatric population has IS. If an individual with IS has a family history of the disease, it is referred to as Familial Idiopathic Scoliosis (FIS). Targeted Next Generation Sequencing (NGS) methods were used to sequence previously identified linkage regions [Miller et al. 2005] on 1p36, 1q31, 9q33, 16p12 and 16q11 in 174 individuals from 20 families with FIS in order to identify variants responsible for FIS in both coding and noncoding regions. Each family included a minimum of 2 affected individuals with an average family size of 7 individuals. A 2 Mb region on 6q24, encompassing *GPR126,* an IS candidate gene previously identified in a genome-wide association study of a large case/control Japanese population and replicated in a Han Chinese and European-ancestry population [Kou et al. 2013], was also sequenced. Standard QC and filtering methods were used to clean the sequence data. Linkage analysis (MERLIN), and association analysis (PLINK DFAM) were performed on 216,229 single nucleotide variants (SNVs). From the tests of intrafamilial association, 93 SNVs were significant at a level of $p < 1 \times 10\text{-}3$, and 5 SNVs were significant at $p < 1 \times 10\text{-}5$ over all targeted regions. Of the five SNVs significant at a critical value of $1 \times 10\text{-}5$, all were in non-coding regions and four were located within 600 Kb of the *GPR126* locus. Rare variants (MAF < 0. 01) located in exonic regions were prioritized based on: 1) location of the variant in a linkage region and/or in linkage disequilibrium (LD) with associated region, and 2) segregation of the missense variant with the disease phenotype in one or more families. Rare missense variants were identified in several loci. Of note, three SNVs found in our sample cause missense changes in *GPR126*; however, these missense variants were not present in all individuals with IS within a family, and were not significant for association analysis of the whole sample ($p < 0.1$). These results suggest that functional variation in non-coding regulatory regions on chromosomes 1, 6 and 16 may contribute to the FIS phenotype.

776W

CANDIDATE GENE POLYMORPHISMS AS RISK FACTORS FOR PRIMARY KNEE OSTEOARTHRITIS IN ASIAN INDIANS. *s. Poornima[1], p. subramanyam[2], q. hasan[2].* 1) Genetics, Kamineni Hospitals, Hyderabad, india, India; 2) orthopedics, kamineni Hospitals, hyderabad, India.

Primary Osteoarthritis (OA) also know ‘’Degenerative arthritis” is a progressive and irreversible pathology, which is considered as a part of the ageing process. It affects all the joints of the body, predominantly affecting large joints. Symptoms of OR are pain, swelling stiffness and limitations of joint movements. It is estimated that ~10 % of the world’s population ≥60 years have symptomatic OA. At present therapy for OA is only palliative and includes the use of pain relieving medicines, physical, exercise and joint replacement surgery. The increasing incidence of OA in individuals below the age of 60 years and the fact that several members of a single family are affected suggests a genetic predisposition. Current evidence indicates an important role of gene polymorphisms in the aetiology of complex diseases like diabetes, cardiovascular disease and arthritis. Three candidate gene polymorphisms (i)- C1104T (rs143383) of Growth Differentiation Factor 5 (GDF5) (ii) Oligomeric Component of Golgi Complex (COG5) A2236G (rs4730250) and (iii) C242T (rs4673) of Cytochrome B-245 alpha chain (CYBA) have been selected for evaluation in the present study, which were selected based on published results (Kerkhof et al , 2010; Evangelaou et al, 2011 and Liu et al, 2013). Each of these genes are considered to be associated with re-modeling of bones and cartilage. Primary osteoarthritis patients diagnosed clinically and radiologically were included in the study after approval from institutional ethical committee. A total of 900 PCRs were carried out from 300 cases and age, sex matched controls. All individuals belonged to the Asian Indian population. Results demonstrate that the GDF5 1104T (1. 3929 95% CI 0. 9925-1. 9546 p<0. 05), COG5 2236G (2. 0997, 95% CI 1. 5162-2. 9078, p<0. 0001) and CYBA 242T (7. 1005 95% CI 4. 9547-10. 1755 p<0. 0001) were significantly associated with primay OA in our population similar to what was observed in Caucasian patients. Multi -dimensional reduction (MDR) analysis for these three gene polymorphisms showed a strong synergistic interaction between CYBA with GDF(26. 32%) and COG5 (16. 15%) suggesting that the these polymorphisms can be used as biomarkers to assess the risk of knee OA in family members of Asian Indian patients, who can be counseled for a preventative life style.

777T

Identification of genetic variants in Wnt-1-induced secreted protein 1 gene associated with bone mineral density in Old Order Amish. *X. Wang[1, 2], L. Yerges-Armstrong[1], Z. L. Deng[2], J. Perry[1], C. Hong[1], A. Parihar[1], H. Wang[1], Y. B. Zhu[1], Z. Y. Hu[1], E. A. Streeten[1], A. R. Shuldiner[1], B. D. Mitchell[1], M. Fu[1].* 1) Division of Endocrinology, Diabetes and Nutrition, University of Maryland School, baltimore, MD 21201, USA; 2) Department of Orthopaedic Surgery, Second Affiliated Hospital of Chongqing Medical University, Chongqing 400010, China.

Wnt-1-induced secreted protein 1 (WISP1) is a novel target of the Wnt pathway for modulating osteogenesis and improving bone strength. WISP1 is highly expressed in osteoblasts and perichondral mesenchyme and is up-regulated during fracture healing. Compared to wild-type mice, WISP1 knockout mice have lower bone mineral density (BMD) and transgenic mice with overexpressed human WISP1 have increased BMD. The aim of this study was thus to investigate the role of genetic variation in WISP1 as a determinant of BMD and osteoporosis-related traits in 1398 Old Order Amish (OOA) individuals. The OOA are a relatively homogenous population in terms of both genetic ancestry and lifestyle characteristics. We genotyped single nucleotide polymorphisms (SNPs) in the WISP1 gene coding regions using the Illumina Human Exome BeadChip and identified 3 missense SNPs passing quality control thresholds (rs35513885, rs72731540 andrs3739261) and having a minor allele frequency >7% in the OOA. The T allele of rs35513885 (A118S) was significantly associated with higher BMD at the total hip (P=2. 07×10-4, P-value with adjustments for age, age2and gender), and hip subregions including the intertrochanter (P=3. 19×10-4) trochanter (P=3. 96×10-4), and femoral neck (P=3. 93×10-4) as well as lower BMD in the lumbar spine (P=4. 91×10-3). We also found nominally significant association between the A allele of rs72731540 (V184I) and lower lumbar spine BMD (P=0. 043). These 2 SNPs were in high LD (D’=0. 91). In conclusion, we have demonstrated that genetic variants in the WISP1 gene are significantly associated with BMD at multiple skeletal sites in the OOA. These results in humans support a role for the WISP1 gene on influencing variation in BMD. Studies further characterizing variation in this gene and replication of this result in additional cohorts are warranted.

778F

A regulatory variant in the dopamine β-hydroxylase (DBH) gene is associated with nicotine dependence, smoking cessation, and pulmonary function. *D. B. Hancock[1], T. E. Thorgeirsson[2], N. C. Gaddis[1], G. W. Reginsson[2], N. L. Saccone[3], S. M. Lutz[4], M. S. Artigas[5], S. Steinberg[2], C. Glasheen[1], F. Gu[6], M. T. Landi[6], T. B. Baker[7], I. P. Hall[8], M. D. Tobin[5], S. J. London[9], N. E. Caporaso[6], K. Stefansson[2], J. E. Hokanson[4], L. J. Bierut[3], E. O. Johnson[1], CHARGE Consortium Pulmonary Working Group and SpiroMeta Consortium.* 1) RTI International, Research Triangle Park, NC, USA; 2) deCODE Genetics/Amgen, Reykjavik, Iceland; 3) Washington University in St. Louis, St. Louis, Missouri, USA; 4) University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA; 5) University of Leicester, Leicester, United Kingdom; 6) National Cancer Institute, Bethesda, Maryland, USA; 7) University of Wisconsin, Madison, Wisconsin, USA; 8) University of Nottingham, Nottingham, United Kingdom; 9) National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina, USA.

Nicotine dependence is a heritable trait that is one of the strongest predictors of smoking cessation. Our study evaluated the association between nicotine dependence and the only genome-wide significant SNP association reported for smoking cessation, rs3025343 near the dopamine β-hydroxylase (DBH) gene (TAG Consortium *Nat Genet* 2010). In a meta-analysis of 20,053 European ancestry participants from 8 independent samples with nicotine dependence defined by the Fagerström Test for Nicotine Dependence, we found that the rs3025343-G allele (frequency=12%) was associated with decreased nicotine dependence risk (*P*=1. 0×10-4, odds ratio [95% confidence interval] = 0. 92 [0. 88-0. 96] for severe vs. mild dependence). Next, using the Brain expression quantitative trait loci (eQTL) Almanac dataset comprised of post-mortem tissue of 10 different brain regions from 134 European ancestry participants, we found that the rs3025343-G allele was associated with reduced *DBH* expression in the dopamine-rich substantia nigra (P=7. 6×10-4), a highly relevant brain region given that DBH functions to catalyze the conversion of dopamine to norepinephrine in the synaptic vesicles of noradrenergic neurons. We next extended the rs3025343 association testing to pulmonary function measures, which are used clinically to diagnose and follow the progression of lung disease and are adversely impacted by cigarette smoking. In meta-analyses of more than 48,000 European ancestry participants from 19 or more samples, the G allele of rs3025343 was positively associated with the ratio of forced expiratory volume in one second to forced vital capacity, accounting for ever smoking history (*P*=2. 5×10-3, Soler Artigas et al. *Nat Genet* 2011) and further accounting for current smoking history, pack-years of smoking, and SNP-by-smoking interaction terms (lowest *P*=2. 1×10-3, Hancock et al. *PLoS Genet* 2012). The observed directions were all consistent with the rs3025343-G allele being associated with (1) reduced *DBH* expression in substantia nigra, (2) decreased risk of nicotine dependence, (3) increased smoking cessation success as previously reported, and (4) higher pulmonary function. Our findings thus highlight rs3025343 as a potential regulatory variant associated with cigarette smoking and an important smoking-related outcome.

779W

Ultra-deep sequencing to identify somatic genetic variations in the brain in Alzheimer disease and during aging. *H. T. Helgadottir[1], L. Lilius[2,3], A. K. Lindström[2,3], C. Graff[2,3], M. Eriksson[1].* 1) Karolinska Institutet, Department of Biosciences and Nutrition, Center for Innovative Medicine, SE-141 57 Huddinge, Sweden; 2) Karolinska Institutet, Department of NVS, Center for Alzheimer Research, Division for Neurogeriatrics, SE-141 57 Huddinge, Sweden; 3) Dept of Geriatric Medicine, Karolinska University Hospital-Huddinge, SE-141 86 Stockholm, Sweden.

The accumulation of mutations in cells over time is predicted to create increased genetic mosaicism in the tissue that may contribute to the declined homeostasis seen with aging. Alzheimer disease (AD; MIM 104300) is a neurological disorder that mainly affects elderly people and causes cognitive dysfunction. The main neuropathological findings are amyloid plaques, neurofibrillary tangles, loss of neurons, and decreased brain volume. However, these findings can also be found in the brains of older unaffected individuals and suggest that neurodegeneration is a more common effect of aging than previously anticipated. AD has been classified into early-onset and late-onset, and although early-onset AD may show autosomal transmission, known mutations account for less than 5% of all AD cases. Genome-wide association studies have revealed several loci conferring risk to AD, with the *APOE* gene (MIM 107741) being the main genetic risk factor. In this study we have used ultra-deep sequencing, aiming for >600X average depth, to investigate somatic mutations in the temporal cortex from AD patients in addition to non-AD controls. Samples included both early-onset AD (45-50 years of age) and late-onset AD (70-90 years of age) as well as brain samples from controls, matched for APOE, gender and age. In addition, DNA samples from blood from all the individuals were included in the study. Custom made arrays from Roche-Nimblegen (SeqCap EZ Choice Library) capturing 2. 85 Mb of the genome was used for sequencing. Our analysis focuses on rare somatic mutations and we hope to identify low-degree somatic mutations that might contribute to aging. We expect the mutations to be more common in older individuals compared to younger and to be able to make an estimate of the prevalence of mosaicism in the brain by comparing different tissues and ages. So far, only few studies have investigated genetic variation in the brain and this will provide an opportunity to detect genetic variation in aging and/or AD, and its contribution to the declined tissue homeostasis.

780T

Genetic Susceptibility for Asthma in Two Resource-Limited Settings in Peru. *W. Checkley[1,2], C. Vergara[5], C. Robinson[1], L. Baumann[1], K. Romero[3], R. Gilman[2], R. Wise[1], K. Barnes[6], N. Hansel[1,4], PURA Study Investigators.* 1) Division of Pulmonary and Critical Care, School of Medicine, Johns Hopkins University, Baltimore, MD; 2) Program in Global Disease Epidemiology and Control, Department of International Health,Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 3) Biomedical Research Unit, A. B. PRISMA, Lima, Peru; 4) Department of Environmental Health Sciences, Bloomberg School of Public Health, Johns Hopkins University, Baltimore,MD; 5) Division of Infectious Diseases, School of Medicine, Johns Hopkins University, Baltimore, MD; 6) Division of Allergy and Clinical Immunology, School of Medicine, Johns Hopkins University, Baltimore, MD.

Genetic and environmental exposures related to urbanization are known to be associated with childhood asthma. Peru has the highest rates of asthma in Latin America and one of the highest worldwide. A higher prevalence and severity of asthma is observed in urbanized environments of Peru compared to rural areas. We sought to determine whether genetic susceptibility contributed to asthma and high total IgE (tIgE) levels in two regions of Peru with disparate degrees of urbanization. We analyzed 620 children from Lima (79 asthmatics, 541 non-asthmatics) and 651 children form Tumbes (20 asthmatics, 631 non-asthmatics) with similar age and gender distributions. Total IgE levels were determined in serum by ELISA. 78 SNPs in the genes ADAM33, CD14, GSTM1, GSTP1, IL4, IL13, IL1RL1, TSLP and IL33 were genotyped using the VeraCode technology. The SNPs were previously associated with asthma in GWAS. Population stratification was corrected using 384 ancestry informative markers. Comparisons of allelic frequencies were performed between asthmatics and non-asthmatics in each site and between non-asthmatics from Lima and Tumbes. A P value<0. 05 was considered as significant. We found that SNPs in IL33 were significantly associated with asthma in Lima (rs10975514, rs10975516, rs1332290) and Tumbes (rs16924243, rs12000491). One SNP in GSTM1 (rs2239892) was significantly associated with asthma in Lima and, polymorphisms in ADAM33 (rs598418, rs487377) and ILR1 (rs10204137) showed a significant association in the case-control analysis in Tumbes. Three markers in IL1RL1 (rs2160203, rs10204137, rs10206753) were associated with high tIgE levels in Lima. Rs11905870 (ADAM33) and rs11466750 (TSLP) were associated with tIgE levels in Tumbes. There was a significant difference in the allelic frequency of SNPs in IL13 (rs1295687), IL33 (rs1412420 and rs12000491), ADAM33 (rs3918395, rs487377), IL4 (rs2243252, rs2243253), ILR1 (rs10204137) and TSLP (rs10062929) between non-asthmatics from Lima and Tumbes after correcting with two PCAs. SNPs in candidate genes were identified as risk factors for asthma and high IgE levels within the populations of Lima and Tumbes. Those polymorphisms also showed a significantly different distribution in children without asthma between both sites. Further studies are needed to elucidate if the differences in the genetic component play a role in the observed differences in the prevalence of asthma between an urban and rural environment.

781F

***SLC2A9* sequence variants and serum uric acid concentrations in American Indians: The Strong Heart Family Study.** *G. Chittoor[1], K. Haack[2], S. Laston[3], LG. Best[4], JW. MacCluer[2], JG. Umans[5], SA. Cole[2], VS. Voruganti[1].* 1) Nutrition and Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis, NC; 2) Genetics, Texas Biomedical Research Institute, San Antonio, TX; 3) South Texas Diabetes and Obesity Institute and Regional Academic Center, School of Medicine, University of Texas Health Science Center at San Antonio/University of Texas Rio Grande Valley, Brownsville, TX; 4) Missouri Breaks Industries Research Inc. , Timber Lake, SD; 5) Medstar Health Research Institute, Hyattsville, MD.

Background: Increased serum uric acid (SUA) is a risk factor for gout, renal and cardiovascular disease. The variation in SUA concentrations is under significant genetic influence. Various population studies have shown strong association of SUA concentrations with solute carrier family 2, member 9 (*SLC2A9*), a uric acid transporter gene. Our objective was to identify *SLC2A9* sequence variants and analyze their association with SUA concentrations in a family-based study of American Indians, the Strong Heart Family Study. **Methods:** We sequenced 96 kb of the *SLC2A9* gene using Illumina's TruSeq Custom Amplicon and MiSeq Sequencer in 1118 founders of multigenerational families. Sequence data were aligned to the Human Genome Reference Sequence version 37. 1 (hg19). Variants were called, recalibrated and QC'd using the Genome Analysis Toolkits (GATK v. 3. 3) Haplotype Caller. Of the 427 variants, 233 were single nucleotide polymorphisms (SNPs), 125 were single nucleotide variants (SNVs), 26 were singletons, and 43 were indels/triallelic. Based on comparison with dbSNP database, 117 SNPs were novel. The minor allele frequencies (MAFs) of all variants ranged between 0. 1% and 49%. The association of SNPs with SUA was estimated using a measured genotype analysis accounting for family relationships in SOLAR. The appropriate significance level was determined to be $p < 3 \times 10\text{-}4$ after correcting for multiple tests. **Results:** Strong association was observed between SUA concentrations and *SLC2A9* SNPs ($p < 6 \times 10\text{-}12$). A total of 96 SNPs were associated at the significance level of < 3 x10-4; of which, for example, rs3775946 (A), rs7663032 (C), rs6826764 (G), rs7696983 (A), rs7669090 (C) have not been previously associated with SUA concentrations, and one SNP (chr_pos: 4_10027969) was found to be novel (MAF between 26 and 42%; the effect sizes ranged from 3. 6 to 4. 8%). Our results also showed that the minor alleles of these SNPs were associated with lower SUA concentrations. None of the rare variants were significantly associated with SUA concentrations. **Conclusions:** Our results indicate the importance of common sequence variants in *SLC2A9* in the regulation of SUA concentrations and demonstrate that rare variants of potentially functional effect remain to be discovered in American Indians.

782W

Association of long intergenic non-coding RNA SNP near *MYLIP* (6p23-p22) with Healthy Aging Index: The Long Life Family Study (LLFS). *M. F. Feitosa[1], R. L. Minster[2], M. K. Wojczynski[1], A. M. Matteini[3], R. Mayeux[4], N. Schupf[5], T. T. Perls[6], K. Christensen[7,8], A. B. Newman[9], M. A. Province[1].* 1) Department of Genetics, Washington University School of Medicine, St. Louis, MO; 2) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, PA; 3) Division of Geriatric Medicine and Gerontology, School of Medicine, Johns Hopkins University, Baltimore, MD; 4) Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, NY; 5) Taub Institute, College of Physicians and Surgeons, Columbia University, New York, NY; 6) Section of Geriatrics, Department of Medicine, Boston University, Boston School of Medicine and Boston Medical Center, MA; 7) The Danish Aging Research Center, Epidemiology, University of Southern Denmark; 8) Department of Clinical Genetics and Department of Clinical Biochemistry and Pharmacology, Odense University Hospital, Odense, Denmark; 9) Department of Epidemiology Graduate School of Public Health, University of Pittsburgh, PA.

Elevated level of low-density lipoprotein (LDL) cholesterol is associated with cognitive decline and dementia, and with systolic blood pressure (SBP) in diabetic patients. These diseases may lead to an increased risk of death. We developed a mortality-optimized healthy aging index (HAI-M) that captures these multiple systems and predicts mortality. SBP, pulmonary vital capacity (FVC), modified-mini-mental-status-examination-score (MMSE), creatinine, and glucose were used to generate HAI-M. We hypothesize that SNPs (N = 30 at p < 5E-08) found to be genome-wide associated with LDL in the Global Lipids Genetics Consortium will contribute to HAI-M variability. Association analysis was performed using family mixed model in 3,534 European descent subjects. We found evidence of association between rs3757354 with HAI-M in men (p = 4. 99E-4, beta = -0. 15 ± 0. 04), but not in women (p = 0. 18). Further, we investigated if rs3757354 was associated with each HAI-M component. Rs3757354 was also significantly associated with FVC in men (p = 1. 40E-3) and with MMSE in all data (p = 4. 12E-3). Rs3757354 is in the exon of a lincRNA and is approximately 2kb upstream of *MYLIP*. It encodes the E3 ubiquitin ligase myosin regulatory light chain-interacting protein, which promotes degradation of the LDLR and participates in the inhibition of neurite outgrowth. In neurons, *MYLIP* induces reelin, which down-regulates *VLDLR*, decreasing VLDL levels. Reelin signaling is also associated with neurological disease. Rs3757354 overlaps several potential regulatory markers: (i) DNase hypersensitivity, (ii) binding motifs for transcriptional factors JUND (protects cells from p53-dependent senescence and apoptosis), and EGR1 (involved in neuropathology), and (iii) enhancer (H3K4Me1). Our findings highlight the biological importance of 6p22-p23 to healthy longevous subjects. Further investigations are needed to determine if rs3757354 regulates *MYLIP* and/or other more remote genes with male-specific effects on HAI-M and FVC but in both sexes for cognitive function.

783T

Variation in kidney structure-related genes in African Americans with type 2 diabetes- associated end-stage kidney disease. *M. Guan[1,3], J. Ma[2,4], J. Keaton[3], M. Stromberg[3], P. Mudgal[3], B. I. Freedman[4], D. W. Bowden[3,5,6], M. C. Y Ng[3,5].* 1) Integrative Physiology and Pharmacology Program, Wake Forest School of Medicine, Winston-Salem, NC, USA; 2) Department of Nephrology, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China; 3) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA; 4) Department of Internal Medicine/Nephrology; Wake Forest School of Medicine, Winston-Salem, NC, USA; 5) Center for Diabetes Research; Wake Forest School of Medicine, Winston-Salem, NC, USA; 6) Department of Biochemistry; Wake Forest School of Medicine, Winston-Salem, NC, USA.

Familial aggregation supports that genetic factors contribute to diabetic and non-diabetic end-stage kidney disease (ESKD) susceptibility in African Americans (AAs). Structural abnormalities in the kidney occur with progression of diabetic kidney disease. The contribution of genetic variants in kidney structure-related genes in AAs with type 2 diabetes (T2D) associated-ESKD has not been systematically evaluated. Therefore, a collection of 49 genes important in kidney structure, including podocyte, glomerular basement membrane (GBM), mesangial cells and matrix, and renal tubule and interstitium was tested. Genotyping of tag SNPs for the selected genes was conducted using an Affymetrix Biobank custom genotyping array. Single variant association analysis was performed for each gene using a linear mixed model based method (MMM) in 2,005 AA T2D-ESKD cases (S1), 1,031 AA non-diabetic non-nephropathy controls (S2), and 630 AA T2D non-nephropathy samples (S3). A total of 19 SNPs from 17 regions demonstrated nominal associations (P<0. 01) with T2D-ESKD (S1 vs. S2) after conditioning on age and gender (Controlling for *APOL1* risk gave similar results). The most significant signal rs201234802 (P<7. 08e-5), a missense variant, was observed at *ATP6V11* (expressed in renal tubules). In addition, 21 SNPs were associated with ESKD (S1 vs. S3) at P<0. 01. Subsequent replication was performed on SNPs with strong evidence of association with T2D-ESKD in an additional 915 AA non-diabetic ESKD cases (S4) and 518 healthy AA controls (S5). 4 SNPs replicated nominal association with ESKD (S4 vs. S5). Meta-analysis combining T2D-ESKD and non-diabetic ESKD (S1+S4 vs. S2+S3+S5) revealed 8 SNPs from 5 loci (*CFH*, *NPHP3*, *ARHGAP24*, *TRPC6*, *COL4A2*) associated with all-cause ESKD (P<0. 05). The top hit was a missense variant rs35485382 (P=0. 0026) located in the *NPHP3* region. Notably, two additional SNPs in the *NPHP3* region (rs78174962, P= 0. 039; rs74504809, P=0. 044) were also associated with all-cause ESKD. These results indicate that variation in genes regulating kidney structure may contribute to the predisposition to T2D-ESKD and non-diabetic ESKD in AAs.

784F

Rs738409 Polymorphism in *PNPLA3* gene is associated with lower insulin resistance in Korean Men. *J. Park.* Family medicine, Seoul National University Hospital, Seoul, Seoul, South Korea.

Background & Aims: The rs738409 polymorphism in adiponutrin is a well-known genetic risk factor for nonalcoholic fatty liver disease (NAFLD) development. When we consider that NAFLD is closely associated with insulin resistance, the association between rs738409 variant and insulin resistance is highly suspected. **Methods**: We screened 1,648 Korean men who visited the Health Promotion Center of Seoul National University in Seoul, South Korea from December 2009 to June 2012 and 1,189 were enrolled in final analysis. Hepatic steatosis was evaluated by abdominal ultrasound and subjects with secondary causes of NAFLD or alcohol consumption of more than 30g/day were excluded from NAFLD classification. Serum glucose and insulin levels were measured using overnight fasting blood sample. The homeostasis model assessment estimated insulin resistance (HOMA-IR) index was calculated by multiplying glucose in mg/dl by insulin in μU/ml/405. The cutoff value of 2. 56 was used for insulin resistance diagnosis as previously reported in Korean men. HOMA-IR index values were natural log-transformed after normality test. Rs738409 genotyping was performed using TaqMan assay on a ViiATM 7 Real-Time PCR System (Life Technologies, California, USA). **Results**: The participants were predominantly middle-aged men (49. 1 ±7. 0 years; range, 30-60 years), and the frequencies of NAFLD and insulin resistance were 43. 7% and 17. 5%. We performed multivariate regression analysis after adjustment for NAFLD status. The rs738409 CG and GG genotype showed lower mean levels of HOMA-IR index compared to the CC genotype (P = 0. 031 and < 0. 001, respectively. Then, we checked the effect of the rs738409 G allele on HOMA-IR index levels in each NAFLD and non-NAFLD subgroup stratified by NAFLD status using one-way ANOVA. In the non-NAFLD group, the CG and GG genotypes compared to the CC genotype showed significantly lower mean levels of HOMA-IR index in an additive manner (Sidak-corrected P = 0. 018 and < . 001, respectively); however, in the NAFLD group, those inverse associations were not shown. **Conclusion**: The adiponutrin polymorphism rs738409 known as a NAFLD genetic risk factor was also associated with decreased levels of insulin resistance assessed by HOMA-IR index, and the inverse association was prominent in subjects without NAFLD.

785W

Genetic Associations between Human Leukocyte Antigen-DQA1 Typing and Gallstone Disease in Han Chinese. *H. Yang[1,2], S. Shih[1,2,3,4], T. Chang[2], H. Wang[1,3], M. Lin[2], Y. Lien[2], Y. Lee[2,4,5,6,7].* 1) Mackay Junior College of Medicine, Nursing and Management, Taipei, Taiwan; 2) Medical Research, Mackay Memorial Hospital, New Taipei City, Taiwan; 3) Division of Gastroenterology, Department of Internal Medicine, Mackay Memorial Hospital, Taipei, Taiwan; 4) Department of Medicine, Mackay Medical College, New Taipei City, Taiwan; 5) Department of Pediatric Endocrinology, Mackay Children's Hospital, Taipei, Taiwan; 6) Department of Pediatrics, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan; 7) Institute of Biomedical Sciences, Mackay Medical College, New Taipei City, Taiwan.

AbstractBackground and Goal: Biliary crystals in gallbladder induce latent and persist inflammatory responses and affect extra-hepatic bile ducts and multiple symptoms called gallstone disease (GSD). The pathology and environmental risk factors of GSD are well documented. However, lack of immune or inflammatory responses of genetic regulation in GSD development. Major histocompatibility complex (MHC) class II, HLA-DQA1 typing is correlated with many gastrointestinal disorders. In this study, we investigated whether typings of the *HLA-DQA1* gene were associated with GSD susceptibility. **Methods:** Genomic DNA was isolated from whole blood samples of 466 patients with GSD and 385 gallstone-free controls. HLA-DQA1 polymorphisms were genotyped using SBT. **Results:** We did not find any association between HLA-DQA1 polymorphisms and gallstone disease.

786T

Genotyping the Nord-Trøndelag Health Study (HUNT) Cohort Using an Exome+GWAS Array with All Stop-gain Variants in 96 Genes. *W. Zhou[1], O. L. Holmen[2,3], S. Vrieze[4], L. G. Fritsche[5], H. Zhang[6], J. Chen[6], M. Boehnke[5], K. Hveem[2,7], C. J. Willer[1,6,8], G. R. Abecasis[5].* 1) Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, Michigan, 48109, United States of America; 2) HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, 7600 Levanger, Norway; 3) St. Olav Hospital, Trondheim University Hospital, Trondheim, Norway; 4) Institute for Behavioral Genetics, University of Colorado, Boulder, CO, 80309, United States of America; 5) Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, Michigan, 48109, United States of America; 6) Department of Internal Medicine, Division of Cardiology, University of Michigan Medical School, Ann Arbor, Michigan, 48109, United States of America; 7) Department of Medicine, Levanger Hospital, Nord-Trøndelag Health Trust, 7600 Levanger, Norway; 8) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan, 48109, United States of America.

The Nord-Trøndelag Health Study (HUNT) in Norway is a population-based health study that has lasted more than three decades (www. ntnu. edu/hunt). DNA is currently available for ~70,300 participants. Previous genetic studies have examined targeted subsets of HUNT study participants to discover or replicate disease and complex trait-association signals, demonstrating the value of genetic analyses of HUNT study samples in uncovering novel biological underpinnings and therapeutic targets for a variety of traits. Typically, we see a high correlation of genetic effect sizes estimated in HUNT and other European samples (r2=0. 92 for lipid and cardiovascular disease associated variants, for example), indicating that association findings can be generalized to other European populations. Currently, we are genotyping ~70,300 HUNT participants using an Exome+GWAS array (HumanCoreExome, Illumina), comprised of >500,000 SNPs. We augmented the array with a catalog of potential stop-gain variants in 96 genes that would result in truncated and incomplete gene products because the effect of these variants on phenotypes may be easy to interpret. These 96 genes targeted for this detailed analysis include genes associated with type 2 diabetes, blood lipid levels, Alzheimer's disease, nicotine/alcohol metabolism, and several genes where mutations have been implicated in serious, but treatable, health conditions. In addition to these, we augmented the array with missense and stop-gain variants identified in previous whole genome sequencing of HUNT study participants, associated with any of >900 complex traits in the NHGRI GWAS catalogue, and with ancestry informative markers. The data resource will be used in genetic studies for a wide range of diseases and complex traits. At the meeting, we will describe initial quality control of the genotype data and comparisons with existing whole genome sequence data.

787F

Coding variants in *RREB1* are associated with type 2 diabetes and fasting glucose levels. *H. Kitajima[1], A. Mahajan[1], M. I. McCarthy[1,2] on behalf of GoT2D and T2D-GENES consortia.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, United Kingdom.

Non-coding variants in *RREB1/SSR1* region have been shown to be associated with type 2 diabetes (T2D, rs9502570) susceptibility and fasting glucose (FG, rs17762454) levels in previous genome wide association studies (GWAS). We recently reported association of a coding variant in *RREB1* (Ser1554Tyr, MAF=21. 1%, *P*=8. 3x10-9) influencing FG levels among non-diabetic subjects. The biological mechanism driving these associations is not known and the relation between these associations driving FG and T2D signals is still unexplored. We sought to assess the impact of coding variation at this GWAS locus on both T2D susceptibility and FG levels in an attempt to highlight causal transcripts and facilitate identification of molecular mechanisms influencing the traits. We combined exome array data with exome sequencing from 92,794 (28,305 T2D cases, 51,549 controls) for testing association with T2D. We tested single variants for association with T2D using a linear mixed model to account for relatedness and combined association summary statistics by meta-analysis. The FG associated coding variant Ser1554Tyr, was associated with T2D at nominal significance (*P*=4. 9x10-5 in effective sample size 82,784). However, another coding variant in *RREB1*, rs9379084, coding for Asp1171Asn, showed a much stronger association with T2D (MAF=11. 1%, *P*=4. 0x10-9, effective sample size 56,339) despite smaller sample size. The non-coding GWAS variants were not present on the array, therefore, we used imputed GWAS data to explore their relationship. The coding variant association at *RREB1* (Asp1171Asn) was conditionally independent of the original common variant GWAS signal rs9502570 for T2D, indicating multiple risk variants at this locus. In addition, based on European haplotypes from the 1000 Genomes Project, Ser1554Tyr was more strongly correlated (*r2*=0. 59) with the T2D GWAS lead variant rs9502570 as compared to the FG GWAS lead rs17762454 (*r2*=0. 08), which might indicate presence of multiple FG association signals in this region. *RREB1,* encodes Ras-responsive element-binding protein 1, a zinc finger transcription factor that binds RAS-responsive elements of gene promoters. These data point to a likely functional role for *RREB1* at this locus, with separate and possibly multiple association signals for T2D, and for FG in non-diabetic individuals, although further studies are needed to confirm this hypothesis.

788W

Identifying Candidate Genes Associated with Non-syndromic Orofacial Cleft Risk in Puerto Rican Hispanics. *C. J. Buxó-Martínez[1], L. M. López-Del Valle[1], J. F. Cordero[2], N. Debs[1], M. I. Salcedo[1], L. García-Fragoso[3], M. Soto[1], K. K. Ryckman[4], N. Mukhopadhyay[5], M. Govil[5], A. Butali[6], M. L. Marazita[5], J. Murray[7].* 1) Office of the Assistant Dean of Research, School of Dental Medicine, University of Puerto Rico, San Juan, PR; 2) Dept. of Human Development, Graduate School of Public Health, University of Puerto Rico, San Juan, PR; 3) Dept. of Pediatrics, School of Medicine, University of Puerto Rico, San Juan, PR; 4) Dept. of Epidemiology, College of Public Health, University of Iowa, Iowa City, IA; 5) Dept. of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 6) Dept. of Oral Pathology Radiology and Medicine, University of Iowa, Iowa City, IA; 7) Dept. of Pediatrics, University of Iowa, Iowa City, IA.

Non-syndromic orofacial clefts, primarily cleft lip with or without cleft palate (nsCL/P) are among the most common birth defects, with an etiology including both genetic and environmental factors. Because there is limited knowledge about the epidemiology and risk factors for nsCL/P in Puerto Ricans, thus the study objective was to identify maternal exposures and genetic characteristics associated with nsCL/P risk in Puerto Rican children. Children aged 0-14 years with nsCL/P and their parents were recruited as case families (n=150, 98 triads with both parents and 52 dyads with 1 parent). Unaffected controls were selected within the same age range (n=160, 73 triads and 87 dyads). Forty-six single nucleotide polymorphisms were selected for this study from candidate genes/loci previously associated with nsCL/P in recent genome-wide association studies (GWAS). Genotyping was done using Taqman and Fluidigm. Transmission disequilibrium test (triads), logistic regression with/without covariates and gene-environment interaction (offspring) analyses were conducted using PLINK. Significant differences were found for maternal exposure covariates: alcohol consumption (p=0. 04) during pregnancy, and folic acid (p=0. 001) and multivitamin (p=0. 04) use 6 months before pregnancy. The TDT results were significant for *ABCA4* rs560426 (p=0. 02) and *CRISPLD2* rs4783099 (p=0. 02). We observed significant differences (p<0. 05) between cases and controls for *PAX7* rs742071, *ABCA4* rs4147811, *IRF6* rs2235371, *DYSF* rs7568834, but *MTHFR* rs1801131 was the only variant that met the Bonferroni significance level of 10-3. There was no evidence of gene-environment interaction between maternal exposure covariates and variants except for *MTHFR* rs1801131 and the peri-conceptional intake of folic acid before pregnancy (p=0. 04). Our findings may support a role for *MTHFR* rs1801131 variant in nsCL/P and provide specific variant allele frequencies and impact in Puerto Rican/Hispanic populations. Interaction with folic acid and multivitamin use and *MTHFR* should be evaluated in larger GWAS in Puerto Ricans since the strength and direction remains uncertain. Supported by the following grants: 1K99DE024571, 3R37DE008559-23S1, 8U54MD007587, R25RR017580, and U54RR026139.

789T

Rare Variants in Previously Identified Linkage Region on Chromosome 7p Associated With Carotid Bifurcation Intima-Media Thickness in Dominican Families. *N. D. Dueker[1], A. Beecham[1], L. Wang[2], S. Blanton[2], C. Dong[2], S. Guo[1], D. Cabral[3], T. Rundek[3], R. L. Sacco[2,3].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL; 3) Department of Neurology, Epidemiology and Public Health, Miller School of Medicine, University of Miami, Miami, FL.

Thrombotic disorders including ischemic stroke (IS) and myocardial infarction (MI) are leading causes of death and disability in the US. Though both IS and MI are suggested to be genetically controlled, only a small proportion of the genetic risk for either has been identified. One method to identify genetic variants influencing IS and MI is to study intermediate phenotypes such as carotid intima-media thickness (cIMT), a subclinical measure of atherosclerosis. We have previously shown cIMT measures to be heritable and found evidence for linkage and association of common variants on 7p with carotid bifurcation IMT (BIF). Therefore, we aimed to further characterize the 7p region, hypothesizing that rare variants in this region are associated with BIF. To test this hypothesis, we sequenced the 1 LOD unit down region on 7p in nine extended families from the Dominican Republic with evidence for linkage to BIF in this region (family specific LOD score > 0. 1). Using these data, we performed the family-based sequence kernel association test (famSKAT) on 68 genes within the 7p region to identify those associated with BIF. Analyses were restricted to single nucleotide variants (SNVs) with minor allele frequency (MAF) <5% and a Bonferroni correction was applied to correct for 68 tests (p<0. 00074 considered significant). In our analysis of exonic rare variants, nucleotide-binding oligomerization domain 1 (*NOD1)* was our most significantly associated gene (p=0. 0009; # SNVs=14). This gene showed nominal evidence for replication in an additional sample of twelve extended families from the Dominican Republic with Exome Array data and family specific LOD score >0. 1 for BIF (p=0. 06; # SNVs=9). *NOD1* is an excellent candidate gene as it is expressed in the heart and atherosclerotic lesions have been found to have increased expression of *NOD1*. Furthermore, a study in mice found that administration of a NOD1 agonist led to fibrosis, cardiomyocyte apoptosis and cardiac dysfunction. In addition to *NOD1*, we also identified 15 moderately-associated genes with p<0. 05, including *AQP1* which encodes aquaporin-1, a protein that is present during vascular development and following vascular injury in rats. *AQP1* also showed nominal evidence for replication in our twelve replication families (p=0. 06; # SNVs=5). Taken together, our study provides suggestive evidence for a role of rare variants within our previously identified 7p region in BIF.

790F

Association of *APOE* E3/E4 With A Convex Facial Profile in Individuals diagnosed with Obstructive Sleep Apnea (OSA). *J. K. Hartsfield[1], L. L. Wachs[2], M. Elmahdy[2], J. Roedig[3], L. A. Morford[1], B. Phillips[4], P. Wardrop[5], J. E. Van Sickels[2], G. T. Kluemper[2].* 1) Center for the Biologic Basis of Oral/Systemic Diseases, Univ Kentucky, Lexington, KY; 2) College of Dentistry, Univ Kentucky, Lexington, KY; 3) Private Practice, Nashville, TN; 4) College of Medicine, Univ Kentucky, Lexington, KY; 5) Saint Joseph Sleep & Wellness Center, Lexington, KY.

Objective: Facial profile convexity has been associated with OSA in some but not all studies. *APOE* E4 has also been associated with obstructive sleep apnea (OSA) in some but not all studies. The purpose of this study was to determine the *APOE* alleles in patients with OSA to study potential relationships between the *APOE* alleles, facial profile (skeletal classification), AHI and/or BMI. We hypothesized that there would be an association between a skeletal Class II (convex) facial profile, AHI and/or BMI with certain *APOE* allele(s). Materials and Methods: One-hundred sixty-eight Caucasian OSA subjects (110 males and 58 females) with an AHI > 15 were recruited from two sleep clinics under IRB approval at the University of Kentucky. Subjects underwent an orthodontic clinical exam and lateral photograph-based profile assessment and measurement to determine their facial (Angle) classification (Class I=straight; Class II=-convex; Class III=concave). Exclusion criteria included edentulism, or previous orthodontic treatment. BMI was calculated from clinical chart information. Genomic DNA was isolated from patient saliva and genotyped by Taqman® methodology with the *APOE* allele-defining SNPs, rs429358 and rs7412. Kruskal–Wallis, and chi-squaretests were used to analyze the data using JMP-PRO-10, with significance at 0. 05. Results: No difference (p=0. 3) was observed in AHI amongst the three facial skeletal classifications. Combining the Class I and III groups into a “non-Class II” vs the Class II group was also not significant for AHI (p=0. 3). Class II subjects had significantly lower (p=0. 004) BMI than skeletal Class I or III classified individuals. Both *APOE* genotypic distributions were in Hardy-Weinberg equilibrium (p= 0. 14). While BMI was not associated with any particular combination of inherited *APOE* alleles (p=0. 4), inheriting the E3/E4 alleles together was statistically significant when comparing the Class II versus “non-Class II” group (p=0. 012). Conclusion: This data is suggestive that BMI is lower in the OSA patients with a Class II facial convexity. *APOE* E3/E4 is associated with Class II facial convexity in OSA subjects, but not with BMI. This finding implies a genetic connection between *APOE* and facial growth only, and not between APOE and BMI in OSA patients. Skeletal remodeling variation with *APOE* E4 could be a mechanism to be investigated.

791W

Large-scale integrative fine-mapping over 22 complex traits with over 3. 3 million phenotypic measurements. *G. Kichaev[1], H. Shi[1], M. Roytman[1], B. Pasaniuc[1,2,3].* 1) Interdepartmental Program in Bioinformatics, University of California, Los Angeles, Los Angeles, CA, 90095; 2) Dept. of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA. 90095, USA; 3) Dept. of Pathology and Laboratory Medicine,David Geffen School of Medicine University of California Los Angeles, Los Angeles, CA. 90095, USA.

Identification of causal variants underlying GWAS risk loci remains one the main challenges in the post-GWAS era. This is typically accomplished in fine-mapping studies across tens to hundreds of thousands of individuals. Such large-scale analyses are often prohibited by constraints in data sharing of individual-level genetic data. Multiple recent works have proposed to circumvent this by operating directly at the summary statistic level thus not requiring individual genetic data. In addition, the recent completion of the Roadmap Epigenomics Project [Kundaje et al. Nature 2015] has provided a wealth of functional genomic information that can aid fine-mapping endeavors if properly modeled. We have previously described fine-mapping frameworks [Kichaev et al. PloS Gen. 2014] capable integrating functional genomic data that operate directly on summary association statistics and demonstrated its efficacy for a small number of traits and functional marks. Here, we describe a large-scale analysis of over 3. 3 million phenotypic measurements spanning 22 diverse dichotomous and quantitative traits. We conduct integrative fine-mapping on approximately 1,700 genetic loci, leveraging over 4,000 functional genomic marks to build trait-specific models of functional genomic architectures to help uncover causal variants. We report a number of plausible functional variants underlying GWAS risk loci and create a searchable database that we make freely available to the community.

792T

Single nucleotide polymorphisms in the *APCS* gene influence geographic atrophy secondary to VEGF inhibition therapy in neovascular macular degeneration. *G. J. McKay[1], G. Silvestri[2], R. E. Hogg[2], L. Toth[2], P. Earle[2], C. C. Patterson[1], U. Chakravarthy[2].* 1) Centre for Public Health, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom; 2) Centre for Experimental Medicine, Queen's University Belfast, Belfast, Northern Ireland, United Kingdom.

Introduction: Age-related macular degeneration (AMD) is a leading cause of vision loss in the elderly mostly due to the development of neovascular AMD (nAMD) or geographic atrophy (GA). Intravitreal injections of anti-vascular endothelial growth factor (VEGF) agents are an effective therapeutic option for nAMD. Following anti-VEGF treatments, increased atrophy of the retinal pigment epithelium (RPE) and choriocapillaries that resembles GA has been reported. We sought to evaluate the underlying genetic influences that may contribute to this process. **Methods:** We selected 68 single nucleotide polymorphisms (SNPs) from genes previously identified as susceptibility factors in AMD, along with 43 SNPs from genes encoding the VEGF protein and its cognate receptorsas this pathway is targeted by treatment. We enrolled 467 consecutive patients (Feb 2009 to October 2011) with nAMD who received anti-VEGF therapy. The acutely presenting eye was designated as the study eye and retinal tomograms graded for macular atrophy at study exit. Statistical analysis was performed using PLINK to identify SNPs with a P value < 0. 01. Logistic regression models with macular atrophy as dependent variable were fitted with age, gender, smoking status, common genetic risk factors and the identified SNPs as explanatory variables. **Results:** Grading for macular atrophy was available in 304 study eyes and 70% (214) were classified as showing macular atrophy. In the unadjusted analysis we observed significant associations between macular atrophy and two independent SNPs in the *APCS* gene: rs6695377: odds ratio (OR) = 1. 98; 95% confidence intervals (CI): 1. 23, 3. 19; P = 0. 004; rs1446965: OR = 2. 49, CI: 1. 29, 4. 82; P = 0. 006 and these associations remained significant after adjustment for covariates. **Conclusions:** VEGF is a mitogen and growth factor for choroidal blood vessels and the RPE and its inhibition could lead to atrophy of these key tissues. Anti-VEGF treatment can interfere with ocular vascular maintenance and may be associated with RPE and choroidal atrophy. As such, these medications, which block the effects of VEGF, may influence the development of GA. The top associated SNPs are found in the *APCS* gene, a highly conserved glycoprotein that encodes Serum amyloid P (SAP) which opsonizes apoptotic cells. SAP can bind to and activate complement components via binding to C1q, a mechanism by which SAP may remove cellular debris, affecting regulation of the three complement pathways.

793F

Identification of a p. Arg138Trp variant in *TM6SF2* associated with triglyceride and total cholesterol levels in American Indians. *A. Nair, M. Kaur, P. Kumar, P. Piaggi, S. Kobes, W. Knowler, C. Bogardus, R. Hanson, L. Baier.* National Institute of Diabetes and Digestive and Kidney Diseases, NIH, Phoenix, AZ.

Prior genome-wide association studies identified strong association between serum triglyceride (TG) and total cholesterol (TC) levels and variants at the gene rich chr19p13 region. In multiple ethnic groups rs58542926, which predicts a Glutamine to Lysine (p. Glu167Lys) change in *TM6SF2,* was pinpointed as the causal variant, and in functional studies the Lys allele decreased TM6SF2 protein levels and *Tm6sf2* knockdown resulted in reduced TG rich lipoprotein secretion from liver resulting in lower plasma TG and TC levels. Accordingly, strong associations between this variant and myocardial infarction (MI) and non-alcoholic fatty liver disease (NAFLD) have been identified with the Lys allele increasing risk for NAFLD and decreasing risk for MI. In the current study we analyzed coding variation in *TM6SF2* for association with lipid levels in Pima Indians, a population with a high prevalence of type 2 diabetes but a relatively low rate of coronary heart disease (in non-diabetics) and comparatively favorable lipid profiles. Analysis of 335 whole genome sequence data from Pima Indians identified two coding variants, rs142056540 (p. Arg138Trp) and the previously reported Glu167Lys variant which were then genotyped in 5494 Pima Indian subjects informative for lipid levels. The Lys allele at Glu167Lys had a lower frequency in American Indians (2%) compared to Europeans (7%) and had a directionally consistent nominal association with lower TC (by 4% per copy, *P*=0. 005) but not with TG levels (*P*=0. 15). In contrast, the Trp allele at the novel Arg138Trp variant was strongly associated with lower serum TG (by 18%, *P*=1. 0×10-13) and TC levels (by 6%, *P*= 9. 7×10-9). The Arg138Trp variant was further genotyped in 2899 Urban American Indians and a meta-analysis of all 8393 American Indians identified a genome-wide significant association with TG and TC levels such that the TG levels were lower by 18% (*P*= 8. 1×10-14) and TC levels were lower by 6% (*P*= 5. 6×10-10) per copy of the Trp allele. The Trp allele occurs in 10% of full heritage Pima Indians but was not identified in Europeans, Asians or Africans and in only 2/128 subjects of Mexican ancestry in the 1000 genome data. In Pima Indians, Arg138Trp variant is not in LD with either the Glu167Lys variant (r2=0, D'=0) or with any other SNP in *TM6SF2* (*n*=335). The Arg residue is highly conserved across different species and the SNP is predicted to be damaging by SIFT. Functional characterization of the Arg138Trp variant is ongoing.

794W

Mesocorticolimbic Genetic Variants are Associated with Scores on Opioid Risk Screening Tools. *T. Onojighofia, N. Anand, B. Meshkin, M. Hafez, S. Kantorovich, D. Holman, M. Hopkins.* Research & Department, Proove Biosciences, Irvine, CA.

Background

For patients with acute, subacute, or chronic pain, opioids are a common and sometimes effective part of the pain management plan. Before opioids are initiated, a comprehensive evaluation of the risk of misuse and abuse-related behaviors must be conducted. Currently available screening tools that are validated for the assessment of potential opioid addiction such as the Opioid Risk Tool (ORT) and the Screener and Opioid Assessment for Pain Patients (SOAPP®-R) are based on patients' self-reports of behavior; personal and family medical history; and substance abuse-related environmental factors. However, there is evidence that genetic factors may determine pain perception and up to half of substance abuse risk.

Statement of Purpose

The objective of this study was to establish a basis for associating the potential for patient abuse of opioid medications with genetic markers.

Methods

For this retrospective, cross-sectional association study, 2488 subjects were randomly selected from 2 pain centers in the United States. Subjects were treated for primarily chronic non-cancer pain conditions; currently using opioid medications; and had completed either the Opioid Risk Tool (ORT) or the Screener and Opioid Assessment for Pain Patients, Revised (SOAPP-R). Test scores were stratified into low, moderate, and high risk for opioid abuse. All subjects were genotyped using Taqman® SNP Genotyping Assays (Life Technologies, Carlsbad, CA, USA). Analyses were conducted to determine associations between risk scores and a panel of 12 single nucleotide polymorphisms (SNPs), all of which encode for proteins expressed in the mesocorticolimbic reward pathway of the brain.

Results

From our panel of 12 SNPs, analyses showed that ANKK1/DRD2 rs1800497 was significantly associated with ORT results (p=0. 032), while OPRM1 rs1799971 was nominally associated with ORT results (p=0. 067). OPRM1 rs1799971 (p=0. 001), OPRK1 rs1051660 (p=0. 014), DRD1 rs4532 (p=0. 014), and GABRG2 rs211014 (p=0. 044) were all significantly associated with SOAPP-R test results.

ConclusionFive SNPs from a panel of genes associated with opioid abuse and opioid abuse risk factors were found to be associated with 2 widely used opioid risk assessment tools. Thus, identifying opioid abuse risk-related genes in chronic non-cancer pain patients may help to strengthen overall assessments for opioid abuse when opioid therapy is ongoing or being considered.

795T

IBD associated SNP rs4256159 at chromosome 3p24 gene desert is correlated with increased colonic expression of *SATB1*. *A. Onoufriadis[1], R. Sarna[1], S. Twelves[1], L. Demandt[1], F. Gracio[2], A. Amar[1], E. De Rinaldis[2], JD. Sanderson[3], PM. Irving[3], C. Mathew[1], NJ. Prescott[1].* 1) Medical and Molecular Genetics, King's College London, London, United Kingdom; 2) NIHR Biomedical Research Centre, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom; 3) Department of Gastroenterology, Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom.

Genome-wide association scans (GWAS) have now identified more than 160 genomic regions that influence susceptibility to the two forms of inflammatory bowel disease (IBD), Crohn's disease (CD) and ulcerative colitis (UC). However, GWAS loci only account for a modest proportion of the disease variance in IBD, with genetic variants that underlie protein coding changes playing a relatively minor role. It is therefore hypothesised that the majority of these IBD loci will contain causal variants that exert their effect by altering gene expression levels. In order to assess this, we sequenced the transcriptome of intestinal biopsy samples from 32 IBD patients and controls and focused on transcripts that map to GWAS loci that are devoid of any annotated coding genes, known as "gene-deserts". We observed intestinal transcription of *SATB1-AS1* (special AT rich sequence binding protein 1 antisense 1), a long non-coding RNA that maps to an IBD-associated gene desert at chromosome 3p24 as well as *SATB1* (special AT rich sequence binding protein 1), a protein-coding gene immediately proximal to the gene desert. In order to assess whether the associated SNP in this region (rs4256159) is correlated with altered expression of *SATB1-AS1* or the nearby *SATB1* gene, we performed qPCR in intestinal tissue and whole blood from 85 individuals genotyped for the SNP. eQTL analysis showed a significant upregulation of intestinal *SATB1* expression in individuals carrying the risk allele (p=0. 02098), whereas no significant correlation with *SATB1-AS1* transcription in colon was observed (p=0. 4692). The average expression based on relative quantification (RQ) data was 1. 6 fold greater for the CT genotype and 2. 48 fold greater for the TT genotype, when compared to the non-risk CC. In contrast, eQTL analysis between the risk SNP and expression of *SATB1* and *SATB1-AS1* in whole blood did not show any correlation, suggesting a tissue specific effect. *SATB1* is involved in chromatin remodelling in T-cell differentiation and over-expression of *SATB1* in regulatory T-cells (Treg) has been shown to cause loss of their immunosuppressive function. Taken together, this work suggests the rs4256159 SNP may be confer susceptibility to IBD due to upregulated *SATB1* expression in intestinal mucosa, thus providing insights into the underlying biology of disease associations in gene deserts.

796F

Gender-specific association of *ADIPOQ* gene polymorphisms with adiponectin levels and the influence of genetic ancestry on the association with obesity in the Jackson Heart Study cohort. *P. Riestra[1], SY. Gebrea[1], R. Xu[1], R. Khan[1], A. Bidulescu[2], A. Correa[3], F. Tekola-Ayele[4], SK. Davis[1].* 1) Cardiovascular Section Social Epidemiology Research Unit . National Human Genome Research Institute. 10 Center Drive Bethesda, MD 20892; 2) Indiana University Bloomington School of Public Health. 1025 E. 7th Street, Suite 111 Bloomington, IN 47405; 3) Jackson Heart Study Jackson Medical Mall. 350 West Woodrow Wilson Av. , Suite 701 Jackson, MS, 39217; 4) National Human Genome Research Institute Center for Research on Genomics and Global Health. 12 South Drive Bethesda, MD 20892.

Background: Despite the important role of adiponectin in regulating the general metabolic homeostasis, the analysis of genetic determinants of adiponectin and the related cardio-metabolic traits in African American population has been very limited and inconsistent. Considering the high genetic admixture of African Americans and thus the important population stratification that may be confounding the genetic-trait associations, the objective of this work was to perform a comprehensive analysis of the associations between ADIPOQQ variants and adiponectin levels and obesity phenotypes in a large population of African Americans from the JHS cohort. **Design and Methods**: Genotype data was available for 2968 JHS participants (1131men and 1837women). SNPs were selected by a Tag-SNP Approach and literature review. The genotype imputation was performed using. IMPUTE2 software and reference phased data from the 1000G project. PLINK software was use for the genetic analysis. Plasma specimens were analyzed by ELISA for adiponectin levels. Global European ancestry estimates were calculated using HAPMIX in analyses supported by the CARe consortium. **Results:** We found a gender dependent association of some ADIPOQ variants and adiponectin levels. Only in women four of the studied polymorphisms (rs 6444174, rs 16861205, rs1403697, rs7641507) were associated with adiponectin levels after Bonferroni correction and controlling for the percentage of European ancestry (PEA), age, income and smoking. These results were consistent with the haplotype analysis. The association between the rs12495941 variant and obesity is modulated by the PEA, so that the relationship between the G allele and a higher incidence of obesity was present in those individuals within the lower PEA group. In addition there was an effect modification of obesity on the association between the ADIPOQ rs 6444174 SNP and BMI so that the presence of the T allele was negatively and significantly associated with BMI (p=0. 0003662) only in participants with a normal weight. **Conclusions:** In this big African American cohort, ADIPOQ variants are associated with adiponectin levels in a gender dependent manner and the relationship of some of these variants with obesity and BMI is modulated by the PEA and the obesity status respectively suggesting that the effects of these polymorphisms are subject to a strong interaction with genetic and environmental factors in African American population.

797W

Fine-scale mapping of genomic heritability using summary association statistics. *H. Shi[1], G. Kichaev[1], B. Pasaniuc[1,2,3].* 1) Interdepartmental Program in Bioinformatics, University of California, Los Angeles, Los Angeles, CA 90095; 2) Dept. of Human Genetics, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA. 90095; 3) Dept. of Pathology and Laboratory Medicine,David Geffen School of Medicine University of California Los Angeles, Los Angeles, CA. 90095.

Genome-wide association studies (GWAS) have reproducibly identified thousands of associations yet the majority of heritability remains largely unexplained. This is in part caused by the fact that many causal variants with low effect size fail to reach significance level due to limited sample size. Heritability estimation and partitioning provides a way of interrogating the entire genetic architecture of the disease without imposing significance thresholds on the strength of association between individual variants and trait. In this work we provide an estimator for heritability at fine-scale genomic windows (less than 100Kb) from GWAS summary statistics while taking linkage disequilibrium into account. Through extensive simulations, we demonstrate that the estimator is unbiased under various window size, sample size, and reference panel sizes. We apply our new estimator to summary association data of over 3. 3 million phenotypic measurements spanning 22 diverse traits to characterize the genetic architecture of common disease. In a preliminary analysis of BMI and height summary statistics from GIANT Consortium, as expected we observe an enrichment of GWAS hits in genomic regions with high heritability. Interestingly, we also observe a number of genomic regions with high heritability but without any GWAS hit.

798T

Lack of Association Between Leptin T-2549G and Adiponectin C-11377G Genotype in Asthmatic-Obese Children. *H. Verdi[1], O. Ozbek[2], Y. Yalcin[1], FB. Atac[1].* 1) Baskent University School of Medicine, Department of Medical Biology, Ankara, TURKEY; 2) Baskent University School of Medicine, Department of Pediatrics, Ankara, TURKEY.

Asthma the most common pediatric inflammatory disorder is a complex syndrome that encompasses multiple phenotypes. The prevalence, severity and hospitalization index for asthma have increased markedly in the last several decades. Hence, asthma is often diagnosed along with obesity which could be due to the systemic inflammation. Leptin and adiponectin are the two adipokines secreted by active adipose tissue. It is known that different genotypes may enroll in the development of both conditions. The aim of this study was to evaluate the leptin -2549 T/G and adiponectin -11377 C/G genotypes in children with asthma and/or obesity and their relations with leptin and adiponectin levels. 402 children were enrolled in study. Asthma was defined according to GINA criteria and children with a BMI above the 95th percentile for age and sex were defined as obese. Patients were investigated in four groups. Group 1: Obese children with asthma (n=76) Group 2: Lean children with asthma (n=105) Group 3: Obese children without asthma (n=96), Group 4: Lean healthy children (n=125). Serum leptin and adiponectin levels were measured by ELISA and genotyping was performed by PCR-RFLP. Both the frequency of leptin -2549 T/G and adiponectin -11377 C/G genotypes did not show difference between study and control group. No association was found between leptin/adiponectin genotype and leptin/adiponectin levels. Not only in the genotype distribution but serum levels of leptin/adiponectin show difference between study and control groups. Mean leptin levels of obese asthmatics were significantly higher (13. 1±9. 1) than lean asthmatics (3. 7±4. 4) ($p<0.001$). Serum adiponectin levels of lean asthmatics (16±7. 1) were significantly higher than obese asthmatics (12. 1±6. 9) and lean healthy children (13. 2±5. 9) ($p<0.001$ and $p<0.05$, respectively). Although adiponectin level in adiponectin -11377 CC genotype (14. 4±6. 6) was higher than -11377 CG group (9. 3±3. 7) in healthy controls, the difference was statistically insignificant ($p=0.08$). No significant difference was defined in our study. Further investigations are required that could potentially open new scenario in the diagnostic algorithm and in the strategic approach, with a more comprehensive assessment of the disease.

799F

The impact of known type 2 diabetes associated variants on circulating levels of GLP-1, GIP and glucagon during an oral glucose tolerance test. *A. Jonsson[1], S. S. Torekov[1,2], T. Schnurr[1], C. T. Have[1], Y. Mahendran[1,4], N. Grarup[1], N. B. Johansen[3,4], K. Færch[3], D. R. Witte[4,5], T. Lauritzen[5], A. Sandbæk[5], J. J. Holst[1,2], M. E. Jørgensen[3], O. Pedersen[1], T. Hansen[1].* 1) NNF Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 2) Department of Biomedical Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 3) Steno Diabetes Center, Gentofte, Denmark; 4) The Danish Diabetes Academy, Odense, Denmark; 5) Department of Public Health, University of Aarhus, Aarhus, Denmark.

Aims/hypothesis: Large-scale genome wide association studies (GWAS) have currently identifiedmultiple genetic determinants of type 2 diabetes. For many of those variants, it is unclear through which mechanisms they exert their effect and little is known whether these genes also affect incretin hormone levels and/or alpha-cell function. Our aim is to examine the impact of known type 2 diabetes associated variants on circulating levels of glucagon-like peptide-1 (GLP-1), glucose-dependent insulinotropic polypeptide (GIP) and glucagon during an oral glucose tolerance test (OGTT). **Methods:** Plasma levels of GLP-1, GIP and glucagon were examined in samples obtained during an OGTT in 1,547 individuals from the ADDITION-PRO cohort. Participants were given a standard 75 g OGTT after an overnight fast of ≥7 hours and blood samples were drawn at 0, 30 and 120 min for assessment of plasma GLP-1, GIP and glucagon levels. The incremental area under the curve (iAUC) of plasma GLP-1, GIP and glucagon were calculated from 0-30 minutes and from 0-120 minutes during the OGTT. Associations between 68 genetic variants and plasma levels of GLP-1, GIP and glucagon were studied using a linear mixed model (EMMAX) implemented in the EPACTS software package by the use of inverse-normalized residuals of the traits adjusted for age, sex and BMI. **Results:** None of the associations of 68 type 2 diabetes risk variants with circulating plasma levels of GLP-1, GIP and glucagon during the OGTT were significant after correction for multiple testing. However, risk alleles in *PPARG* (rs1801282) and *ADCY5* (rs11708067) were associated with increased iAUC of plasma glucagon from 0 to 30 min during the OGTT (Beta [SE]: 0. 171 [0. 054], $P=0.0015$ and 0. 130 [0. 043], $P=0.0025$, respectively), showing reduced repression of glucagonrelease after an oral glucose load. **Conclusion/interpretation:** Known type 2 diabetes risk variants seem not to be strongly associated with alteredpost-OGTT releaseof GLP-1, GIP or glucagon. Large-scale genome-wide association studies of incretin hormone and glucagon release as well as studies of incretin hormone and glucagon action are needed in order to explain the genetic influence on these traits.

800W

Assessing a genetic risk score for type 2 diabetes with abdominal obesity in a Korean population. *J. Lee, Y. Lee, S. Park, M. J. Go, B. J. Kim.* Korea Center for Disease Control and Prevention, Osong, South Korea.

Type 2 diabetes mellitus (T2DM) is a complex disease and metabolic disorder. For few years, genome-wide association studies (GWASs) have found many novel susceptibility loci for T2DM, and have demonstrated that some genetic variants have different effect according to the ethnics. We evaluated single-nucleotide polymorphisms (SNPs) in or near 73 susceptibility loci for T2DM, identified through GWASs of groups with European ancestry and Asian populations using 1,201 cases with T2DM and 328 control individuals aged 40 to 49. We selected 10 significant susceptibility loci and calculated a genetic risk score (GRS) by combining the number of risk alleles. We demonstrated that a GRS based on these 10 susceptibility loci (GRS-10) for T2DM was an important risk factor in the Korean population. Addition of GRS-10 to a clinical model resulted in a small but significant increase in the area under the curve (AUC). Next, we performed abdominal obesity-stratified analyses and assessed the GRS-10 in abdominal obesity and non-abdominal obesity groups.

801T

Genetic Variation in the Maltase-Glucoamylase Gene (*MGAM*) Associates with Measures of Body Mass Index in American Indian Adults and Children. *Y. Muller, P. Piaggi, C. Okani, G. Wiessner, S. Kobes, P. Chen, W. Knowler, R. Hanson, C. Bogardus, L. Baier.* Phoenix Epidemiology and Clinical Research Branch, National Institutes of Health, Phoenix, AZ.

Intestinal maltase-glucoamylase (*MGAM*) is an alpha-glucosidase which hydrolyzes linear alpha-1,4-linked oligosaccharide in starch digestion. It plays a crucial role in the production of glucose in the human lumen. The antidiabetic drug Miglitol which is an alpha-glucosidase inhibitor reduces HbA1c and ameliorates obesity in patients with type 2 diabetes (T2D). Therefore, *MGAM* was analyzed as a potential candidate gene for obesity and T2D in Pima Indians. Forty five SNPs tagging 610 common variants (minor allele frequency ≥0. 05; r2≥0. 85) across a ~211 kb genomic region (Chr7:141645679-141856547 which covers ~50kb upstream and downstream of *MGAM*) were genotyped as part of a custom Axiom array (Affymetrix) in a population-based sample of 3494 Pima Indians from a longitudinal study. Genotypes were analyzed for associations with maximum body mass index (BMI) measured in adulthood, maximum BMI z-score in childhood, % body fat and T2D. Conditional analyses demonstrated that two intronic tag SNPs (rs1880924 and rs58864948, D'=0. 07, r2=0) in *MGAM* independently associated with BMI. These two SNPs were further genotyped in a replication sample of 4060 mixed-heritage American Indians. SNP rs1880924 [risk allele frequency (RAF) = 0. 17] associated with adult BMI in both samples (full-heritage Pima Indians: b=2. 6% per copy of T allele, p=8. 2 x 10-4; mixed-heritage American Indians: b=2. 6%, p=8. 6 x 10-4). Combining both samples provided the strongest evidence for association with adult BMI (b=2. 5%, p=6. 3 x 10-6, n=7697). The rs1880924 allele associating with higher BMI in adulthood also associated with a higher childhood BMI z-score (b=0. 08 SD, p=0. 004, n=5350), and a higher % body fat (b=0. 89%, p=0. 05, n=744), but did not associate with T2D [odds ratio=1. 11(0. 99-1. 24) per copy of T allele, p=0. 08, n=7679]. The second SNP rs58864948 (RAF = 0. 23) similarly associated with BMI in adulthood (b=1. 6%, p=0. 001), BMI z-score in childhood (b=0. 049 SD, p=0. 05) and % body fat (b=1. 06%, p=0. 008). When the risk alleles at each of these two SNPs were added together, each additional risk allele increased BMI by 2% in adulthood (p =8. 7x10-8) and BMI z-score by 0. 07 SD in childhood (p=7. 2 x 10-4). Although neither SNP achieved genome-wide significance, both SNPs associated with BMI after accounting for multiple testing (45 tag SNPs require a p≤0. 001 for significance). We conclude that common variation in *MGAM* modestly affects BMI measured throughout lifetime in American Indians.

802F

Analysis of common variants in *SLC19A2* identifies a modest association with type 2 diabetes in Pima Indians. *M. Traurig, A. Moulton, S. Kobes, R. L. Hanson, W. C. Knowler, C. Bogardus, L. Baier.* NIH, Phoenix, AZ.

The *SLC19A2* gene encodes a transporter protein involved in thiamine uptake. *Slc19a2* knockout mice develop diabetes and rare mutations in human *SLC19A2* cause thiamine-responsive megaloblastic anemia (TRMA) syndrome often characterized by early-onset diabetes. In previous linkage studies, the region on chromosome 1 which contains the *SLC19A2* locus was linked to type 2 diabetes in Pima Indians and several other ethnic groups. Whole genome sequencing of 335 Pima Indians identified one rare (mAF=0. 014) missense variant (rs150049339, Thr365Ser) in exon 4, however it was not associated with type 2 diabetes in ≈3500 Pima Indians (P=0. 32). In the current study, we are assessing whether common variants in *SLC19A2* are associated with type 2 diabetes in Pima Indians. We are utilizing genotypic data from a recently designed custom Pima Indian specific Affymetrix Axiom array which captures 92% of all variation with a mAF ≥5% in the Pima Indian genome. Currently, 3,637 full-heritage Pima Indians (Stage 1) have been genotyped with this array which contains 9 tag SNPs in *SLC19A2*; genotyping in an additional 4,060 mixed-heritage American Indians (Stage 2) is ongoing. To date, 2 of the 9 tag SNPs, rs10919175 and rs2056926, across the *SLC19A2* locus have been analyzed in both samples. Rs2056926 was associated with diabetes in both samples and the combined analysis (N=7,697) had a P=2. 1×1 0-6 and OR=1. 28 per copy of the C allele. However, the diabetes association did not replicate in a separate sample of American Indians from the FIND study. Rs2056926 is located in the promoter region 227 bp upstream of the transcription start site; therefore, we examined whether the variant could function as a cis-acting expression quantitative trait locus (eQTL) for *SLC19A2*. Genotypic data for rs2056926 were merged with adipose (N=192) and skeletal muscle (N=204) gene expression data from non-diabetic Pima Indians. A nominal association between rs2056926 and *SLC19A2* expression was observed in adipose (P=0. 03, B=-0. 29, SD per C allele) and the diabetes risk allele was associated with lower *SLC19A2* expression; however, it is unclear whether this weak association is a false positive. Expression levels for *SLC19A2* in the adipose and skeletal muscle samples were very low; therefore, additional functional studies using tissues (e. g. liver) with higher *SLC19A2* expression, which are not available in our study of Pima Indians, are needed to determine whether rs2056926 affects *SLC19A2* transcription.

803W

The genetic and regulatory architecture of the *ERBB3*- type 1 diabetes susceptibility locus. *S. Kaur[1], A. H. Mirza[1,2], T. Floyel[1], C. A. Brorsson[1], H. B. Mortensen[1,2], J. Størling[1], F. Pociot[1,2], The Hvidoere Study Group.* 1) Herlev University Hospital, Herlev, Denmark; 2) University of Copenhagen, Denmark.

Aims- Epidermal growth factor receptor (EGFR) signaling pathway is suggested to play a role in β-cell growth and islet development, and *ERBB3* gene is an important member of this pathway. The genetic and regulatory architecture of the *ERBB3* locus is not well explained with respect to type 1 diabetes. The aim of our study was to explore the role of *ERBB3* locus in type 1 diabetes. **Methods-** We genotyped 984 children with new-onset type 1 diabetes and examined if genetic variation of *ERBB3* (rs2292239) affects residual β-cell function. Furthermore, we systematically explored the regulatory architecture of the *ERBB3* locus, examined the expression of *ERBB3* and its potential regulators in human islets and investigated the effect of *ERBB3* knockdown on apoptosis in INS-1E cells. **Results-** rs2292239 strongly correlated with residual β-cell function and metabolic control in children with type 1 diabetes. We identified evolutionary conserved long non-coding RNAs (lncRNAs) expressed in tissue-specific manner, multiple *CTCF* transcription factor binding sites and *cis*-eQTL signals, suggesting their putative regulatory roles in modulating *ERBB3* expression. Additionally, we found that *ERBB3*, *CTCF*, and antisense lncRNA *NONHSAG011351* are expressed in human islets and that *ERBB3* is down-regulated by pro-inflammatory cytokines. Importantly, knockdown of *ERBB3* in insulin-producing INS-1E cells decreased the expression of CTCF and diminished basal and cytokine-induced apoptosis as shown by reduced Caspase-3 activation. **Conclusions-** Our data suggests an important functional role of *ERBB3* and its regulators in the b-cells and may constitute novel targets to prevent b-cell destruction in type 1 diabetes.

804T

Impact of *APOL1* genetic variation on HIV-1 infection and opportunistic infections in African-American patients with HIV-1. *P. An[1], E. Sezgin[1,2], G. D. Kirk[2], M. L. Van Natta[2], D. A. Jabs[2,3,4], A. Dinh[5], J. B. Kopp[5], C. A. Winkler[1].* 1) Basic Research Laboratory, Center for Cancer Research, National Cancer Institute, Leidos Biomedical Research, Inc, Frederick National Laboratory for Cancer Research, Frederick, MD; 2) Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD; 3) Departments of Ophthalmology, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Departments of Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 5) Kidney Disease Section, NIDDK, NIH, Bethesda, MD.

APOL1, encoding apolipoprotein L1, confers resistance to *Trypansoma brucei*, the cause of African trypanosomiasis. *APOL1* G1 (S342G:I384M) and G2 (6 bp deletion) alleles restore resistance to trypanosomes causing acute human trypanosomiasis and are found only in individuals with recent African ancestry. *APOL1* variant alleles are recessively associated with increased risk for HIV-associated nephropathy (OR 27-89); ~12-14% of the AA population carry high-risk genotypes. *In vitro* APOL1 inhibits HIV-1 replication by several mechanisms; the *in vivo* role of *APOL1* variants on HIV-1 infection, progression, and disease outcomes has not been investigated. We determined *APOL1* genotypes for AA enrolled in the ALIVE HIV natural history and LSOCA AIDS cohorts. In the ALIVE cohort (n = 749), 7. 7% of AA HIV+ carried high-risk genotypes compared to 12% of at-risk HIV- individuals (OR 0. 60, 95% CI: 0. 33-1. 08; P = 0. 09). Individuals with the three *APOL1* genotypes showed similar HIV-1 viral load levels (P> 0. 40). *APOL1* genotypes were not associated with the rate of progression from HIV seroconversion to AIDS in survival analyses (P > 0. 70). However, high-risk genotypes were associated with a 71% decrease in the odds of opportunistic infection (OI) (OR 0. 29, 95% CI: 0. 09, 0. 97; P = 0. 04). In the LSOCA cohort (n = 784), *APOL1* genotypes were not associated with baseline CD4+ T-cell counts and HIV-1 viral loads. Carriage of high-risk genotypes was associated with a non-significant decrease in the odds of prevalent OI at enrollment (OR 0. 66, 95% CI: 0. 40, 1. 11; P = 0. 11), which was mainly driven by a protective association with fungal infection (OR 0. 54, 95% CI: 0. 32, 0. 93; P = 0. 02). In summary, *APOL1* G1/G2 variants were not associated with viral load or AIDS progression, but *APOL1* association with resistance to HIV acquisition and OI cannot be excluded. These results, requiring further validation, suggest potential novel anti-pathogen roles for APOL1. Funded by National Cancer Institute Contract HHSN261200800001E; National Institute on Drug Abuse (R01-DA04334, R01-12586); National Eye Institute (U10 EY 08052 to Icahn School of Medicine at Mount Sinai; U10 EY 08057 to Johns Hopkins University Bloomberg School of Public Health; U10 EY 08067 to University of Wisconsin, Madison, School of Medicine).

805F

Immunochip analysis identifies amino acid residues in five separate HLA genes driving the association between the MHC and primary biliary cirrhosis. *R. Darlay[1], K. L. Ayers[1,2], G. F. Mells[3], V. A. Money[4], J. Z. Liu[5,6], M. A. Almarri[5,7], G. J. Alexander[8], D. E. Jones[9], R. N. Sandford[3], C. A. Anderson[5], P. T. Donaldson[9], H. J. Cordell[1], UK-PBC Consortium.* 1) Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, Tyne and Wear, United Kingdom; 2) Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, USA; 3) Academic Department of Medical Genetics, Cambridge University, Cambridge, UK; 4) School of Biological and Biomedical Sciences & Department of Chemistry, Biophysical Sciences Institute, Durham University, Durham, UK; 5) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK; 6) New York Genome Center, New York, USA; 7) Department of Forensic Science and Criminology, Dubai Police HQ, Dubai, United Arab Emirate; 8) Department of Hepatology, Cambridge University Hospitals National Health Service (NHS) Foundation Trust, Cambridge, UK; 9) Institute of Cellular Medicine, Medical School, Newcastle University, Newcastle upon Tyne, UK.

Primary biliary cirrhosis (PBC) is a classical autoimmune liver disease characterized by progressive auto-immune destruction of intrahepatic bile ducts. The strongest genetic association is seen with *HLA-DQA1*04:01*, with at least three additional independent HLA haplotypes contributing to susceptibility. To identify functional, potentially causal variants within the HLA region, we used dense SNP data in 2861 PBC cases and 8514 controls to impute classical HLA alleles and amino acid polymorphisms using the software packages HIBAG and SNP2HLA. Through stepwise analysis we demonstrate that association in the HLA region is largely driven by variation at five separate amino acid positions: position 11 of HLA-DPb1 (p = 4. 72e-60), position 74 of HLA-DRb1 (p = 2. 95e-40), position 57 of HLA-DQb1 (p = 1. 73e-21), position 156 of HLA-C (p=2. 83e-12) and position -13 in the signal peptide of HLA-DQa1 (p = 7. 64e-12). Three dimensional modelling and calculation of electrostatic potentials was performed to explore the effect of these key residues on these molecules. Two of the associated residues were shown to affect the electrostatic charge of the peptide binding groove. An aspartic acid at residue 57 of HLA-DQb1, which protects against PBC, induces a negative charge in pocket 4 of the peptide binding groove, whereas an arginine at residue 156 of HLA-C, which is associated with an increased risk of PBC, induces a positive charge within the peptide binding groove.

806W

Identification and characterization of functional genetic variants in Dupuytren's disease. *J. Du[1,2], K. Becker[1,2], J. Altmüller[1], H. Thiele[1], S. Tinschert[3], R. Casper[1], P. Nürnberg[1,2], HC. Hennies[1,2,4], The German Dupuytren Study Group.* 1) Cologne Center for Genomics (CCG), University of Cologne, Germany; 2) Cologne Excellence Cluster on Cellular Stress Responses in Aging-associated Diseases (CECAD), University of Cologne, Germany; 3) Institute of Clinical Genetics, Dresden University of Technology, Germany; 4) Center for Dermatogenetics, Div. of Human Genetics, Medical University of Innsbruck, Austria.

Dupuytren's Disease (DD) is a progressive, aging-associated fibromatosis disorder of the palmar aponeurosis, leading to progressive flexion contractures of fingers. DD is the most frequent genetic disorder of connective tissue and current treatment of DD consists largely of surgical or enzymatic removal of the contracted connective tissue, which is, however, associated with risk of neurovascular injury and recurrence. Hence, unraveling the molecular etiology of DD is needed to provide insight into potential therapeutic targets for treatment of DD. Here, we aim to identify causative variants, including both coding and noncoding variants, which directly predispose to DD. In a collaborative study we have previously identified nine disease-associated chromosomal regions using genome-wide association studies (GWAS). We have now further analyzed a 500kb GWAS-identified candidate region on chromosome 7 encompassing the SNP rs16879765 for variants that could explain the strongest disease-association signal by targeted next-generation genomic sequencing of 96 patient samples. The target region was enriched by the SureSelectXT2 kit from Agilent and sequenced on the Illumina HiSeq 2000 platform. Sequencing data was analyzed with the Varbank pipeline (v2. 13). The identified variations were analyzed by SIFT, Polyphen2, Fathmm, GWAVA and Haploregv3. A missense variant was detected in *EPDR1* in four patient samples and predicted to be deleterious. Multiple putative regulatory variants were identified in *EPDR1 and SFRP4* using ENCODE and Roadmap data. We are currently testing the role of these variants on gene expression, functional properties of fibroblasts derived from DD tissue, and regulatory pathways potentially involved in DD pathogenesis. We expect our experiments to pinpoint genetic variants that underlie the manifestation of DD, which peaks at around 60 years of age, and gene networks to unravel pathomechanisms leading to the complex and disfiguring disorder.

807T

Improved methods for multi-trait fine mapping of pleiotropic GWAS risk loci. *M. Roytman[1], G. Kichaev[1], H. Shi[1], R. Tamimi[2,3], S. Lindstrom[3], B. Pasaniuc[1,4,5].* 1) Bioinformatics Interdepartmental Program, University of California Los Angeles; 2) Department of Medicine, Harvard Medical School; 3) Department of Epidemiology, Harvard School of Public Health; 4) Department of Human Genetics, Geffen School of Medicine, University of California Los Angeles; 5) Department of Pathology and Laboratory Medicine, Geffen School of Medicine, University of California Los Angeles.

Genome-wide association studies have identified thousand of variants that alter risk of disease with many loci being implicated in multiple diseases. For example it is widely known that mammographic density and breast cancer share multiple GWAS risk loci [Lindstrom et al Nat Comm 2014]. In this work, we propose methods for the identification of causal variants at loci shown to be at risk for multiple correlated traits. By integrating association data across multiple correlated traits we increase power to find causal variants that impact both traits. Although our proposed methods make an assumption that the same causal variants impact both traits, we allow for heterogeneity in the effects on the two traits. Importantly, our approach only requires summary association statistics thus avoiding the usual restrictions that accompany sharing of individual data. We analyze data on breast cancer and mammographic density, demonstrating that our approach is an improvement on existing methods that only examine one trait. Moreover, recent works have shown that incorporating functional annotations across the genome improves fine mapping performance. Integrating this data into our approach will further enhance our power to detect causal variants.

808F

Devising an imputation reference panel for fine-mapping regulatory variants in autoimmune diseases. *H. Westra[1,2], S. Onengut-Gumuscu[3,4], A. Lee[5], W-M. Chen[3,4], A. R. Quinlan[3,6], P. K. Gregersen[5], S. S. Rich[3,4], S. Raychaudhuri[1,2,7,8].* 1) Brigham and Women's Hospital, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, USA; 4) Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, USA; 5) The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, New York, United States of America; 6) Departments of Human Genetics and Biomedical Informatics, University of Utah, Salt Lake City, Utah, USA; 7) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Smith Building, 1 Jimmy Fund Way, Boston, MA 02115, USA; 8) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.

Over the last few years, over 40 loci have been associated with Type 1 Diabetes (T1D) through GWAS, linkage and candidate gene studies. More recently, the associated regions have been analyzed using the ImmunoChip genotyping platform, which has increased the resolution of the associated SNPs at the already identified loci (Onengut-Gumuscu *et al*, Nature Genetics, 2015). Using H3K4me3 ChIP-seq data for 34 cell types from the ENCODE project, we have previously observed significant enrichment for T1D associated variants in CD4+ T-cells (Trynka *et al*, Nature Genetics, 2014). However, variants used in this enrichment experiment were imputed using the 1000 Genomes Project reference panel, which consists mainly of healthy individuals. We reasoned that variants in these regions may have a higher allele frequency in T1D cases, and as such an imputation panel composed of cases would result in a higher imputation quality. We also reasoned that novel variants, such as insertion/deletion polymorphisms, could underlie these functional regions. We fine-mapped the T1D-associated loci by targeted sequencing followed by variant calling using the Genome Analysis Toolkit (GATK) in 160 T1D trios and 384 rheumatoid arthritis (RA) cases. The sequenced individuals were previously genotyped on ImmunoChip. We selected 900 regions of 1 kilobases, which collectively target 54 loci from the T1D and RA ImmunoChip studies that overlap H3K4me3 peaks in CD4+ T-cells. We detected 4,566 novel variants, of which 72 having each allele observed at least 5 times, imputed these variants into 12,097 T1D cases, and 14,342 controls genotyped on ImmunoChip, and performed association analysis using logistic regression. When comparing the association results with 1000 genomes imputation, we observed similar association p-values ($r > 0.99$) and comparable imputation quality scores ($r > 0.63$). We conclude that our sequencing panel can be used to reliably impute variants in regulatory regions in auto-immune diseases.

809W

Genetic modifiers of age at onset of Alzheimer disease in PSEN1 mutation carrier families. *J. H. Lee[1, 5], R. Cheng[1], B. Vardarajan[1, 2], R. Lantigua[1, 4], G. Tosto[1], D. Reyes-Dumeyer[1], A. Piriz[1], M. Medrano[6], I. Z. Jiménez-Velázquez[7], R. Mayeux[1, 2, 3, 5].* 1) Sergievsky Ctr/Taub Inst, Columbia Univ, New York, NY; 2) Departments of Neurology, College of Physicians and Surgeons, Columbia University, New York, NY; 3) Departments of Psychiatry, College of Physicians and Surgeons, Columbia University, New York, NY; 4) Departments of Medicine, College of Physicians and Surgeons, Columbia University, New York, NY; 5) Departments of Epidemiology, School of Public Health, Columbia University, New York, NY; 6) School of Medicine, Pontificia Universidad Catolica Madre y Maestra, Santiago, Dominican Republic; 7) Department of Internal Medicine, University of Puerto Rico School of Medicine, San Juan, Puerto Rico.

We previously reported that three candidate genes – *SORBS2, SNX25, PDLIM3* -- modified age at onset of Alzheimer disease (AD) in families that carry the G206A mutation in *PSEN1*. Moreover, SNPs in these genes were associated with delayed or earlier age at onset in both mutation carriers and non-carriers in Caribbean Hispanics. We postulate that these genetic factors may influence the risk of AD or may influence memory performance. Here we extend our earlier study to examine 4q35 that contains the *SORBS2* gene using a larger set of families. To identify genetic modifiers, we studied 305 family members from 45 carrier families that had at least two mutation carriers. We genotyped the samples using 700k SNPs chips by Illumina. Using the variance component approach as implemented in SOLAR, we performed a genome wide linkage to confirm and refine the previous identified linkage peak at 4q35. For this purpose, we used age at onset for affected family members and age at last examination for unaffected family members. However, environmental risk factors, such as a history of head injury and educational levels, did not explain the variation in age at onset of AD. This study confirms our earlier linkage finding at 4q35 by identifying a SNP located within *SORBS2* (LOD=3. 25) and further nominates an additional candidate gene *CDKN2AIP* (LOD=4. 28). The minor allele for the *SORBS2* gene was associated with delayed age at onset by ~10 years, whereas other SNPs were associated with earlier age at onset by ~10 years. Here, we report candidate genes that are associated with variation in age at onset of AD in *PSEN1* mutation carrier families. This study illustrates the complex nature of genotype-phenotype relations even in a relatively 'simple' form of early onset AD, and provides the possibility of identifying protective genetic variants.

810T

Genotype and phenotype diversity of Indian populations. *A. Mishra, k. Thangaraj.* Medical and Evolutionary Genetic, Centre for Cellular and Molecular Biology, Hyderabad, Telangana, India.

Human skin color varies remarkably both within and between populations. This variation has been mainly depend upon amount, type and distribution of melanin in the melanin producing cells (melanosomes). Middle gangatic belt (middle east geographical region) of India were characterized by the huge phenotype variations (skin colour). Here we study the *SLC24A5*, which has been shown to explain almost one third of the pigmentation variations. This study was conducted in three cohorts of Indian subcontinent. The first cohort (1167 subjects; 27 diverse populations) were selected for phenotype study, second cohort (448 subjects; 25 population) for phenotype-genotype study and third cohort (1825 subjects; 52 populations of India) for phylogeography study. Phenotype study (cohort-1) suggest wide variation in pigmentation with melanin indices (MI) ranging from 23–90. 4. The highest MI was observed among Manjhis (70. 54 ±6. 53); while, it was lowest among Brahmins (MI-47. 37 ± 6. 05). There was no significant difference between the MI of males and females (p value=0. 1036). Genotype study (cohort-2) have identified two polymorphisms rs1426654 (non-synonymous mutation, exon-3; G<A) and rs2470102 (synonymous mutation, intron-8; G>A). Here, LD pattern is varying from strong (r2 = 0. 81) to moderate (r2 = 0. 30). In studied SNPs, the genotype and allele frequencies of the subjects were in Hardy-Weinberg equilibrium (P<0. 05). Association analysis of both SNPs with pigmentation (MI; Cohort-2) using sex and population as covariates suggest a significant association with skin pigmentation; rs1426654 (p = 4. 493x10-7) and rs2470102 (p = 5. 79x10-7) respectively. The effect of sex was found insigificant (p=0. 69) but the effect of population was highly significant (p value<2. 2x10-16). The presence of one copy of derived allele for SNPs rs1426654 and rs2470102 led to decrease in skin pigmentation of approx 2. 3 and 3. 5 melanin units respectively. In silico approaches suggests that rs2470102 might affect pigmentation through micro-RNA binding processes. Phylogeography study (cohort-3) of both the SNPs suggests wide presence of the derived alleles over the Indian subcontinent albeit their frequencies vary across the populations. Our results suggest that both the polymorphisms (rs1426654 and rs2470102) play an important role in skin pigmentation variation among South Asians.

811F

Hypospadias and SRD5A2 gene: Haplotype frequency estimation and mutation in Iranian patients. *M. Rahimi[1], Mir. Omrani[2,3], R. Mirfakhraie[3], B. Emamalizadeh[1], Z. Fazeli[1], M. Rouzrokh[4], S. Omrani[4,5].* 1) Department of Medical Genetics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 2) School of Paramedical Sciences , Faculty of Paramedical sciences, Shahid Beheshti University, Tehran, Iran; 3) Department of Medical Genetics, Faculty of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 4) Department of Pediatric Surgery, Mofid Children's Hospital, school of Medicine, Tehran, Iran; 5) School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Hypospadias is the most common congenital malformation of the male external genitalia affecting 1 in 250 live male births in the Iranian population. However, little is known the causing genes involved in the disease. The aim of the present study was to investigate mutations in steroid 5α-reductase-2 (SRD5A2) gene in Iranian children with hypospadias. In addition, we studied the association between two functional polymorphisms including V89L and A49T in relation to hypospadias. **Subjects and methods***:* A total of 109 boys with hypospadias and 109 healthy controls were enrolled. We used direct sequencing in order to detect mutations in exons of 1, 4 and 5 and the flanking intronic regions of the SRD5A2 gene. Furthermore the PCR-RFLP was used to investigate the association of V89L and A49T Polymorphisms with hypospadias disease. **Results**: The sequencing of exons 1, 4 and 5 of SRD5A2 gene revealed two new mutations including S74T mutation in one patient (0. 91%), and c. 230-231insA in two patients (1. 83%). In addition, the previously identified G196S mutation was detected in one patient. Moreover, strong association were observed between SRD5A2-V89L (OR=5. 8, 95% CI: 3. 94-26. 25; P<0. 001) and SRD5A2-A49T (OR=10. 2, 95% CI: 3. 8-8. 8; P<0. 001) polymorphisms and hypospadias in the patients. Haplotype analysis indicated that the L-A haplotype confers significant risk of hypospadias in Iranian patients (D'= 0. 364; P= 0. 0009). **Conclusion**:In conclusion, the results of the present study suggested that SRD5A2 mutations and polymorphisms may have a critical function on hypospadias in Iranian population.

812W

The association of HLA-DQA1 gene alleles frequencies and Kawasaki Diseases. *H. Chi, NC. Chiu, MR. Chen, FY. Huang, YJ. Lee.* MacKay Children"s Hospital, No 92, 2nd section, Chung-Shan North Rd. , Taipei, Taiwan.

Background:Kawasaki disease (KD) is a systemic vasculitis caused by unknown infectious agents, host immune dysregulation, and genetic susceptibility in children. The human leukocyte antigen (HLA) system plays an important role in regulating immune response because it takes part in the defense against pathogens. Therefore we designed a case-control study to investigate the association between the HLA-DQA1 allele and KD. Material and methods:The diagnosis of KD was according to the diagnostic criteria of the American Heart Association. Treatment consisted of intravenous gamma globulin (IVIG) and aspirin as soon as the diagnosis was made. All patients had two-dimensional echocardiography during the febrile stage and again after discharge from the hospital. A CAL was defined as the inner diameter of a coronary artery being ≥3 mm in KD patients <5 years old and ≥4 mm in those ≥5 years old or a localized caliber of >150% when compared with the adjacent vessel diameter. The controlincluded hospital personnel and individuals who underwent routine health examinations or minor surgery. None had a history of KD or autoimmune diseases. The DQA1 gene was genotyped using sequencing-based typing. The alleles frequencies of HLA-DQA1*0101, *0102, *0103, *0201, *0301, *0401, *0501 and *0601 were analyzed. Results:The patients were 459 unrelated children (264 males, 195 females) with KD. The age at diagnosis was 1. 8±1. 6 (0. 1–10. 2) years. Among them, 140 (96 males, 46 females) had CAL. The controls were 385 subjects (136 males, 249 females). The frequencies of HLA-DQA1*0301 and HLA-DQA1*0501 were 0. 330 and 0. 184 in theKD group, similar to those of the controls (0. 314 and 0. 205 respectively, both $P > 0.05$). There were no significant difference of HLA-DQA1 alleles frequencies betweenKD patients with CALs and control as well as patients with CALs and those without CALs. Conclusion:The alleles frequenciesof HLA-DQA1 gene were not associated with KD and CALsin Han Chinese children.

813T

Fine-mapping analyses to identify additional variants influencing fasting metabolite concentrations. *R. Li-Gao[1], H. G. de Haan[1], R. de Mutsert[1], A. van Hylckama Vlieg[1], K. Willems van Dijk[2,3], F. R. Rosendaal[1], D. O. Mook-Kanamori[1,4,5].* 1) Clinical Epidemiology, Leiden University Medical Center, C-7, Postbus 3600, 2300 RC, Leiden, the Netherlands; 2) Department of Human Genetics, Leiden University Medical Center, S4-P, P. O. Box 9600, 2300 RC, Leiden, the Netherlands; 3) Department of Endocrinology, Leiden University Medical Center, S4-P, PO Box 9600, 2300 RC, Leiden, the Netherlands; 4) Department of Public Health and Primary Care, Leiden University Medical Center, V-7, PO Box 9600, 2300 RC, Leiden, the Netherlands; 5) Epidemiology Section, BESC Department, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia.

In previous genome-wide association studies (GWAS) fifteen loci (*FADS1, ACADS, ACADM, ACADL, ELOVL2, CPS1, PLEKHH1, SYNE2, SPTLC3, ETFDH, SLC16A9, ACSL1, SCD, SLC22A4* and *PHGDH*) with strong associations to a panel of fasting metabolite concentrations were identified. To discover independent secondary signals, we performed a high-resolution fine-mapping analysis in each gene region, by conditioning on the previously reported GWAS tag variants in multivariable linear regression models. The analyses were performed with data from the Netherlands Epidemiology of Obesity (NEO) study, a population-based cohort study. Both Illumina HumanCoreExome array genotyping and fasting metabolite concentrations by Absolute*IDQ*TM p150 kit were available for 477 participants (mean (SD) age:(56. 3 (5. 7)) year, BMI: (30. 4 (5. 0)) kg/m2, 56. 4% men). Imputation was carried out by 1000 Genomes Phase one integrated variant release v3 using IMPUTE2. After Bonferroni correction, we observed eight secondary association signals in four gene regions (*FADS1*, *ACADM*, *CPS1*, *PHGDH*). The strongest associations were found at *FADS1* and *CPS1* prior to conditioning. At *FADS1* rs188885524 (MAF=1. 7%, Pbefore= 5. 3×10−15) located in 5' untranslated region showed an, albeit decreased, association with PCaaC36:3-to-PCaaC36:4 ratio after conditioning on the GWAS variant rs174547 (Pconditional=3. 2×10−5 , r2=0. 04). At *CPS1*, three independent association signals with Glycine-to-PCaeC38:2 ratio were observed after conditioning on the tag variant rs2216405: rs186668263 (MAF=0. 5%, Pbefore=7. 3×10−14, Pconditional=3. 4×10−13 , r2=0. 0), rs2371008 (MAF=1. 7%, Pbefore=1. 2×10−7, Pconditional=3. 3×10−6 , r2=0. 0) , rs145922665 (MAF=0. 8%, Pbefore=2. 5×10−12, Pconditional=5. 5×10−12 , r2=0. 01). Furthermore, decomposition of metabolite ratios to individual metabolites resulted in additional independent association signals. More specifically, we identified ten additional association signals in *CPS1* with glycine concentrations, of which two variants (rs2371008 (Pbefore=4. 8×10−23, Pconditional=8. 6×10−19) , rs145922665 (Pbefore=1. 2×10−22, Pconditional=4. 9×10−19)) showed stronger association signals with glycine than with Glycine-to-PCaeC38:2 ratio. In conclusion, high-resolution fine-mapping of previously reported GWAS signals identified eight novel secondary variants associated with fasting metabolite concentrations. Replication of these findings will be sought.

814F

White matter tract laterality indices as endophenotypes to assess genetics of brain asymmetry. *M. S. Naslavsky[1,3], S. A. M. Ezquina[1], C. Malcher[1], G. L. Yamamoto[1], Y. A. O. Duarte[1], M. L. Lebrão[1], M. Thiebaut De Schotten[2], M. Catani[2], E. Amaro Jr. [1,3], M. Zatz[1].* 1) University of Sao Paulo, Sao Paulo, SP, Brazil; 2) Institute of Psychiatry, King's College London, London, UK; 3) Brain Institute, Albert Einstein Hospital, Sao Paulo, SP, Brazil.

Brain laterality is a phenotype common to many animals in which several functions are activated asymmetrically by each hemisphere. In humans, handedness and language are two main examples of lateralized phenotypes in the brain. Dominance for handedness and language is leftward in most humans and it is not fully mirrored in left-handers, but the distribution of this phenotype is more dispersed than in right-handers. The genetic components behind human brain laterality are largely unknown and contribute with medium to small overall effect. Disruption of brain asymmetry patterns is associated with neuropsychological disorders such as dyslexia, schizophrenia and autism. *LRRTM1* is a maternally imprinted gene that was previously associated with lateral hand skill measures in a group of schizophrenic/schizoaffective individuals. The aim of this study was to verify enrichment of variants in *LRRTM1* gene in extreme phenotypes of arcuate fasciculus laterality indices. These indices were derived from white matter fractional anisotropy (FA) values obtained by diffusion tensor imaging in magnetic resonance of a subgroup of a Brazilian population-based sample which were exome-sequenced. The laterality indices based on FA values in four pairs of white matter tracts (arcuate fasciculus: long, anterior and posterior segments; and frontal aslant tracts) were derived and showed a normal skewed distribution. Gene variant enrichment analysis was tested with sequence kernel association test (SKAT) clustering the extremes of the distribution of the arcuate fasciculus long segment and *LRRTM1* variants, with n=30 per extreme ends. Although no significant p-value was obtained, one *LRRTM1* missense polymorphism (rs6733871, p. Asn330Ser) present in two copies in combination with other three rare variants showed a trend of enrichment in leftward lateralization, including non-right-hander individuals. This variant is described with frequency >0. 2 among different populations and exclusive to humans but highly conserved among mammals. Additional analyses with the other language-related white matter tracts laterality indices and *LRRTM1* variation are in progress to provide evidence on *LRRTM1* role in brain asymmetry. Further variant collapsing methods in a genome-wide scale may be useful to identify gene candidates as key components of development and maintenance brain asymmetry.

815W

***TPH2* variants are associated with the Metabolic Syndrome and Obesity.** *V. Causer, S. Reis, I. Halder.* University of Pittsburgh, Pittsburgh, PA.

Tryptophan hydroxylase 2 (*TPH2*) gene encodes a protein that catalyzes the first and rate limiting step in the serotonin pathway and *TPH2* variants have been associated with various psychological outcomes including depression. Studies have shown a bidirectional relationship between depression and metabolic syndrome. We hypothesized that *TPH2* variation may also be associated with Metabolic Syndrome (MS) and/or its components. Study subjects included 771 European Americans recruited in the University of Pittsburgh HeartSCORE study. Subjects were between 45-74 years in age, 64% female, and represented all Framingham risk strata. Information on individual MS components, demographic measures and genotypes for ~49,000 SNPs (as part of an Illumina gene chip) were available for all participants. We analyzed 24 *TPH2* SNPs on the gene chip. Each SNP was in Hardy Weinberg Equilibrium and had minor allele frequency >0. 05. Using Haploview we identified 5 linkage blocks and selected one known tag SNP from each block: rs7963803, rs11179002, rs4760750, rs4760820, and rs12231356 to capture 99% of the variation in *TPH2*. Using PHASE, we estimated five haplotypes in our sample that had frequencies > 0. 05. MS was modeled as a latent variable unifying the individual components of the syndrome. In structural equation models (SEM) adjusted for age and gender (implemented in MPlus) we found two haplotypes CGCCG (Hap1) and CGCGG (Hap2) to be significantly associated with the latent MS variable (p=0. 003, 0. 013). We confirmed the associations with Hap1 and Hap 3 by testing against a continuous metabolic risk score (MRS; defined as mean of standardized MS components) and each MS component (Blood pressure, Waist circumference, Triglycerides, HDL and glucose) after adjusting for appropriate covariates. Hap1 was significantly associated with MRS (p=0. 001) and with waist (p < 0. 001), but not with other MS components. Of the five tag SNPs only rs4760820 was significantly associated with MRS (<0. 001) and waist circumference (p= 1. 62x10-6). In four SNP long moving window haplotypes using all 24 *TPH2* SNPs both MRS and waist circumference remained significantly associated with all four haplotypes containing rs4760820 (MRS: P = 0. 0003, waist: P = 0. 0001). Our results indicate that *TPH2* variations which have previously been associated with depression are also associated with MS primarily through effects on central adiposity.

816T

Metabochip study of body composition measures in black South Africans reveals novel insights into obesity risk loci. *Z. Lombard[1,2], V. Pillay[3], L. M. Hendry[1], A. Choudhury[2], H. Soodyall[3], S. A. Norris[4], N. J. Crowther[5], M. Ramsay[2] for the AWI-Gen Research Group, as members of The H3Africa Consortium.* 1) School of Molecular & Cell Biology, Faculty of Science, University of the Witwatersrand, Johannesburg, South Africa; 2) Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa; 3) Division of Human Genetics, School of Pathology, Faculty of Health Sciences, National Health Laboratory Service & University of the Witwatersrand, Johannesburg, South Africa; 4) MRC/Wits Developmental Pathways for Health Research Unit, Department of Paediatrics, School of Clinical Medicine, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; 5) Department of Chemical Pathology, School of Pathology, Faculty of Health Sciences, National Health Laboratory Service & University of the Witwatersrand, Johannesburg, South Africa.

Obesity is a predisposing factor for many diseases, and results from an interplay between genetic and environmental factors. Strong evidence exists for the role of genetic variation in the etiology of obesity and body composition, and although these findings have been widely replicated, there remains a paucity of data from African populations. Likewise, the heritability of body composition measures has not been comprehensively investigated in Africans. We sought to establish whethergeneticvariants on the Illumina Metabochip array were associated with different body composition measures in black South Africans. Additionally, the heritability of these traits in this population was assessed. Participants from the Birth to Twenty cohort study, a longitudinal birth cohort from South Africa, were the focus of this study, and included 1240 participants (median age = 17. 9 years) and 1033 of their female caregivers (median age = 40 years) (total N=2273). The association between SNPs and a given body composition measure was estimated using linear regression, with adjustment for covariates (sex, age and height) using PLINK 1. 9. Body composition measures investigated included anthropometric measures commonly used to define obesity, such as BMI, waist- and hip circumference and waist-to-hip ratio (WHR). In addition, dual-energy X-ray absorptiometry (DXA)-derived whole body composition measures were evaluated, including fat mass, lean-tissue mass and percentage body fat. Narrow-sense heritability was estimated using genotyped SNP data with GCTA. Strong associations observed in these analyses, are between variants near the *SEC16B* gene with fat mass (P=9. 48 x 10-7; b=1. 79kg) and percentage body fat (P=7. 9 x 10-7; b=1. 35kg). These data suggest the presence of a universal risk locus near *SEC16B* for body fat mass, as it has previously been associated with obesity measures in European, African American and Asian populations. Sex-specific associations are observed in males only near *TRPM7* with both hip- (P=1. 28 x 10-8; b=1. 80cm) and waist (P=6. 19 x 10-8; b=1. 64cm) circumference, and in females only in *FTO* (P=2. 91 x 10-5; b=0. 02) with WHR. Preliminary heritability results are comparable to estimates derived from Europeans. These results provide evidence that the genetic etiology of obesity-related traits may differ by ancestry and sex, and these factors should therefore be considered when assessing the genetic determinants of body composition and related disorders.

817F

Association of rs6967330, a functional variant of *CDHR3*, with asthma exacerbation and rhinovirus infections in African Americans. *M. March[1], P. M. A Sleiman[1], J. McElwee[2], D. Diogo[2], F. Mentch[1], K. Thomas[1], C. Hou[1], C. Kim[1], H. Hakonarson[1].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Merck Research Laboratories, Department of Genetics & Pharmacogenomics (GpGx), Boston, MA.

Asthma is a chronic inflammatory condition of the lungs, characterized by acute episodes of breathing difficulties and hypersensitivity of the lungs to a variety of common environmental stimuli and allergens. Among the most common causes of hospitalization during childhood are acute asthma exacerbations. Considerable effort has been spent examining the genetic underpinnings of the asthma phenotype; very few studies have been conducted investigating the contribution of genetic variation to asthma exacerbations. A recent GWA study (Bønnelykke *et al.* , 2014) associated a coding variant of the *CDHR3* (Cadherin-related family member 3) gene with an asthma exacerbation phenotype defined by hospitalizations in a Danish childhood cohort. It was also shown that cell surface expression of the CDHR3 protein was enhanced in the disease-associated variant, which contains a cysteine-to-tyrosine substitution in the extracellular domain. A recent functional study (Bochkov *et al.* , 2015) showed that the disease associated variant of the CDHR3 protein greatly enhanced infection rates of rhinovirus C in a cell culture model system. This observation provides an excellent model explaining the GWA result, as rhinovirus C infection is a common cause of asthma exacerbation. To further investigate this hypothesis, we chose to investigate the relevant *CDHR3* variant (rs6967330) in the cohort assembled at the Center for Applied Genomics at the Children's Hospital of Philadelphia. We analyzed rs6967330 in our African American cohort (approximately 10,700 subjects) using a phenome-wide association study (PheWAS) based on ICD9 codes. This analysis revealed that rs6967330 is associated with the PheWAS phenotype "Asthma with exacerbation" to a phenome-wide significant level ($p < 9.0 \times 10^{-5}$), a phenotype for which we have 1688 cases and 6285 controls. To follow up the PheWAS result, we compiled a case/control cohort for phenotypes associated with rhinovirus infections based on ICD9 codes associated with diagnoses. Despite the small size of our cohort (178 cases), we found that rhinovirus infection associated to a nominal level ($p < 0.05$) with the genotype at rs6967330 in African American children. The results demonstrate that the indicated variant of CDHR3 is relevant to asthma exacerbations in African Americans, a previously unknown association, and the data provide support for the interaction of this variant with rhinovirus infection in triggering asthma exacerbations.

818W

Tissue and sex-specific gene expression patterns during pubertal development. *H. Hou[1,2], L. Uusküla-Reimand[1], M. Makarem[1,3], C. Corre[1], A. Metcalf[1], C. Bellissimo[1], A. Goldenberg[1,5], M. Palmert[1,4], M. Wilson[1,2].* 1) Department of Genetics and Genome Biology, Hospital for Sick Children, Toronto; 2) Department of Molecular Genetics, University of Toronto; 3) Faculty of Medicine, University of Toronto; 4) Department of Pediatrics, Hospital for Sick Children, Toronto; 5) Department of Computer Science, University of Toronto.

Puberty is a postnatal developmental process achieved by all reproductively competent adults. Age of puberty onset varies among the general population and is associated with adverse health outcomes, including risks for breast cancer, cardiovascular disease and behavioral disorders. Studies of rare mutations in patients with pubertal disorders have revealed a complex regulatory system involving the hypothalamic-pituitary-gonadal (HPG) axis. Recent GWAS studies identified more than 100 loci associated with age at menarche (AAM); however only a few of these loci fall within or near genes implicated in rare pubertal disorders (*MKRN3*, *LEPR*, *TACR3*, *PCSK1*). Little is known about the mechanisms by which these loci affect pubertal timing.

We first asked if genes associated with AAM GWAS loci and disorders of puberty could give us insight into the cell types and tissues that influence pubertal timing. Using transcriptome data from more than 75 normal human tissues/cell types, we found that the expression of these puberty-associated genes was significantly enriched for the pituitary. Interestingly the second most enriched tissue was the pineal gland. For example gene expression levels of *TMEM38B* (the second most statistically significant GWAS hit), *CA10*, *RXRG*, *PLCL1* were most highly expressed in the pineal gland.

We isolated HPG axis tissues, the pineal gland and liver at five developmental time points that span the pubertal transition in male and female mice. Using high-throughput qPCR we then profiled the expression of 183 mouse orthologs of genes known to underlie pubertal disorders and/or are implicated in the regulation of pubertal timing by GWAS. We confirmed previously reported temporal expression profiles such as the decrease in hypothalamic expression of *Mkrn3* prior to puberty. We also found evidence of pineal gland expression for several genes including *Tmem38b*. Overall, the number of significant changes in gene expression between adjacent time points was greatest prior to puberty. Excluding the gonads, pituitary gene expression was found to be the most dynamic between sexes and, interestingly, many of the pituitary sex differences became apparent at the onset of puberty and increased thereafter. Overall our results give new insights into the temporal, tissue and sex specific expression patterns of genes that regulate pubertal timing.

819T

Polymorphisms of genes involved in extracellular matrix homeostasis of tendons and the risk of rotator cuff tears. *M. F Leal[1,2], E. A. Figueiredo[1], P. S. Belangero[1], C. Cohen[1], L. C. Loyola[1,2], M. C. Smith[2], C. V. Andreoli[1], A. C. Pochini[1], B. Ejnisman[1], M. Cohen[1].* 1) Departamento de Ortopedia e Traumatologia, Universidade Federal de São Paulo, 04038-032, São Paulo, SP, Brazil; 2) Disciplina de Genética, Departamento de Morfologia e Genética, Universidade Federal de São Paulo, 04023-900, São Paulo, SP, Brazil.

Rotator cuff tears is common orthopedic disease. The tears occurs in part due to a degenerative process in tendons of the shoulder. The histopathological changes associated with rotator cuff tears include thinning and disorganization of collagen fibers. Several studies, including from our research group, have been demonstrate the deregulated expression of genes involved in the extracellular matrix homeostasis in the ruptured tendon. Recently, two studies showed that genetic variants might have a role in the tears etiology. Therefore, polymorphisms in genes involved in the extracellular matrix homeostasis can contribute to an extended tendon degeneration process, which explains why subsets of patients are more susceptible to rotator cuff tears. In this study, we evaluated if the polymorphisms rs6728999 (*FN1*), rs2104772 (*TNC*), rs1800012 (*COL1A1*), rs3196378 (*COL5A1*), rs11568785 (*TGFBR1*), rs1590 (*TGFBR1*), rs17731 (*KLF6*), rs243865 (*MMP2*), rs243866 (*MMP2*), rs679620 (*MMP3*), rs522616 (*MMP3*), rs17577 (*MMP9*), rs2252070 (*MMP13*), rs4898 (*TIMP1*) and rs2277698 (*TIMP2*) were associated with the risk of rotator cuff tears. We genotyped 210 patients with full-thickness rotator cuff tears and paired 411 controls (individuals without this disease and without any familiar history of rotator cuff tears) using TaqMan inventoried assays in a real-time PCR system. The Chi-square test was used to compare the genotypes differences between groups. Univariate and multivariate logistic regression were performed to investigate whether the studied polymorphisms may contribute to the risk of rotator cuff tears risk. The frequencies of genotypes were in the Hardy-Weinberg equilibrium, except for the polymorphism rs4898. The rare variants of rs17731 (adjusted p = 0. 005) and of rs3196378 (adjusted p = 0. 025) were associated with the risk of rotator cuff tears by multivariate logistic regression (the p-value was adjusted by gender and age). Therefore, these genetic variants of *KLF6* and of *COL5A1* may have a role in the rotator cuff tears etiology.

820F

HLA association analysis and haplotype analysis identify susceptible alleles of HLA class II to tuberculosis. *L. Toyo-oka[1,2], S. Mahasirimongkol[2], H. Yanai[1,3], T. Mushiroda[4], S. Wattanapokayakit[2], N. Wichukchinda[2], N. Yamada[5], S. Nedsuwan[6], P. Kantipong[6], A. Takahashi[7], M. Kubo[8], P. Sawanpanyalert[9], K. Tokunaga[2].* 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Tokyo, Japan; 2) Department of Medical Science Center, National Institute of Health, Department of Medical Sciences, Ministry of Public Health, Nonthaburi, Thailand; 3) Fukujuji Hospital, Japan Anti-Tuberculosis Association (JATA), Kiyose, Japan; 4) Laboratory for Pharmacogenomics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 5) Research Institute of Tuberculosis, Japan Association of Tuberculosis, JATA, Kiyose, Japan; 6) Chiangrai Prachanukroh Hospital, Ministry of Public Health, Chiang Rai, Thailand; 7) Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 8) Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 9) Food and Drug Administration, Ministry of Public Health, Nonthaburi, Thailand.

Tuberculosis (TB) is chronic respiratory infection with high burden to global health especially in Africa and south-east Asia. Genome-wide association study (GWAS) in Thai and Japanese populations identified at-risk locus in 20q12 in young-onset (< 45 years old) TB. Human leukocyte antigen (HLA) class II is considered to play important roles for TB based on its pivotal roles in host interactions with M. tuberculosis and its strong association with leprosy, an infectious disease caused by M. Leprae. However, previous association studies of HLA class II are usually under power due to small number of cases investigated. *HLA-DQB1*05:03* was suggested as strong risk factor for tuberculosis in Cambodia with additional functional confirmation. In Thai population, *HLA-DQB1*05:02*, *HLA -DQB1*06:01* and *HLA -DQB1*03:01* was reported with 82 cases and 160 controls. In GWAS, not all *HLA-DRB1* alleles could be captured by SNP genotyping due to complexity of this genomic region. The purpose of this study is to investigate the associations of HLA class II, *HLA-DRB1* and *HLA-DQB1* with TB. Subjects were recruited through a TB surveillance system in Chiang Rai province in Thailand after obtaining individual informed consent. Diagnosis of TB was by clinical features with microbiological confirmation. HIV seropositive individuals were excluded. In summary, 396 TB patients and 282 controls were selected and their blood samples were drawn for DNA extraction. *HLA-DRB1* and *HLA-DQB1* were determined by PCR-SSO (sequence-specific oligonucleotide probes), WAKFlow. We conducted HLA association analysis and haplotype analysis utilizing alleles and haplotypes with frequencies of >5% and obtained corrected *p*-values (*Pc*) adjusted by Bonferroni correction. In HLA association analysis, *HLA-DRB1*16:02* and *HLA-DRB1*14:01* showed lowest *Pc* of 0. 002 with odds ratio (OR) at 2. 02 (1. 36-3. 06) in dominant model. Moreover, *HLA-DRB1*14:01* was associated with *Pc* of 0. 042 with OR of 0. 58 (0. 39-0. 86) in dominant model while none of *HLA-DQB1* alleles showed association in additive, dominant and recessive model. Haplotype analysis revealed that *DRB1*16:02 - DQB1*05:02* (*Pc*: 0. 002 and OR: 2. 02 (1. 36-3. 06) in dominant model) and *DRB1*14:01 -DQB1*05:02* (*Pc*: 0. 013 and OR: 0. 55 (0. 37-0. 82) in dominant model) were risk and protective haplotypes, respectively. These results suggest that *HLA-DRB1* is more critical than *HLA-DQB1* because *HLA-DQB1*05:02* did not affect the tendency and magnitude of OR of *HLA-DRB1* allele.

821W

Survey of Rare and Common Genetic Variation in Putative Risk Loci for Primary Open-angle Glaucoma. *R. P. Igo[1], J. N. Cooke Bailey[1], W. K. Scott[2], R. J. Sardell[2], M. A. Pericak-Vance[2], L. R. Pasquale[3], M. A. Hauser[4], T. Gaasterland[5], J. L. Wiggs[3], J. L. Haines[1], NEIGHBORHOOD Consortium.* 1) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 3) Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, MA; 4) Department of Ophthalmology, Duke University Medical Center, Durham, NC; 5) Scripps Genome Center, Scripps Institution of Oceanography, University of California, San Diego, San Diego, CA.

Glaucoma, a progressive degenerative disease of the eye, affects over three million people in the United States and is the second leading cause of blindness. The late-onset form of the disease has a complex mode of inheritance: more than a dozen genetic risk loci for glaucoma have been found to contribute modest effects, but the greater part of the heritability still remains to be explained. To identify new genetic determinants for glaucoma, and to better characterize suspected risk loci, we genotyped 6,044 unrelated primary open-angle glaucoma (POAG) cases and controls of European descent within the NEIGHBORHOOD Consortium for over 240,000 mostly exonic genetic variants on the Illumina HumanExome BeadChip, enriched for variation likely relevant to POAG by the addition of 15,800 custom variants in genomic regions implicated in glaucoma. Here, we report results from single-variant and gene-level tests of association between POAG and 301 diallelic polymorphisms within and near 14 genes in previously reported risk loci, on a sample of 3,208 POAG cases and 2,836 controls. Association analysis for single common variants was conducted by logistic regression, adjusting for age at recruitment, sex, DNA source and three principal components for ancestry, using an additive genetic model (on the log odds ratio scale). For rare variation, gene-based tests using SKAT-O were performed including all variation with minor allele frequency less than 0. 1 or 0. 02, using the default weighting scheme. Association at studywide significance (nominal $p < 0.00017$) was detected for one or more common variants in *AFAP1, ATXN2, CAV1, CDKN2B-AS1, SIX6* and *TMCO1,* including missense variants in *ATXN2* (W262R) and *SIX6* (H141N). Genes *FNDC3B* and *AFAP1* contained multiple rare coding variants enriched in either cases (OR > 2) or controls (OR < 0. 5), and thus had suggestive SKAT-O results ($p = 0.005$ and 0. 008 respectively). Insight into the role of both rare and common variation at POAG risk loci will improve our understanding of the genetic basis of this common eye disease. Grant support: NIH/NEI R0122305; 1X01HG007486-01.

822T

Variants in the HLA-DQB1 region confer susceptibility to Age-Related Macular Degeneration. *E. Jorgenson[1], R. Melles[2], T. Hoffmann[3, 4], X. Jia[5], L. Sakoda[1], M. Kvale[3], Y. Banda[3], C. Schaefer[1], N. Risch[1, 3, 4], L. Shen[1].* 1) Division of Research, Kaiser Permanente, Oakland, CA; 2) Department of Ophthalmology, Kaiser Permanente Northern California Redwood City Medical Center, Redwood City, CA 94063, USA; 3) Institute for Human Genetics, University of California San Francisco, San Francisco, CA, 94143 USA; 4) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA 94143, USA; 5) Department of Neurology, University of California San Francisco, San Francisco, CA 94143, USA.

Age-related Macular Degeneration (AMD) risk variants in the complement system point to the important role of immune response and inflammation in the pathogenesis of AMD. While the Human Leukocyte Antigen (HLA) region plays a central role in regulating immune response, previous studies of genetic variation in HLA genes and AMD have been limited by sample size or incomplete coverage of the HLA region by first generation genotyping arrays and imputation panels. Here, we conducted a large-scale HLA fine-mapping study with 4,841 AMD cases and 23,790 controls of non-Hispanic white ancestry from the Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. Genotyping was conducted using custom Affymetrix Axiom arrays, with dense coverage of the HLA region. Classic HLA polymorphisms were imputed using SNP2HLA, which utilizes a large reference panel to provide improved imputation accuracy of variants in this region. We examined a total of 6,937 SNPs and 172 serotypes, incorporating established AMD risk variants in a forward-stepwise conditional analysis, which revealed novel associations with two nonsynonymous SNPs, rs9274390 and rs41563814 (OR=1. 21; p=1. 40×10-11) corresponding to amino acid changes at position 66 and 67 in *HLA-DQB1*, respectively, and the DQB1*02 serotype (OR=1. 20; p=3. 9×10-10) with the risk of AMD. Similar effects were observed with non-exudative AMD and CNV AMD subtypes. The effect of rs9274390 was consistent whether it was carried on the DQB1*02 haplotype or not, suggesting that the association of this SNP was primary. These findings support a role of HLA class II alleles on the risk of AMD.

823F

Heterogeneity in expression of HLA class II genes provides evidence for unexpected functions of their products. *E. Hesselberg, M. Houtman, K. Shchetynsky, L. Padyukov.* Department of Medicine, Karolinska Institutet, Stockholm, Sweden.

Introduction. The strongest genetic risk factor for developing rheumatoid arthritis (RA) was described within the *HLA-DRB1* locus. Several genetic variations cumulatively defined as shared epitope (SE) alleles were repeatedly shown to have especially high influence in the development of seropositive RA. However, there is very little knowledge regarding different products from the genes in this locus. We therefore investigated *HLA-DRB1* RNA expression for alternative splicing isoforms. **Methods.** Generic PCR based on flanking exons of HLA-DRB1 was run on cDNA from human PBMCs, osteoclasts, B-cell line, and fibroblasts. Amplicons were sequenced and the most robust product was detected using qPCR on a set of whole-blood cDNA from a cohort of RA patients and healthy controls (n = 95 + 95) for the validation of the isoform. **Results.** We detected a non-canonical 219bp-length isoform lacking premature stop codons across different cell types. The short isoform lacked regions translating Ig domain and antigen-binding domain. This isoform was repeatedly detected in healthy and sick individuals. Interestingly, this isoform was detected in different types of cells, including professional and non-professional antigen presenting cells. **Conclusion.** A short *HLA-DRB1* isoform exists in multiple cell types and its functionality should be further studied. The absence of antigen-binding domain indicates the possibility of unexpected functions of the product from this gene that should be investigated in the context of autoimmune diseases.

824W

Identification and functional analysis of chronic pancreatitis-associated deep intronic variants in the SPINK1 gene. *WB. Zou[1,2,3], A. Boulling[1,2], E. Masson[1,4], Z. Liao[3], ZS. Li[3], JM. Chen[1,2,5], C. Férec[1,2,4,5].* 1) Institut National de la Santé et de la Recherche Médicale (INSERM), U1078, Brest, France; 2) Etablissement Français du Sang (EFS) - Bretagne, Brest, France; 3) Department of Gastroenterology, Changhai Hospital, the Second Military Medical University, Shanghai, China; 4) Laboratoire de Génétique Moléculaire et d'Histocompatibilité, Centre Hospitalier Universitaire (CHU) Brest, Hôpital Morvan, Brest, France; 5) Faculté de Médecine et des Sciences de la Santé, Université de Bretagne Occidentale (UBO), Brest, France.

Over the past two decades, genetic variants have being increasingly reported to be associated with or cause hereditary, familial and idiopathic forms of chronic pancreatitis. However, to date, mutational analyses of the disease-causing or -predisposing genes (e. g. PRSS1, SPINK1, CTRC) have focused on the coding sequences and exon/intron boundaries. Given the certain presence of functional variants in deep intronic sequences, this practice must underlie a fraction of the "missing heritability" of chronic pancreatitis. As a first step to address this issue, we performed Sanger sequencing of the intronic sequences of the SPINK1 gene in a subset of French patients with familial chronic pancreatitis; all of these patients had previously been found not to carry a known disease-predisposing variant. We found deep intronic variants in 13 patients. To determine whether these variants are of pathogenic relevance or not, we sought to perform RT-analysis of each variant in the context of the entire genomic sequence of the SPINK1 gene (i. e. from the translation initiation codon to the termination codon). We have successfully introduced most of the deep intronic variants into the pcDNA3. 1/V5-His-TOPO vector harboring the entire SPINK1 gene by means of directed mutagenesis. The resulting mutant vectors were used to transfect HEK 293 cells; and total RNA were extracted for RT- PCR analysis. One (i. e. c. 87+363A>G) of the so far tested variants was found to cause aberrant splicing. Findings from this analysis will have implications for designing mutation screening strategies in clinical practice as well as for interpreting data that will be obtained from the application of next-generation sequencing.

825T

Modeling Autism by Candidate Gene Editing in Human Induced Pluripotent Stem Cells. *E. Deneault, K. Zaslavsky, R. K. Yuen, T. O. Thompson, J. Ellis, S. W. Scherer.* Sickkids, Toronto, Ontario, Canada.

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder that is poorly understood at the gene level. As live neurons from human brain are not easily accessible, we generated a series of new iPS cell lines derived from fibroblasts obtained from members of seven different ASD families affected by loss-of-function mutations in ASD-associated genes. The seven candidate genes affected by these mutations are *AGBL4*, *CAPRIN1*, *CNTN5*, *DLGAP2*, *EHMT2*, *KAL1* and *NRXN1*. We used the double-nicking type II CRISPR/Cas9 system to correct mutations in patient-derived iPS cells, hence providing isogenic controls to highlight the role of these genes in autism. Guide RNA (gRNA) sequences were devised in order to minimize the likelihood for off-target binding. In the presence of a single-stranded oligonucleotide (ssODN) template, cellular HDR machinery replaced the disease mutations for the wild-type sequences. Edited alleles were detected using specific DNA probes along with absolute quantification by droplet digital PCR (ddPCR). The frequency of modified alleles was amplified to 100% using sibselection steps in a 96 well format (Miyaoka Y, Nat. Methods, 2014;11[3]). We also adapted this editing system to specifically knock-out 8 additional ASD-associated genes, for which skin fibroblasts were not available to derive iPS cells from ASD patients. These include *AFF2*, *ASTN2*, *ATRX*, *CACNA1C*, *CHD8*, *KCNQ2*, *SCN2A* and *TENM1*. We hypothesized that artificial disruption of a candidate gene in iPS-derived neurons would promote a similar phenotype to that observed in iPS-derived neurons from probands. We preferentially targeted the earliest exon that is common to the different transcripts for each candidate gene. We inserted a 60bp DNA fragment that includes all-frame termination codons, a rare restriction site (*Mre*1) and a V5 epitope tag to possibly reveal truncated forms of proteins. This stop tag is synthesized as ssODN flanked by homology arms that are specific to each candidate gene. All iPS cell lines generated above are currently cultured in vitro under conditions favouring neuronal differentiation. RNAseq, immunostaining and electrophysiology will shed light on the effects of mutated ASD-associated genes on neuronal expression patterns and identity, dendritic complexity and synaptic function.

826F

Prevalence of HLA-DQ alleles, *PTPN22* gene functional variant R620W, and three autoantibodies in Kuwaiti Arab children with Type-1 Diabetes Mellitus. *M. Z. Haider[1], M. A. Rasoul[1,2], S. A. Mathews[1], H. Al-Kandari[3], M. Al-Mahdi[4], J. Sukumaran[1], G. S. Dhaunsi[1,2].* 1) Department of Pediatrics, Faculty of Medicine, Kuwait University, Safat, Kuwait; 2) Department of Pediatrics, Mubarak Al-Kabeer Hospital, Jabriya, Kuwait; 3) Department of Pediatrics, Farwania Hospital, Farwania, Kuwait; 4) Department of Pediatrics, Adan Hoapital, Al-Adan, Kuwait.

A complex interaction between susceptibility genes, immune mediators and environmental factors play a critical role in genetic predisposition to type-1 diabetes mellitus (T1DM). We have determined the prevalence of HLA-DQ alleles, protein tyrosine phosphatase N22 (*PTPN22)* gene C1858T functional variant and three autoantibodies in Kuwaiti children with T1DM. This study included 239 Kuwaiti children with TIDM and 133 controls (healthy and ethnically matched). The diagnosis of T1DM was based on the ISPAD criteria. HLA-DQ alleles were determined by sequence-specific PCR in 178 patients. The genotypes for *PTPN22* gene variant C1858T (R620W) were identified by PCR-RFLP. The presence of autoantibodies (ICA, INS and GAD) were determined by radioimmunoassay. Nine different combinations of HLA-DQ alleles were detected in patients. In 55% patients, the genotype was either homozygous for DQ2 or in combination with a DQ8 allele. In 36% patients, the genotype was homozygous DQ8 or with other alleles. Collectively, 91% of the patients had either DQ2 or DQ8 alleles. The variant genotype of the *PTPN22* gene was detected in homozygous/heterozygous combination in 39% patients compared to 27% in controls. The homozygous TT-genotype was detected in 8% patients compared to 0. 99% in controls (p <0. 001). In patients with TT-genotype of *PTPN22* gene, 93% had at least one DQ2 allele and 60% carried either a DQ2 or a DQ8 allele. In T1DM patients with TT-genotype, GAD autoantibody was detected in 83%, INS-Ab in 67% and ICA-Ab in 54% cases respectively. Our data demonstrate that the HLA-DQ2/DQ8 alleles and the variant T-allele of *PTPN22* gene constitute significant determinants of genetic predisposition to T1DM in Kuwaiti Arabs.

827W

Revealing the functional effect that relates *PTPN2* SNPs to rheumatoid arthritis. *M. Houtman, K. Shchetynsky, L. Padyukov.* Rheumatology Unit, Department of Medicine Solna, Karolinska Institute and University Hospital, Stockholm, Sweden.

Background: Rheumatoid arthritis (RA) is a common chronic autoimmune disorder that has a strong genetic component. Over 100 risk loci have been confirmed in genome-wide association studies and the major risk factor are HLA-DRB1 shared epitope (SE) alleles. Outside the HLA region, one of the recently identified candidate genes for RA is protein tyrosine phosphatase non-receptor type 2 (*PTPN2*), which was previously also shown in association with T1D. However, the functional consequences of genetic variations in the *PTPN2* region remain undefined. We aimed to understand the functional mechanisms that connect variations in the *PTPN2* region with the risk of RA development. **Methods:** We conducted an association study in the RACI cohort (6573 seropositive RA cases and 15870 controls, Eyre *et al.* , 2013), computed gene-gene interactions using attributable proportion for SE alleles with 21 SNPs within *PTPN2*, performed RNA expression analysis for *PTPN2* in blood and analyzed DNA methylation data (689 individuals, Liu *et al.* , 2013). The expression cohorts comprise 69 RA cases and 77 healthy controls (whole blood) and 44 RA cases and 43 healthy controls (peripheral blood mononuclear cells). **Results:** The interaction analysis between SNPs within *PTPN2* and HLA-DRB1 SE alleles identified an additional RA-associated SNP: rs657555 (meta-analysis in RACI cohort, p = 3. 35e-06; OR = 1. 1352; Q = 0. 3993). The expression of total *PTPN2* and *PTPN2* splice variants, *TC45* and *TC48*, did not vary significantly in individuals with different rs657555 genotypes. In addition, the expression of other genes from the locus within 1 Mb from rs657555 were not different in relation to genotypes in our and publicly available data. We identified a significant association independent of disease status between rs657555 genotypes and four CpG sites close to *PTPN2*: cg23544223, cg23598886, cg09945482, and cg24737193 (all p < 5e-9). These CpG sites are located in a CpG island 7 kb downstream of *PTPN2* and provide a potential functionally important link between associated genotypes and RA. **Conclusion:** This study of a complex autoimmune disease integrates genetic association and interaction of HLA SE alleles with the architecture of the *PTPN2* locus, expression and DNA methylation data and could provide a novel insight into disease mechanisms.

828T

Uncommon*PGRN* deletion and FTD responsible for phenotype variability in three familial cases. *E. Vitale[1], S. Napoletano[1], S. Pappata'[2], M. T. Gentile[3], L. Colucci-D'Amato[3], G. Della Rocca[4], A. Maciag[5], A. Puca[5], D. Grossi[6], G. Milan[7], A. Postiglione[8].* 1) Institute of Protein Biochemistry (IBP), CNR-National Research Council, Napoli, Napoli, Italy; 2) Institute of Bioimaging and Biostructures , CNR, Naples-Italy; 3) Department of Environmental, Biological, Pharmaceutical Science and Technology, Second University of Naples, Caserta-Italy; 4) Villa Camaldoli Foundation Clinic, Naples, Italy; 5) Department of Medicine - University of Salerno- Salerno, Italy; 6) Department of Psychology, Second University of Naples, Caserta, Italy; 7) Geriatric Clinic "Frullone" ASL Napoli 1 - Naples, Italy; 8) Department of Clinical Medicine & Surgery, University of Naples "Federico II", Naples-Italy.

Mutations in Progranulin (PGRN) gene have been genetically associated with fronto-temporal dementia (FTD), although the neurobiology of this secreted glycoprotein is still unclear. We identified three familial cases having a rare +5 deletion in exon six of PGRN gene (g. 101349_101355delCTGCTGT) as causative of frameshift mutation. This mutation is not considered frequent and already described in two apparently sporadic case of FTD only, but never associated with familial cases. Our patients showed also heterogeneous clinical phenotypes, such as behavioral variant (bv-FTD) in men and primary progressive aphasia (PPA) in women. Quantitative RT-PCR of *PGRN* gene expression in WBC shows an unusual increasing of expression, while protein analyses of plasma concentration detected a PGRN protein deficiency. Using allele-specific PCR mutation-primers (ARMS), this mutation was searched in three hundreds healthy controls matched by age, sex and geographic regions and no deletion was ever found. The findings provide convincing evidence of the putative role of *PGRN* in the genetic etiology of FTD and show a link between FTD and ex 6-del *PGRN* mutations, which probably are more frequent than previously considered. The described mutation express two different gender linked phenotypes bv-FTD in men and PPA in women. Although we believe that this is the product of a founder effect, we yet cannot prove at this stage. Our findings imply that PGRN is essential for neuronal survival and even partial loss of PGRN eventually leads to neurodegeneration.

829F

Role of endothelial nitric oxide synthase (*eNOS*) G894T polymorphism in ischemic stroke susceptibility: A case-control study in North Indian population. *A. Kaur[1], K. kaur[1], A. Uppal[2].* 1) Department of Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India; 2) Uppal Neuro Hospital, Amritsar, Punjab, India.

Introduction: Genetic variations in *eNOS* gene might have some functional implication in pathogenesis of ischemic stroke through reduced basal NO production and may affect individual's susceptibility to ischemic stroke. The only common variation (G894T) in *eNOS* gene leading to an amino acid change in the mature protein is guanine to thymine substitution at exon 7 replacing glutamate by aspartate at position 298. *eNOS* protein containing Asp at position 298 to may be subjected to selective proteolytic cleavage in endothelial cells and vascular tissues, leading to reduced nitric oxide synthase activity. **Objective:** The present study aims to elucidate the association of G894T variant of eNOS gene with ischemic stroke using case-control approach. **Methodology:** The case-control study included total of 239 individuals comprising of 139 patients (74 males and 65 females) with mean age of 62. 84±11. 36, from Uppal Neuro hospital Amritsar, clinically and radiologically diagnosed with ischemic stroke and 100 healthy controls (48 males and 52 females) with mean age of 58. 19±10. 74, validated by Questionnaire for verifying Stroke Free Status. All the subjects were genotyped using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) technique and the data so obtained was statistically analyzed. **Results:** The frequency of *eNOS* 894T allele was slightly higher inpatients as compared to controls (16% vs 8. 55%, p=0. 075). Both the cases and controls were found to be in Hardy Weinberg equilibrium (cases: χ^2=3. 010, p=0. 083; Controls: χ^2=0. 013, p=0. 91). Among the three genetic models, only co-dominant model (TT vs GT / GT vs GG) showed mild increased risk (OR=1. 66, 95% CI, 0. 95-2. 89, p=0. 064) of ischemic stroke susceptibility in patients. Overall, the distribution of *eNOS* G894T genotypes differed insignificantly between patients and controls (χ^2=3. 891, p=0. 143). There was no significant risk of developing disease as revealed by odds ratio (OR) (crude OR; GT: OR=1. 47, 95% CI, 0. 76–2. 83; TT: OR=4. 82, 95% CI, 0. 57–40. 87) in carriers of variant genotype from the reference group (GG). **Conclusion:**Although the co-dominant model showed weak association of *eNOS* G894T variant with ischemic stroke, overall,this polymorphism couldn't predict the risk of ischemic stroke in North Indian population, which can be attributed to small sample size of the study group.

830W

The type 1 diabetes susceptibility gene *CLEC16A* restrains NK cell function by modulating expression of NK receptors via C Vps-HOPS complex, CART and autophagy. *R. Pandey[1], M. Bakay[1], S. Yoeun[1], J. Kushner[2], J. S. Orange[3], H. Hakonarson[1,4].* 1) The Centre for Applied Genomics, Children's Hospital Of Philadelphia, Philadelphia, PA; 2) Section of Pediatric Diabetes and Endocrinology, Department of Pediatric Medicine, Endocrine-Metabolism, Texas Children's Hospital, Houston, TX; 3) Section of Immunology, Allergy, and Rheumatology, Department of Pediatric Medicine, Texas Children's Hospital, Houston, TX; 4) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

Type 1 Diabetes is a multi-factorial childhood disease with a strong genetic component. Several GWAS had showed association of CLEC16A with T1D, which encodes a protein of unknown function. We reported previously that protective alleles of CLEC16A [A/A] are associated with higher levels of mRNA. We hypothesize that CLEC16A functions in NK cells to restrain secretory functions including cytokine release and cytotoxicity. Therefore, defining the role of CLEC16A in NK cells will provide insight into the pathogenesis of T1D. To address this, we investigated the expression of CLEC16A in human immune cells and NK cell lines on mRNA level by RT-PCR and on protein level by Western blot. We examined the consequence of knockdown of CLEC16A in NK cell lines and ex-vivo NK cells using siRNA and over-expression of this protein in NK cell line with a GFP expression reporter using a retroviral expression system. Cytotoxicity and INF-γ production were decreased in NK cells over-expressing CLEC16A. siRNAs mediated knockdown resulted in 70% reduction of Clec16A protein levels and 35% increase in cytotoxicity compared to control siRNAs. Clec16a over-expressing NK cells formed smaller number of conjugates for all time points in conjugation assay with no difference in CD107a expression. Sub-cellular localization studies revealed cytosolic localization. Mechanistic studies through reciprocal co-immunoprecipitation experiments reveals that CLEC16A associates with Vps16A, a subunit of class C Vps-HOPS complex and CART: an Hrs/Actinin-4/BERP/Myosin V Protein Complex, and modulates receptor recycling and expression through expansion of autophagosomes by ubiquitin-like conjugation pathways ending with Atg5. Clec16a knockdown in UBC-Cre-Clec16aloxP mice resulted in increase in NK cell cytotoxicity of the YAC-1 targets in comparison to control. Phenotypic analysis of splenocytes from TAM and control littermates revealed up regulation of NKp46, Ly49H and CD226 activating receptors. Splenocytes of Clec16aloxP knockdown TAM mice exhibit decreased mitochondrial membrane potential. Taken together, our results indicate that CLEC16A serves a role in restraining two major functions of NK cells, cytotoxicity and cytokine release. Our mechanistic studies reveal that *CLEC16A* restrains NK cell functions by modulating expression of NK receptors via C Vps-HOPS complex, CART and autophagy**.** Owing to its role in autophagy, Clec16a impacts the function of NK cells and the risk of autoimmunity.

831T

Capture Hi-C reveals a novel causal gene, *IL20RA*, in the pan-autoimmune genetic susceptibility region 6q23. *G. Orozco[1], A. McGovern[1], S. Schoenfelder[2], P. Martin[1], K. Steel[3], J. Massey[1], D. Plant[1], P. Fraser[2], A. Barton[1], J. Worthington[1], S. Eyre[1].* 1) Arthritis Research UK Centre for Genetics and Genomics, University of Manchester, Manchester, United Kingdom; 2) Nuclear Dynamics Programme, The Babraham Institute, Cambridge, United Kingdom; 3) Immunobiology Department. King's College London. New Hunt's House, Guy's London, United Kingdom.

The majority of genetic variants that predispose to complex diseases map to non-coding enhancer regions. There is evidence supporting a role of these non-coding variants in transcriptional regulation through long-range interactions with their target genes. The chromosomal region 6q23 is associated with a number of autoimmune diseases. Associated SNPs lie a large distance from any gene (>180kb). The aim of this work was to identify causal disease genes at the locus by studying long range chromatin interactions. The recently developed Capture Hi-C (CHi-C) method was used to characterize chromatin interactions involving the 6q23 region, in a hypothesis free manner, at high resolution in human T and B cell lines. Both the intergenic region associated with disease and all promoters within 500kb of associated SNPs were targeted. Fragments containing the disease associated intergenic SNPs displayed strong interactions with three distinct regions: the *IL20RA* gene, upstream of *IFNGR1* and a region downstream of the *TNFAIP3* gene containing the lncRNAs RP11-10J5. 1 and RP11-240M16. 1. Additionally, *IL20RA* interacts with these lncRNAs and the promoter of the *TNFAIP3* gene. *TNFAIP3* also demonstrated interactions with the same lncRNAs. A number of interactions were cell type specific, and they were all validated using 3C-qPCR. The most strongly associated SNP in the 6q23 region, rs6920220, is in tight LD (r2<0. 8) with eight other SNPs. Using RegulomeDB, we determined the most plausible causal SNP to be rs6927172, as it maps to an enhancer in both B and T cells, is in a DNase hypersensitivity cluster, shows transcription factor binding and is in a conserved region. To further characterize the functional effect of rs6927172 we used human primary T cell gene expression data to examine expression quantitative trait loci. The risk allele was modestly associated with an increased expression of *IL20RA* mRNA (p=0. 03). Chromatin immunoprecipitation demonstrated binding of chromatin marks of active enhancers (H3K4me1 and H3K27ac) to rs6927172. In conclusion, we have compelling evidence that the autoimmune risk variant, rs6927172, is within a complex gene regulatory region, involving *IL20RA*, *TNFAIP3* and regulatory elements, such as lncRNAs. These results show that CHi-C can help identify GWAS causal genes and suggest novel therapeutic targets; indeed, anti-IL-20 monoclonal antibody therapy has recently been shown to be effective in the treatment of rheumatoid arthritis.

832F

The genetic architecture of human ear morphology. *K. Adhikari, A. Ruiz-Linares, CANDELA Consortium.* Genetics, Evolution and Environment, University College London, London, London, United Kingdom.

The human pinna, consisting primarily of cartilage, is highly variable within and across populations, as anthropological studies trying to characterize pinna shape in major continental populations have provided conflicting findings over the years. The ear lobe, not consisting of cartilage but of areolar and adipose tissue, is a modern human trait not present in other primates. So far the genetic architecture of morphological variation in human pinna is poorly understood. In the Candela project comprising over 5,000 volunteers from five Latin American countries, we scored a set of classical anthropological traits on the outer ear and observed genome-wide significant associations for lobe size, lobe attachment, folding of antihelix, helix rolling, ear protrusion and antitragus size. Seven genomic regions were identified, including a functional variant in the *EctodysplasinA receptor (EDAR)* gene, a key regulator of embryonic skin appendage development, and an evolutionarily conserved binding site for the transcription factor Cartilage paired-class homeoprotein 1 (*CART1*). Varying effect size, direction and prevalence of these variants explain why traditional anthropological studies could not observe consistent continental patterns so far. Associations were followed up with functional studies including replication in mouse models. Human pinna shape is now being used for forensics and user identification in electronic equipment, where these findings can be particularly interesting.

833W

Genetic variants associated with gene expression in the lung and asthma susceptibility. *J. C. Berube[1], E. Lavoie-Charland[1], N. Gaudreault[1], L. Sbarra[1], C. Henry[1], D. S. Postma[2], D. D. Sin[3], K. Hao[4], D. C. Nickle[5], W. Timens[2], P. D. Paré[3], M. Laviolette[1], L. P. Boulet[1], Y. Bossé[1,6].* 1) Institut universitaire de cardiologie et de pneumologie de Québec, Quebec City, Canada; 2) University of Groningen, University Medical Center Groningen, GRIAC Research Institute, Groningen, Netherlands; 3) The University of British Columbia Centre for Heart Lung Innovation, St Paul's Hospital, Vancouver, Canada; 4) Icahn School of Medicine at Mount Sinai, New York, NY, United States; 5) Merck & Co, MRL, Seattle, Washington, United States; 6) Department of Molecular Medicine, Laval University, Quebec City, Canada.

Pathogenic changes occurring in airways of asthma patients generate chronic symptoms such as cough, shortness of breath, chest tightness and wheezing. Studying genetic regulation of lung transcriptome may lead to the identification of new variants associated with asthma susceptibility. The top 160 SNPs associated with gene expression in the lung of 1,111 lung tissue samples patients were selected and genotyped in 965 individuals of the Quebec City Case-Control Asthma Cohort. Allele frequencies were compared between asthma patients and controls. Ten SNPs were found to be significantly associated with asthma (P <0. 05). Rs17138154 on chromosome 7 was the SNP most strongly associated with asthma (P = 1. 26E-03). This protective asthma allele was associated with lower expression of *LOC493754* in the lung (eQTL p value = 1. 87E-45). The lung eQTL-SNP rs3807807 associated with the expression of *ICA1* (eQTL p value = 1. 84E-46) was also associated with asthma (P = 1. 08E-02). In this case, the asthma risk allele showed less expression of *ICA1* in the lung. Rs2249828, regulating the expression of *PIGP* in the lung (eQTL p value =4. 18E-122), was also associated with asthma (P = 1. 71E-02). Two SNPs in linkage disequilibrium with rs2249828 (rs2032088, r2 = 1. 0; rs2835624, r2 = 0. 92), showed an association with asthma in the meta-analysis of the European GABRIEL Consortium (P <0. 05). This project took advantage of a large-scale lung eQTL dataset to identify new functional genetic variants associated with asthma. The identification of lung eQTLs associated with asthma is not only elucidating new susceptibility genes, but also providing vital information about the genetic mechanisms leading to abnormal function of the genes that promote the development of asthma.

834T

Temporal Expression Profiling Identifies Pathways Mediating Effect of Causal Variant on Phenotype. *S. Gupta[1], A. Radhakrishnan[1], P. Raharja-Liu[2], G. Lin[3], L. Steinmetz[3,4,5], J. Gagneur[2], H. Sinha[1].* 1) Department of Biological Sciences, Tata Institute of Fundamental Research, Mumbai, India; 2) Gene Center, Ludwig-Maximilians-Universitét, Munich, Germany; 3) European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany; 4) Department of Genetics, Stanford University School of Medicine, Stanford, California, United States of America; 5) Stanford Genome Technology Center, Stanford University, Palo Alto, California, United States of America.

Even with identification of multiple causal genetic variants for common human diseases, understanding the molecular processes mediating the causal variants' effect on the disease remains a challenge. This understanding is crucial for the development of therapeutic strategies to prevent and treat disease. While static profiling of gene expression is primarily used to get insights into the biological bases of diseases, it makes differentiating the causative from the correlative effects difficult, as the dynamics of the underlying biological processes are not monitored. Using yeast as a model, we studied genome-wide gene expression dynamics in the presence of a causal variant as the sole genetic determinant, and performed allele-specific functional validation to delineate the causal effects of the genetic variant on the phenotype. Here, we characterized the precise genetic effects of a functional *MKT1* allelic variant in sporulation efficiency variation. A mathematical model describing meiotic landmark events and conditional activation of *MKT1* expression during sporulation specified an early meiotic role of this variant. By analyzing the early meiotic genome-wide transcriptional response, we demonstrate an *MKT1*-dependent role of novel modulators, namely, *RTG1/3*, regulators of mitochondrial retrograde signaling, and *DAL82*, regulator of nitrogen starvation, in additively effecting sporulation efficiency. In the presence of functional *MKT1* allele, better respiration during early sporulation was observed, which was dependent on the mitochondrial retrograde regulator, *RTG3*. Furthermore, our approach showed that *MKT1* contributes to sporulation independent of Puf3, an RNA-binding protein that steady-state transcription profiling studies have suggested to mediate *MKT1*-pleiotropic effects during mitotic growth. These results uncover interesting regulatory links between meiosis and mitochondrial retrograde signaling. In this study, we highlight the advantage of analyzing allele-specific transcriptional dynamics of mediating genes. Applications in higher eukaryotes can be valuable for inferring causal molecular pathways underlying complex dynamic processes, such as development, physiology and disease progression. .

835F

Lung disease-associated SNPs and lung eQTLs in the major histocompatibility complex (MHC). *M. Lamontagne[1], C. Couture[1], DS. Postma[2], DD. Sin[3], K. Hao[4], D. Nickle[4], PD. Pare[3], W. Timens[2], M. Laviolette[1], Y. Bossé[1,5].* 1) Centre de recherche de l'Institut universitaire de cardiologie et de pneumologie de Québec, Québec, Canada; 2) University of Groningen, University Medical Center Groningen, GRIAC research institute, Groningen, The Netherlands; 3) The University of British Columbia Center for Heart Lung Innovation, St. Paul's Hospital, Vancouver, BC, Canada; 4) Merck & Co. Inc. , Rahway, New Jersey, USA; 5) Département de Médecine Moléculaire, Université Laval, Québec, Canada.

The extended Major Histocompatibility Complex (xMHC) is a key locus for the immune system. This region was associated with lung diseases (asthma, cystic fibrosis, interstitial lung disease, lung cancer) and pulmonary function in previous GWAS. As for many GWAS-nominated loci, the causal gene(s) underpinning the genetic association with lung diseases and function in the xMHC locus remain unknown. This study has two goals. First, identify lung expression Quantitative Trait Loci (eQTLs) within the xMHC. Second, identify new lung diseases susceptibility genes by overlaying lung eQTL results in the xMHC with previously reported lung function and respiratory disease SNPs in this region. Non-tumor lung specimens and blood-DNA samples from 1,038 patients recruited at three academic sites were included in this study. The xMHC was delimited by the *HIST1H2AA* and *KIFC1* genes on chromosome 6 including 6,872 genotyped SNPs and 271 probe sets for eQTL analyses. Lung eQTLs were identified in the discovery set (Laval n=409; P-value ≤ 2. 68e-08) and then replicated in two replication sets (UBC n=287; Groningen n=342; P-value ≤ 0. 05). The GWAS Catalog was downloaded and GWAS-SNPs located in the xMHC and associated with lung diseases were selected. A total of 5,790 significant lung eQTLs were detected in the xMHC. These eQTLs consisted of 2,807 SNPs and 66 associated probe sets testing 50 transcripts. The majority of eQTLs were local (n= 5,516), but 274 distant-acting eQTLs were also found. eQTL-SNPs were pruned to 663 independent SNPs (LD with r2 < 0. 8). Among the 28 GWAS SNPs associated with lung diseases in the xMHC, 14 were associated with the expression levels of at least one transcript. Nineteen unique eQTLs were found with SNPs previously associated with asthma (6 SNPs; 8 transcripts), two with interstitial lung disease (1 SNP; 2 transcripts), four with lung cancer (2 SNPs; 4 transcripts) and 13 with pulmonary function (5 SNPs; 10 transcripts). Three genes were associated with GWAS SNPs from two traits/diseases: *AGPAT1* – asthma and interstitial lung disease, *HLA-DQB1* – asthma and interstitial lung disease, *ZFP57* – lung cancer and pulmonary function. This study highlights possible causal genes underlying the genetic association with lung diseases and pulmonary function in the xMHC region.

836W

Impact of common variation at diabetes locus*MTNR1B* on sleep, circadian, and melatonin physiology. *J. M. Lane[1,2,3], AM. Chang[2,3,4,5], D. Aeschbach[2,4,7], SW. Cain[2,4], CA. Czeisler[2,4], EB. Klerman[2,4], SW. Lockley[2,4,8], M. StHilaire[2,4], SA. Shea[2,4], JF. Duffy[2,4], OM. Buxton[2,4,5,6], S. Redline[2,4], FAJL. Scheer[2,4], R. Saxena[1,2,3], Care Consortium.* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Brigham and Women's Hospital, Division of Sleep and Circadian Disorders, Boston, MA; 3) Broad Institute, Program in Medical and Population Genetics, Cambridge, MA; 4) Harvard Medical School, Division of Sleep Medicine, Boston, MA; 5) Pennsylvania State University, Department of Biobehavioral Health, University Park, PA; 6) Harvard School of Public Health, Department of Social and Behavioral Sciences, Boston, MA; 7) German Aerospace Center, Institute of Aerospace Medicine, Cologne, Germany; 8) Monash University, School of Psychological Sciences, Clayton, Victoria, Australia.

Genome-wide association studies of type 2 diabetes suggest a causal link between sleep (*MTNR1B* rs10830963*)* and circadian rhythms (*CRY2* rs11605924*)* and development of diabetes. The causal mechanism of these associations, however, is unknown. In this study, we aimed to determine how both common and rare genetic variation in Melatonin-receptor 1B (*MTNR1B*) predisposes to risk of type 2 diabetes. We undertook three experimental approaches. First, given that prior studies point towards pancreatic islets as the site of action, we tested the effect of the *MTNR1B* ligand, melatonin, on glucose-stimulated insulin secretion in both rat beta cell lines and human islets. Second, we tested for an association between *MTNR1B* rs10830963 and measures of circadian rhythm timing (phase), magnitude (amplitude), length (period) and melatonin physiology using hourly plasma melatonin concentrations and body temperature collected over ≥24 hours in a unique cohort of subjects of European ancestry from intensive in-laboratory sleep studies (n= 193). Third, we tested for an interaction between objectively measured wake time and rs10830963 on glucose metabolism in multi-ethnic subjects from MESA (n=1,501). Our experimental results confirmed previously reported suppression of glucose-stimulated insulin secretion during acute and chronic melatonin administration in an *MTNR1B* dependent fashion in rat pancreatic beta-cells and human islets. Second, the *MTNR1B* variant significantly associated with a delayed timing of melatonin offset (Beta(SE) 1. 37h (0. 55h), p=0. 015) and a longer duration of melatonin secretion (41min (19min), p=0. 032) in our studies, demonstrating that risk allele carriers have altered melatonin physiology. Finally we found nominal evidence of a SNP x wake-time interaction (p=0. 024) effect on risk of type 2 diabetes. Taken together, our results are consistent with a mechanism whereby *MTNR1B* diabetes risk allele carriers have suppressed morning insulin response due to elevated melatonin levels, and food intake concurrent with elevated melatonin levels during this time adversely affects glucose metabolism. This points towards the need for genotype-based intervention trials using light (to suppress melatonin production) or changes in meal timing (to reduce melatonin during glycemic load). Therefore, our work defining mechanisms by which this variant influences melatonin physiology can now be explored for direct translational potential, an important goal for T2D genetics.

837T

The role of conditional genetic effects in determining human longevity. *S. Ukraintseva, K. Arbeev, L. Arbeeva, A. Kulminski, D. Wu, I. Akushevich, E. Stallard, A. Yashin.* Duke University, Durham, NC.

One of the paradoxes in genetics of human longevity is that many clinically relevant variants for major diseases (cancer, CHD, AD, etc.) are often present in genomes of longest-lived people in similar or sometimes even higher frequencies as in individuals with shorter lifespans. Here we discuss several biological mechanisms for the conditional effects of genes on aging and health related phenotypes which together could explain why carrying the genetic "risk factors" may not necessarily compromise longevity. These mechanisms include: (i) trade-off-like effects of genes on two or more health traits; (ii) age-specific influence of genes on vulnerability to death; (iii) gene-gene interaction; and (iv) gene-environment interaction, among other factors. Review of current knowledge suggests that the conditional effects of genes on aging and health related phenotypes are common and may result in situations in which disease "risk allele" can also be a pro-longevity variant.

838F

Integrative *In-Silico* Functional Characterization of Celiac Disease Susceptibility Variants. *N. A. Shaik[1], O. Rashidi[2], B. Babajan[2], J. Al-Aama[1,2], R. Elango[2].* 1) Genetic Medicine, King Abdulaziz University, Jeddah, Saudi Arabia, Select a Country; 2) Princess Al-Jawhara Center of Excellence in Research of Hereditary Disorders, King Abdulaziz University, Jeddah, Saudi Arabia.

Background & Aim: Celiac disease (CD) is a gluten intolerance disorder with an important genetic component. Recent GWAS and fine mapping studies in European CD patients have identified up to 57 non-HLA CD susceptible loci, majority of which lie either in non-coding or intergenic regions. Deciphering their pathogenic relevance to CD, through functional biology assays is a challenging task. Therefore, this study objective was to adopt integrative computational methods to uncover the plausible mechanisms through which CD specific loci contribute to the disease. **Methodology:** In this study, we investigated 1065 LD variants (r2 ≥0. 8) of 57 CD lead susceptible loci using CADD, GWAVA, RegulomeDB and FATHMM tools to rank them from deleterious to benign levels based on integrative scoring system developed. The mRNA levels of corresponding various ranked genes were tested in publicly available gene expression profiles of intestinal tissues and lymphocyte data. Pathway analysis was done to identify the potential pathways that are likely to be influenced by highly deleterious CD loci. **Results:** The integrative prediction analysis of 1065 LD variants revealed 22 SNPs (2% of 1065 LD) lying in 13 genes as highly deleterious (Rank I), followed by 46 SNPs (4. 3% of 1065 LD) of 29 genes which are deleterious (Rank II). Of the six genes with positive immunological function and altered expression levels, only *CCR2, CSK* genes fall under Rank 1 category. These two contribute to pathways which are essential to positively regulate the in vivo activation of leukocytes and T cell lymphocytes, which may further trigger the cascade of immunological events responsible for intestinal gluten intolerance. **Conclusion:** Using *in-silico*functional prediction tools, this study found 2 SNPs ranked in 1st category can contribute to celiac disease through pathways which are not well implicated previously. Although experimental validation is needed, our study shows the effective utility of integrative scoring system in prioritizing non-coding SNPs based on their corresponding gene expression levels, as a useful starting point to screen for putatively celiac causal SNPs from the crowd of GWAS data.

839W

Examining the causal effect of Vitamin D on childhood caries: A Mendelian Randomization study. *T. Dudding[1,2], SJ. Thomas[2], K. Duncan[2], DA. Lawlor[1], NJ. Timpson[1].* 1) MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol, Avon, United Kingdom; 2) School of Oral and Dental Science, University of Bristol, Bristol, Avon, United Kingdom.

Dental caries is the localized destruction of susceptible tooth tissues by acidic by-products from bacterial fermentation of dietary products. Although caries in primary teeth is decreasing in the United Kingdom, it is still prevalent with 31% of 5 year olds having obvious decay experience in 2013. Previous studies have reported an inverse association between vitamin D and childhood dental caries, but whether this is causal is unclear. Mendelian randomisation (MR) is an analytical method used to provide evidence for causal associations between potentially modifiable risk factors and health outcomes. We undertook an MR study, using genetic variants known to influence circulating 25-hydroxyvitamin D (25(OH)D) levels, in 5,544 European origin children from the South West of England, to determine the causal effect of circulating 25(OH)D on dental caries. Data on caries and related characteristics were obtained from parental completed questionnaires between 38 and 91 months and clinical assessments in a random 10% sample at 31, 44 and 61 months. In multivariable confounder adjusted analyses we found no strong evidence for an association of 25(OH)D with caries experience, severity or onset, or having a general anesthetic (GA) for dental problems. Known relationships between vitamin D and *CYP2R1, DHCR7* and *GC* were replicated and when combined these SNPs were confirmed as a strong genetic instrument for vitamin D with per allele changes of a clinically relevant magnitude. These variants also showed no associations with factors potentially confounding the vitamin D – dental caries relationship. In MR analysis the odds ratios per 10 nmol/L increase in 25(OH)D were 0. 88 (95% confidence interval (CI): 0. 78, 1. 00; *P*=0. 054) for caries experience and 0. 94 (95% CI: 0. 73, 1. 21; *P*=0. 62) for GA. Our MR study suggests that there might be an inverse causal effect of 25(OH) D on dental caries. However, our estimates are imprecise and a larger study is required to determine a robust effect. With the use of a genetic instrument, a life-course vitamin D exposure model was used to test associations between vitamin D and dental caries in a population that had poorly suited vitamin D measures. In contrast to a recent meta-analysis we found little convincing evidence of an association between vitamin D and caries experience, onset or severity. This study highlights the opportunity to apply genetics based causal methods to dental epidemiology.

840T

Allele-specific enhancers mediate associations between *LCAT* and *ABCA1* polymorphisms and HDL cholesterol levels. *A. D. Howard[1], X. Wang[1], M. Prasad[1], A. D. Sahu[2], R. Aniba[2], S. Hannenhalli[2], Y. C. Chang[1].* 1) University of Maryland School of Medicine, Baltimore, MD 21201, USA; 2) Center for Bioinformatics and Computational Biology, University of Maryland, College park, MD 20742, USA.

For any given trait the majority of associated SNPs identified through genome-wide association studies (GWAS) lie within noncoding regions and are likely to mediate associations via allele-specific enhancer activity. In order to identify likely causal regulatory SNPs underlying the GWAS associations, we developed a bioinformatics pipeline that harnesses epigenetic signatures of enhancers to identify putative enhancers that contain a SNP in linkage disequilibrium (LD) with a GWAS SNP. We applied our approach to GWAS for HDL cholesterol levels. For functional validation, we focused on 8 putative enhancers that contain SNPs associated with HDL cholesterol. Three of the eight putative enhancers demonstrated enhancer activity in liver-derived cell lines HepG2 and Huh7. These enhancers are within introns of the *LCAT* and *ABCA1* genes, which encode lecithin-cholesterol acyltransferase and ATP binding cassette transporter A1, respectively, and play crucial roles in HDL biogenesis. Additionally, two of the eight putative enhancers demonstrated enhancer activity in HEK293 cells. Not only are these putative enhancers cell line-specific and size-specific, more importantly, their *in vitro* enhancer activities are also allele-specific. For example, while the HDL-associated SNPs span over 188 kb, the putative enhancer (743-bp) within *LCAT* containing the rs1109166 T-allele demonstrated 3- and 15-fold higher activity in HepG2 and Huh7, respectively, compared to the same sequence harboring the C-allele. To further elucidate the mechanism of the enhancer-promoter interaction, we sought to determine whether these putative enhancers are in close physical proximity to the promoters of their respective genes, *in vivo*. Using chromosome conformation capture we found that one of the predicted enhancers is in physical proximity to the *ABCA1* promoter. Ongoing efforts are focused on exploring the biological consequences these polymorphic enhancers may have on promoter activity and subsequent perturbation of the HDL biogenesis pathway. Our results demonstrate that through this approach, the regions that contain GWAS signals, often hundreds of kilobases in size with multiple SNPs serving as statistical proxies to the true functional site, can be narrowed down to much smaller regions and, importantly, provide an experimentally testable hypothesis for the underlying mechanism linking genetic variants to complex traits.

841F

A functional regulatory variant associated with type 2 diabetes and fasting glucose levels at the *ADCY5* locus. *T. S. Roman[1], B. Wolford[2], R. Welch[3], M. Morken[2], M. L. Buchkovich[1], Y. Wu[1], S. Khetan[4], A. Uyar[4], R. Kursawe[4], D. Ucar[4], S. C. J. Parker[5], F. S. Collins[2], M. L. Stitzel[4], K. L. Mohlke[1], NISC Comparative Sequencing Program.* 1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 4) The Jackson Laboratory for Genomic Medicine, Farmington, CT; 5) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI.

Genome-wide association studies (GWAS) have identified more than 76 loci associated with type 2 diabetes (T2D), however many of the molecular and biological mechanisms are unknown. Many of the variants identified by GWAS are located within noncoding or intergenic regions, suggesting that these variants may influence gene regulation. One association signal for both T2D and fasting glucose levels is located within introns of *ADCY5*. The A allele of rs11708067 has been associated with higher fasting glucose levels and increased risk of T2D ($P<5x10-8$). *ADCY5* encodes adenylate cyclase 5, which catalyzes the production of cyclic AMP (cAMP) from ATP (Vatner 2013). Intracellular cAMP, ATP, and Ca2+ play crucial roles in insulin secretion in pancreatic beta cells (Shibasaki 2004). A previous study established a role for *ADCY5* in glucose metabolism and demonstrated that human islets from individuals homozygous for the rs11708067-A T2D risk allele showed decreased *ADCY5* expression compared to islets from heterozygous donors (P<. 05, Hodson 2014). We sought to identify the functional variants responsible for the association signals. Of the 14 variants in strong linkage disequilibrium (r2> . 8, 1000 Genomes phase 1 CEU) with rs11708067, only rs11708067 overlaps both FAIRE and DNase peaks in human pancreatic islets, suggesting that it could have regulatory function in islets. rs11708067 also overlaps a strong enhancer chromatin state in human islets (Parker and Stitzel 2013). We tested rs11708067 for allelic differences in enhancer activity by cloning a 231-bp candidate regulatory element in both orientations with respect to a minimal promoter luciferase vector and performing transcriptional reporter assays in 832/13 rat insulinoma cells. We observed 1. 3 to 3-fold enhancer activity for rs11708067-G, and lower activity for rs11708067-A, with significant differences between the alleles in both orientations ($P=. 0017$, $P=. 03$). Furthermore, we observed an allelic imbalance in H3K27ac ChIP-seq reads overlapping rs11708067 (26 reads containing rs11708067-G vs. 3 reads containing rs11708067-A, $P=7. 4x10-6$) and evidence of differential protein binding in EMSAs. RNA-seq data from 82 published (Fadista 2014) and 32 new islet samples showed decreased *ADCY5* expression for rs11708067-A ($P=6. 5x10-13$). Taken together, these data suggest that the rs11708067-A risk allele acts to decrease *ADCY5* expression, which may lead to an increase in fasting glucose levels and increased risk of T2D.

842W

Understanding the genetic involvement in Retinopathy of Prematurity by a functional genomics and protein profiling. *I. Kaur[1], S. Rathi[1], S. Jalali[2], H. Syed[1], GR. Musada[1], P. Chhablani[3], R. Kekunaya[3], D. Balakrishnan[2], S. Chakrabarti[1].* 1) Kallam Anji reddy Molecular Genetics lab, Prof Brien Holden Eye Research Centre, L V Prasad Eye Institute, Hyderabad, N/A = Not Applicable, India; 2) Smt. Kannuri Santhamma Centre for Vitreo Retinal Diseases, L V Prasad Eye Institute, Hyderabad, India; 3) Jasti V Ramanamma Children's Eye Care Centre, L V Prasad Eye Institute, Hyderabad, India.

ROP a vasoproliferative condition of eye, is a leading cause of blindness among preterm infants with low birth weight (BW) and gestational age (GA). The underlying genetic mechanism contributing to ROP are unclear. The present study aimed to identify gene(s) responsible for ROP by a functional genomics and protein profiling. **Methods:** 384 SNPs spanning 27 candidate genes were screened in a cohort of 400 preterm infants (GA ≤35weeks and BW ≤1700g) including 200 clinically characterized cases of ROP and 200 No-ROP preterm infants by customized genotyping. Allele and genotype frequencies, linkage disequilibrium and haplotype analysis were done to delineate the ROP-associated variants. Global gene expression profiling was performed in 15 preterms infants with different stages of ROP and 9 No-ROP and 3 full term infants with no retinal disease by using Illumina bead chip array having ~47,000 transcripts. The expression profiles were analyzed by the Genome Studio-software and fold change was computed. Meta-analysis of differentially expressed genes were performed by Gene Ontology (GO) and pathway analysis using Panther software. The vitreous humor levels of 27 candidate proteins were assessed in patients with severe ROP (n=30) and congenital cataract (n=30) by multiplex bead arrays. Differentially expressed proteins were further validated by western blotting and zymography. **Results:** Significant differences were noted in the alleles and correponding haplotype frequencies of few SNPs in *TSPAN12, CFH, C2/BF, IHH*, and *MMP9*, between cases and controls that withstood Bonferroni correction for multiple testing (p=1. 3x10-4). Other studied genes did not exhibit any association to ROP. Compared to controls, ROP patients exhibited significant (p<0. 01) increased vitreous levels of MMP9, CFH, C3, C4, Prealbumin, SAP, APO A1 and APOC3. These results were further validated by western blotting and zymography. A total of 142 genes were differentially expressed (115 upregulated and 27 downregulated) between the ROP patients and no-ROP preterm controls. Of these, 19 genes expressed differentially with >2 fold change. Major pathways identified in ROP pathogenesis using differentially expressed genes included inflammation mediated by chemokine and cytokine, endothelin and toll like receptor signaling. **Conclusions:** Significant genetic associations and differential gene expression observed in ROP suggests the possible underlying mechanisms in the development and progression of ROP.

843T

A Lupus-associated variant in purine nucleoside phosphorylase (*PNP*) causes cell cycle abnormalities. *Y. Ghodke-Puranik[1], J. M. Dorschner[1], D. M. Vsetecka[1], S. Amin[2], M. Makol[2], F. Ernste[2], T. Osborn[2], K. Moder[2], V. Chowdhary[2], M. A. Jensen[1], T. B. Niewold[1].* 1) Department of Immunology and Division of Rheumatology, Mayo Clinic, Rochester, MN; 2) Division of Rheumatology, Mayo Clinic, Rochester, MN.

Introduction: Systemic lupus erythematosus (SLE) is a multi-system, autoimmune disease characterized by autoantibodies to nucleic acids and nucleosomal proteins. The type I interferon pathway is dysregulated in SLE and IFN-αlevels are high in patients. We performed a genome-wide association study and found a missense SNP (C297T:gly51ser) in the *PNP* gene that associates with high IFN levels in SLE (rs1049564; P=1. 24 x10-7). PNP is a key enzyme of purine metabolism. PNP deficiency leads to dysregulated deoxynucleotide levels, a block in DNA synthesis, and defective immunity. Previously, the rs1049564 SNP was thought not to impact PNP function. To show a functional link between rs1049564 and SLE, we performed in vitro experiments using HAPMAP cell lines and patient leukocyte samples with and without the PNP variant. **Methods:** To determine if the rs1049564 variant alters PNP function, we exposed 6 HAPMAP cell lines (2 homozygous (CC), 2 heterozygous, and 2 homozygous (TT)) to a dose curve of deoxyguanosine (dGuo; nucleotide precursor and a substrate for PNP). One day later cells were analyzed for cell cycle phase using Click-iT plus EdU chemistry combined with FxCycle violet stain and flow cytometry. Various pharmacological agents were tested for their ability to reverse the cell cycle block caused by the PNP variant. A similar approach was used to study PNP function in SLE patient samples with and without the rs1049564 variant. Colorimetric enzyme and ELISA assays were used to measure the functional activity of the different PNP isoforms. **Results:** We find that the rs1049564 variant causes increased S phase block and cell death in lymphoblastoid cells. Cell lines homozygous for rs1049564 (TT) had 2 fold increases in S phase block compared to cell lines without the PNP variant (CC). The cell cycle block caused by the PNP variant could be reversed pharmacologically and similar findings were observed in SLE patient cells. **Conclusion:** These results suggest that the rs1049564 *PNP* polymorphism is a loss of function variant that leads to altered PNP function and subsequent S-phase block in select cell subsets within the lymphocyte compartment. This may lead to increased frequencies of circulating apoptotic lymphocytes, and higher type I IFN levels in human SLE. These findings have pharmacogenomic implications, as the S-phase block can be rescued in our in vitro experiments, suggesting a potential for personalized therapeutics.

844F

Deciphering the biological role of *C1orf106*, a susceptibility gene in inflammatory bowel disease. *C. Levesque[1, 2], G. Charron[1], A. Alikashani[1], C. Beauchamp[1], L. Villeneuve[1], S. Foisy[1, 2], G. Boucher[1], P. Goyette[1], J. D. Rioux[1, 2].* 1) Montreal Heart Institute, Montreal, Canada; 2) Université de Montréal, Montreal, Canada.

Inflammatory bowel diseases (IBD, [MIM266600]) involve chronic inflammation of the digestive tract and include ulcerative colitis (UC) and Crohn's disease (CD). IBD may reflect a defect in the homeostasis of immune system and intestinal epithelium. The latter forms a physical and biochemical barrier to commensal and pathogenic microorganisms. A dysfunction in the epithelial barrier may lead to a sustained immune response against bacteria of the gut flora. Genome wide association studies have identified 163 susceptibility regions in IBD. Among these, the *1q32* region associated with risk of both CD (p<2x10-11) and UC (p<6x10-7), contains the gene *C1orf106*. Our targeted re-sequencing study has identified a low-frequency variant, Y333F (p=0. 009) in C1orf106, a protein of unknown function and in which tyrosine333 is predicted to be phosphorylated. Its substitution by a phenylalanine could thus have an impact on its biological activity. For this reason, we prioritized C1orf106 for functional analysis to determine its potential causality. Our microarray analysis of human tissues and cell lines demonstrated that *C1orf106* is mostly expressed in small intestine and colon. It is also detectable in monocytic cell lines but more highly expressed in colonic epithelial cells lines. Furthermore, its expression increased by 40% during differentiation of colonic epithelial Caco-2 cells into polarized epithelium. As it is known that intestinal epithelial cells participate in, and are influenced by inflammatory processes, we examined the impact of IL-1b treatment. Specifically, while IL-1b treatment increased mRNA expression of IL-6 and IL-8 in both proliferative and differentiated Caco-2 cells, it had no effect on *C1orf106* expression suggesting that it is not regulated by IL-1b pro-inflammatory cascade. To provide further biological context, we generated HEK293T cells stably overexpressing C1orf106 and demonstrated that it is colocalized with zonula occludens 1 (ZO-1) tight junction (TJ) marker. TJ play an essential role in the establishment of epithelial barrier and require phosphorylation signals for its assembly. C1orf106 localization at these regions suggests its possible implication in epithelial barrier homeostasis. We are now producing stable overexpression of Y333F in colonic epithelial cells to look for possible changes in C1orf106 localization, TJ assembly and stability. Our results suggest that *C1orf106* may be the causal gene in the *1q32* IBD susceptibility region.

845W

Characterising the causal mechanism at the 5q11 susceptibility locus associated with rheumatoid arthritis. *K. McAllister[1], G. Orozco[1], N. Nair[1], J. Massey[1], A. Anderson[2], J. Diboll[2], J. Isaacs[2], A. Pratt[2], J. Worthington[1], S. Eyre[1].* 1) Arthritis Research UK Centre for Genetics and Genomics, University of Manchester, Manchester, United Kingdom; 2) Institute of Cellular Medicine (Musculoskeletal Research Group), Newcastle University, Newcastle upon Tyne, United Kingdom.

The third strongest association to rheumatoid arthritis (RA) is intronic to the *ANKRD55* gene, on chromosome 5q11, with more compelling candidate genes found in the region. Here well-powered genetic data was used to fine-map the region, and through bioinformatic data and functional experiments, we generate robust evidence for the causal variant and causal gene at the 5q11 locus. As part of the ImmunoChip project, 60 SNPs around the lead GWAS marker at the 5q11 locus, rs6859219, were analysed for their association in 11,475 cases and 15,870 controls. Conditional logistic regression and haplotype analysis were conducted in order to fully characterise the genetic architecture of the region. Bioinformatic tools were used in order to examine and prioritise SNPs based on their regulatory potential. Correlation of the putative variants with expression of nearby genes (eQTL) was carried out in whole blood (n=67), CD4+(n=185) and CD8+ T (n=21) cell subsets ; capture Hi-C was used for the first time in RA to look at potential chromatin interactions and chromatin immunoprecipitation (ChIP) experiments for histone markers of enhancers, H3K4me1 and H3K27ac were carried out in both B and T cell lines. Genetic fine-mapping using ImmunoChip data refined the association at the 5q11 locus to a single signal, rs71624119 (p = 5. 59E-20). Bioinformatic prioritisation of SNPs implicated two SNPs, rs10065637 and rs6859219, in high LD (R2>0. 9) with rs71624119, with the strongest evidence of regulatory activity, located within an intronic enhancer element. eQTL evidence was obtained for correlation of these SNPs with expression, showing only *ANKRD55* in whole blood p =3. 85E-05 (n=67); *ANKRD55* (p = 6. 21E-11) and *IL6ST* (p = 6. 54E-04) in CD4+ and no evidence for eQTL in CD8+ T cells. Capture Hi-C data shows that the disease-associated SNPs are brought into close vicinity of both *IL6ST* and *IL31RA,* over 300kb away. ChIP data was suggestive of a greater enrichment for the histone mark of enhancer activity, H3K4me1, with carriage of the risk allele at both rs10065637 and rs6859219. Although the RA associated loci at 5q11 contains multiple candidates for causality, our preliminary findings strongly implicate the associated SNPs being located in an intronic enhancer element regulating expression of *ANKRD55*, a gene of unknown function, with additional weaker evidence for *IL6ST* regulatory activity in CD4+ cells.

846T

To determine the biological role of GPR65, a gene associated to chronic inflammation. *V. Mercier[1,2], G. Charron[1], D. Devost[3], A. Alikashani[1], C. Beauchamp[1], P. Goyette[1], G. Boucher[1], S. Foisy[1,2], L. Villeneuve[1], T. Hébert[3], J. D. Rioux[1,2].* 1) Montreal Heart Institute, Montreal, Quebec, Canada; 2) Université de Montréal, Montreal, Quebec, Canada; 3) McGill University, Montreal, Quebec, Canada.

Inflammatory bowel diseases(IBD [MIM 266600]), mainly comprising ulcerative colitis (UC) and Crohn's disease (CD), result in the chronic inflammation of the gastrointestinal tract. Genome-wide association studies have associated the *14q31* locus, including the genes *galactosylceramidase (GALC)* and *G protein-coupled receptor 65 (GPR65* [MIM 604620]*)*, to both phenotypes. *GPR65* encodes a pH-sensing G protein-coupled receptor (GPCR). The most associated variant in this region (rs8005161; p=2,35x1014) is correlated (r2=1) to a missense coding variant of *GPR65* (rs3742704: Ile231Leu). We determined that *GPR65* is expressed in lymphoid and mucosal tissues as well as in immune cell lines and human primary immune cells. We also found that its expression is significantly increased in inflamed biopsies from UC patients compared to non-inflamed biopsies and biopsies from healthy controls. Upon activation by low pH, GPR65 stimulates accumulation of cAMP, formation of stress fibers and activation of the RhoA pathway. Thus *GPR65* is a good candidate causal gene. Our objective, therefore, was to define pathways downstream of activation of GPR65 and to evaluate the impact *GPR65**231Leu on these pathways. We used HEK 293 cells stably expressing GPR65 and deficient for either Gas/olf, Gαq/11 or Gα12/13 as it is known that Gαs activates adenylyl cyclase, Gαq/11 leads to the release of intracellular Ca2+, which can also activate certain isoforms of adenylyl cyclase, and that Gα12/13 regulatesactincytoskeletalremodeling. We demonstrated that cAMP accumulation upon activation of GPR65 is, at least partly, due to the Gαs pathway and only slightly or not at all to the Gαq/11 pathway. We have revealed, using bioluminescence resonance energy transfer (BRET)-based biosensors that the Gα12 pathway is also triggered upon acid activation of GPR65. Ongoing analyses of this pathway includes evaluating; (i) stress fiber formation via phalloidin staining, (ii) cellular morphology by microscopy and (iii) activation of small GTPases (RhoA, Cdc42 and Rac1) with G-LISA activation assays using G proteins deficient cells stably expressing *GPR65*. In conclusion, we demonstrate that GPR65 activates both Gαs and Gα12/13 and potentially has a causal role in IBD.

847F

***In vivo* Modeling Implicates *APOL1* in Nephropathy: Evidence for Dominant Negative Effects and Epistasis under Anemic Stress.** *B. R. Anderson[1], D. N. Howell[2], K. Soldano[1], M. E. Garrett[1], N. Katsanis[1], M. J. Telen[3], E. E. Davis[1], A. Ashley-Koch[1].* 1) Center for Human Disease Modeling, Duke University, Durham, NC; 2) Department of Pathology, Division of Pathology Clinical Services, Duke University, Durham, NC; 3) Division of Hematology, Department of Medicine, Duke University Medical Center, Durham, NC.

African Americans have a disproportionate risk for developing nephropathy. This disparity has been primarily attributed to coding variants (G1 and G2) in apolipoprotein L1 (*APOL1*); however, there is little functional evidence supporting the role of this protein in renal function. The G1 allele consists of two nonsynonymous variants in perfect LD, rs73885319 and rs60910145 (encoding S384G and I384M), while the G2 variant consists of a six base pair deletion removing amino acids N388 and Y389 (~21% and ~13% allelic frequency in African Americans for G1 and G2, respectively). Here, we combined genetics and *in vivo* modeling to examine the role of *apol1* in glomerular development and pronephric filtration and to test the pathogenic potential of *APOL1* G1 and G2 variants. Translational suppression or CRISPR/Cas9 genome editing of *apol1* in zebrafish embryos results in podocyte abnormalities and glomerular filtration defects. Complementation of *apol1* morphants with wild-type human *APOL1* mRNA rescues these defects. However, the *APOL1* G1 risk allele does not ameliorate defects caused by *apol1* suppression and the pathogenicity is conferred by the *cis* effect of both individual variants of the G1 risk haplotype (I384M/S342G). *In vivo* complementation studies of the G2 risk allele also indicate that this variant cannot restore normal protein function. Moreover, *APOL1* G2, but not G1, expression alone promotes developmental kidney defects, suggesting a possible dominant-negative effect of the altered protein. In sickle cell disease (SCD) patients, we reported previously a genetic interaction between *APOL1* and *MYH9*. Testing this interaction *in vivo* by co-suppressing both transcripts yielded no additive effects. However, upon genetic or chemical induction of anemia, we observed a significantly exacerbated nephropathy phenotype. Furthermore, concordant with the genetic interaction observed in SCD patients, *APOL1* G2 reduces *myh9* expression *in vivo*, suggesting a possible interaction between the altered APOL1 and *myh9*. Our data indicate a critical role for *APOL1* in renal function that is compromised by nephropathy-risk encoding variants. Moreover, our interaction studies indicate that the *MYH9* locus is also relevant to the phenotype in a stressed microenvironment and suggest that consideration of the context-dependent functions of both proteins will be required to develop therapeutic paradigms.

848W

Association analysis of *TNF*α (-308G>A) polymorphism in Polycystic Ovary Syndrome cases. *R. Kaur[1], S. Pahuja[1], A. Beri[2], A. Kaur[1].* 1) Department of Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India; 2) Beri Maternity Home, Amritsar, Punjab, India.

Introduction: Chronic low grade inflammation is closely related with the incidence of polycystic Ovary Syndrome (PCOS). Tumor necrosis factor alpha (TNFα), which is a pro-inflamatory cytokine is an important mediator of inflammation and is elevated in women with PCOS. TNF-α is secreted by corpus luteum and plays a key role in reproductive physiology, including regulation of ovarian steroidogenesis. -308G>A polymorphism in promoter region of *TNFα* has been shown to be associated with hyperandrogenism, which is a major characteristic in PCOS cases. **Objective:** To analyse the association of *TNFα* (-308G>A) polymorphism in women of 17-40 years of age, suffering from Polycystic Ovary Syndrome and age matched controls. **Methods:** The present case-control study included a total of 91 subjects, including 41 cases and 50 controls. All the cases were recruited from Beri Maternity Home, Amritsar (August 2014 to May 2015). The cases were selected on the basis of detailed ultrasound report and PCOS symptoms as diagnosed by doctor. Age matched healthy controls without any PCOS symptoms were included. Genotyping for *TNFα* (-308G>A) polymorphism in all subjects was performed using polymerase chain reaction based restriction fragment length polymorphism. Statistical analysis was performed using SPSS package (version 17). **Results:** In 41 PCOS cases, the frequency of GG, AG and AA genotype was 73. 17%, 26. 82% and 0% compared to 82%, 18% and 0% in 50 controls. The mutant genotype (AA) was not observed in analysed cases and controls. The major allele (G) and minor allele (A) frequency was 87% and 13% in cases and 91% and 9% in controls. Although minor allele frequency was slightly higher in cases as compared to controls, however the distribution of genotype and allele frequency were not observed to be statistically significant (α^2=0. 571, p=0. 448 and α^2=0. 503,p=0. 478, respectively). The differences in means values between cases and controls were observed to be statistically significant (p=0. 031). **Conclusion:** In the present study *TNFα* (-308G>A) polymorphism has not been observed to be associated with PCOS. This may be attributed to small sample size.

849T

Regulatory variant alters expression of CRISPLD2, a Nonsyndromic Cleft Lip and Palate gene. *L. Mailli[1,2], B. T. Chiquet[2,3], E. C. Swindell[1,2], Q. Yuan[2], A. Letra[3], J. T. Hecht[1,2,3].* 1) Graduate School of Biomedical Sciences, University of Texas Health Science Center, Houston TX; 2) Department of Pediatrics, University of Texas Medical School at Houston, Houston TX; 3) University of Texas School of Dentistry at Houston, Houston TX.

Nonsyndromic cleft lip/palate (NSCLP) is the most common craniofacial birth defect occurring in 1/700 newborns and affecting 4000 babies in the US each year. NSCLP has a multifactorial etiology, wherein genetic variation and environmental factors contribute to causation. Genetic variation in the *CRISPLD2* gene was associated with NSCLP in our nonHispanic white and Hispanic families and confirmed in three other studies. Importantly, *CRISPLD2* is expressed in the mandible, palate and nasopharynx during mouse craniofacial development and knock-down of *CRISPLD2* in zebrafish causes abnormal palate and jaw development. Environmental factors such as retinoic acid, vitamin D and glucocorticoids are known to affect *CRISPLD2* expression. SNP rs1546124, which showed the strongest association in our nonHispanic white families (p = 0. 0006), is located 51 base pairs upstream of the ATG start site, suggesting a potential regulatory function. *In silico* analysis predicted several transcription factors binding differentially to the SNP region depending on the allele present, including a vitamin D receptor-binding site for the ancestral C allele and several retinoic acid binding sites for the alternate G allele. Electrophoretic mobility shift assay revealed allele-specific binding affinities, with the ancestral allele binding more strongly than the alternate allele. Luciferase reporter assays showed that the presence of the C allele was associated with increased promoter activity in COS7 and HEK293 cells (p=0. 000002). Additional *in vivo* and *in vitro* studies are underway to evaluate whether retinoic acid, glucocorticoid or vitamin D treatments influence allele-specific *CRISPLD2* expression. Our results suggest that rs1546124 is a functional SNP that influences *CRISPLD2* expression and may contribute to NSCLP.

850F

Identification of Genotype Driven Gene Expression Changes in Human Kidneys. *Y. Ko[1], H. Yi[1], M. Liang[1], N. Ledo[1], F. Chinga[1], C. Brown[2], K. Susztak[1].* 1) Renal Electrolyte and Hypertension Division, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Department of Genetics, University of Pennsylvania.

Introduction: There are more than 5 million sequence variants in humans. Some genetic variations influence transcript levels and therefore have the ability to interfere with cell and organ function. Such genetic variants are called expression quantitative trait loci (eQTL). The goal of our study was identify such functionally important genetic variants at the genome-wide level.

Methods: This analysis requires the collection of large number of human tissue samples with genotype and transcript level data. Here we used 99 human kidney samples of Central European descents. RNAseq and genotype data was normalized and genotype data was imputed using 1,000 Genome reference data. The association between genotype and transcript levels was performed using Matrix eQTL software and was limited to *cis*-eQTLs where the genetic variant and transcript levels were within 2 megabases distance.

Results: We identified 164 significant target genes (we call these eGenes) and 7590 significant SNPs (eSNPs) that passed the threshold for statistical significance after multiple testing correction using adjusted *p*-values<1. 0E-10. Next we compared variants that influence gene expression in the kidney to those that have been published for other organs using the publicly available Genotype-Tissue Expression project (GTEx). Of the 164 eGenes, 103 were common between the kidney and other organs, indicating that there are cell type specific and cell type independent eGenes. Using kidney specific epigenome maps, we found that eSNPs were enriched on kidney specific regulatory elements, including promoters and enhancers. We also found a significantly greater overlap between kidney eGenes and polymorphisms that are associated with CKD development, compared to other traits (digestive, nervous, immune system diseases, hematological measurement, cardiovascular, and metabolic disease).

Conclusion: We identified transcript level changes associated with genotypic variations. These results can highlight kidney specific regulatory elements and may also help to identify target genes for polymorphisms associated with kidney function related traits. .

851W

A polymorphic transposable element regulates *FADS1* expression. *C. Wadelius, M. Bysani, S. Enroth, A. Ameur, H. Nord, M. Cavalli, M. Essand, U. Gyllensten, G. Pan.* Science for Life Lab, Dpt of Immunol, Gen and Path, Uppsala University, Uppsala, Sweden.

Background: The FADS region may be associated to more diseases and phenotypes than any other, except for the HLA. The reported associations include the metabolic syndrome, fasting glucose, inflammatory bowel disease, triglycerides, LDL and HDL cholesterol, brain development and many others. The association was previously mapped to an LD block with 28 known SNPs and we set out to identify the functional one.

Methods and results: Using liver tissue we generated chromatin signals which define regulatory elements and then tested SNPs in them to see if the two variants differed in activity using luciferase assays. We found that all 28 SNPs could be explained by rs174557 which is located in a transposable element in intron 1 in *FADS1*, present in chimpanzee and human but no other species. This SNP showed significant difference between alleles in luciferase assays. PacBio single molecule sequencing detected multiple variants of the polymorphism and luciferase assays verified one main variant with high functional activity and several low activity variants. The regulatory element has three binding sites for the transcription factor (TF) SREBP1, one for SP1, and one at the regulatory rs174557 for the repressor PATZ1. We have verified the binding of each TF in vivo in HepG2 cells using ChIP-qPCR and also mutated all of the binding sites, alone or in combination and thus validated their function using luciferase assays. We have also overexpressed each TF and shown that activation is lost when the TF binding site is mutated. Using mutation and competition experiments it was found that the repressor PATZ1 competes in binding with the activating complex of SREBP1 and SP1 which is due to PATZ1 and SP1 having overlapping binding sites. When SREBP1 and PATZ1 are overexpressed in HepG2 cells there is increased ChIP-signal for each TF. MCF7 cells are heterozygous at rs174557 and after over-expression of PATZ1 allelic bias is shifted towards repression. **Conclusions:** The association of many diseases and traits to the FADS1 region can be explained by a multi-allelic polymorphism in intron 1 of FADS1 bound by the repressor PATZ1 in competition with an activating complex. The results are interesting from evolutionary and genetic perspectives and the gene variants also help explain why we react differently to modern diet as compared to the diet in ancient times.

852T

Novel human DNA Identikit using SNVs in the Fluidigm Nano fluidic Dynamic Arrays. *D. Bercovich[1], Y. Plosky[2], A. Stroler[1,2], Y. Dekel[1], I. Chernovitky[4], R. Gershoni[3].* 1) Human Molecular Genetics, Tel Hai College, Tel Hai, Israel; 2) GGA lab, Kazerin, Israel; 3) Rambam Mdical Center, Hifa, Israel; 4) Afula Medical Center, Israel.

In today's crime scenes, law enforcement authorities use face imaging (or identikit pictures), in order to identify and track down suspects. Usually, face imaging is based on eyewitness's accounts which are notoriously unreliable. occasionally , there is no eyewitness's at all. DNA found at crime scene can be a clincher in prosecuting a suspect when there is a match to the police databases. When there is no match, this important evidence might turn to be useless. In this study, 43 known human SNVs were tested on 251 DNA samples in order to determine a suspect's eye, hair/skin color, hair shape, racial mix (European /African), possible weight/ height, tendency to boldness, blood type (ABO) and left-right hand tendency. In addition, gender identification was performed using X & Y chromosome PCR primers and age determination (decades resolution) using RT-PCR . a correlation of 80-90% was found for SNPs determination of eye colour (Brown and blue respectively) or , dark hair colour , and 87- 90% for European African descent respectively. An average of 83% correlation was found to one of the four ABO blood types (A-77%, B-78%, AB-100%, O-91%) and 99. 99% for Gender determination and age +/- 8. 4 years (R^2 = 0. 7257). The SNV's proposed in this work, proved to be suitable for novel DNA Identikit that can be very helpful for further potential suspect identification. .

853F

Functional characterisation of the osteoarthritis susceptibility locus marked by the polymorphism rs10492367 at chromosome 12p11. 22. *K. Johnson, L. N. Reynard, J. Loughlin.* Newcastle University, Newcastle-upon-Tyne, Tyne and Wear, United Kingdom.

Osteoarthritis (OA) is a polygenic, multifactorial arthritis characterised by the progressive loss of articular cartilage in synovial joints. It is an age-associated disease but has aetiological routes during the life course. The arcOGEN GWAS reported that the rs10492367 G/T SNP marks an intergenic region on chromosome 12p that is associated with hip OA in Europeans, with an odds ratio of 1. 14 and a *p*-value of 1. 48 x 10-8. We hypothesised that the OA association signal acts by regulating the expression of *PTHLH* and/or *KLHL42*, the two most proximal genes. To investigate the functionality of the region we used luciferase reporter assays, showing differential enhancer activity between the alleles of rs10492367 and of several SNPs in high LD (r2 > 0. 8) with it. Through electrophoretic mobility shift assays and chromatin immunoprecipitation, we identified binding of SUB1, RELA and TCF3 to rs10492367. Knock-down of SUB1 in human articular chondrocytes (*n* = 3) revealed a down-regulation of *PTHLH*. We next sought to quantify gene expression in hip cartilage from OA patients who had undergone hip arthroplasties (*n* = 21). We observed decreased *KLHL42* expression (*p* = 0. 03) relative to age-matched non-OA controls (*n* = 18). The association signal did not correlate with expression of either gene. We confirmed this by interrogating the allelic expression imbalance of the genes using pyrosequencing. We used data from an Illumina BeadChip array to identify ten CpG sites from within the association interval that were significantly differentially methylated between OA hip (*n* = 17) and non-OA control cartilage (*n* = 42): six CpG sites had lower levels of methylation and four CpG sites had higher levels of methylation in the OA hip donors. rs10492367 genotype did not correlate with the methylation of any of the CpG sites analysed. We therefore speculated that the effect of the association signal is on *PTHLH* and/or *KLHL42* during joint development. We utilised microarray data and qPCR from mesenchymal stem cells to confirm expression of *PTHLH* and *KLHL42* during chondrogenesis and osteoblastogenesis, highlighting the potential for either of the genes to be regulated during skeletogenesis. Following an extensive interrogation of the OA association locus, our studies have defined a functional role for rs10492367. We have characterised gene expression and regulation, and suggest the region mediates OA susceptibility during joint development rather than in end-stage diseased cartilage.

854W

Long-range modulation of *PAG1* expression by 8q21 allergy risk variants. *C. T. Vicente, S. L. Edwards, K. M. Hillman, S. Kaufmann, H. Mitchell, L. Bain, D. M. Glubb, J. S. Lee, J. D. French, M. A. R. Ferreira.* QIMR Berghofer Medical Research Institute, Brisbane, QLD, Australia.

The gene(s) whose expression is regulated by allergy risk variants is unknown for many loci identified through genome-wide association studies. Addressing this knowledge gap might point to new therapeutic targets for allergic disease. The aim of this study was to identify the target gene(s) and the functional variant(s) underlying the association between rs7009110 on chromosome 8q21 and allergies. Eight genes are located within 1 Mb of rs7009110. Multivariate association analysis of publicly available exon expression levels from lymphoblastoid cell lines (LCLs) identified a significant association between rs7009110 and the expression of a single gene: *PAG1* (*P*=0. 0017), 732 kb away. Analysis of histone modifications and DNase I hypersensitive sites in LCLs identified four putative regulatory elements (PREs) in the region. Chromosome conformation capture confirmed that two PREs interacted with the *PAG1* promoter, one in allele-specific fashion. To determine if these PREs were functional, LCLs were transfected with *PAG1* promoter-driven luciferase reporter constructs. PRE3 acted as a transcriptional enhancer for *PAG1* exclusively when it carried the rs2370615:C allergy predisposing allele, a variant in complete linkage disequilibrium with rs7009110. As such, rs2370615, which overlaps RelA transcription factor (TF) binding in LCLs and was found to disrupt Foxo3a binding to PRE3, represents the putative functional variant in this locus. Our studies suggest that the risk-associated allele of rs2370615 predisposes to allergic disease by increasing *PAG1* expression which may promote B-cell activation and have a pro-inflammatory effect. Inhibition of *PAG1* expression or function may have therapeutic potential for allergic diseases.

855T

On the Prediction of Risk for Autism from Common Variants. *L. Zhu[1], P. Chaste[2, 3], Y. Song[2], K. Roeder[1], B. Devlin[2].* 1) Dept. of Statistics, Carnegie Mellon University, Pittsburgh, PA; 2) Dept. of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA; 3) FondaMental Foundation, Créteil, France.

Rare and common genetic variations contribute to risk for autism spectrum disorder (ASD). Indeed certain rare variants are so penetrant that their predictive value is beyond doubt. Here we explore prediction from common variation, more challenging because individual variants have small impact. To aid interpretation, we also explore prediction of a related quantitative trait, head circumference, which is often found to be somewhat enlarged in ASD subjects. Two ASD data sets were evaluated, the Simons Simplex Collection (SSC) and the Autism Genome Project (AGP). For head size it was relatively straightforward to develop a significant predictor, although the heritability explained was small, < 1%. One can obtain accurate prediction using parental head size, but common variants do not add to accuracy. It proved much more difficult to predict ASD, consistent with a loss of information for a binary versus continuous outcome. Prediction was only slightly better than a random guess when using pseudo-controls (parental non-transmitted alleles); it gained more predictive power when using unaffected subjects as controls, but was still far from being accurate; and, consistent with theory, its performance was significantly affected by whether ASD individuals carried a highly-penetrant rare variant. Some, although modest traction was garnered by analysis of common variants affecting expression of genes previously implicated in risk based on rare variant studies. Predicting risk from common variants currently is hard, over time it can be refined and in the process it will surely refine our understanding of the genetics of ASD. .

856F

Genomic variation in neuron-specific cell adhesion molecule "bar codes": candidate mechanims for disease and phenotype influences on the brain connectome. *G. R. Uhl[1,2,3], J. Drgonova[2], D. Walther[2].* 1) Research, NMVAHCS, Albuquerque NM , NM; 2) NIDA IRP, NIH; 3) Depts of Neurology, Neuroscience and Mental Health, Johns Hopkins Medical Institutions, Baltimore Md.

Recent reannotation reveals 474 human genes likely to encode cell adhesion molecules (CAMs). Most CAMs are expressed in brain and in neurons. When combined combinatorially, this number of gene products is sufficient to specify the large number of likely brain synapses. Many CAM genes are identified by repeated modest (10-2 < p < 10-8) GWAS signals for brain phenotypes. Eighty three are identified in at least three independent samples for addiction phenotypes. Dopamine neurons, implicated in reward from most addictive substances, express about 20 of these addiction-associated CAM genes (and more than 200 of the 474 human CAM genes in RNA seq datasets). Mice with altered expression of at least four studied addiction-associated CAM genes display altered place preference for stimulants, a reliable index for human abuse liability. Interestingly, several changedose-response relationships for reward. CDH13 and PTPRD, some of the most robstly addiction-related CAM genes, display haplotypes associated with sizable differencs in levels of expression in postmortem brain samples from unrelated individuals. CDH13 alterations produce changes in cerebral cortical dopamine innervation and neurochemistry. These examples undescore the ways in which individual differences in CAM bar codes can change neuronal connectivities, alter the connectome, and underpin substantial genetic influences on complex brain phenotypes.

857W

Genetic interaction analysis of reported longevity-associated epistatic effects in healthy oldest old Super-Seniors. *L. C. Tindale[1,2], S. Leach[1], A. R. Brooks-Wilson[1,2].* 1) Genome Sciences Centre, BC Cancer Agency, Vancouver, Canada; 2) Simon Fraser University, Burnaby, Canada.

Longevity and healthy aging are complex traits for which the genetic contribution is not well understood. Genome-wide association studies (GWAS) of longevity have found few replicable variants, and centenarians do not appear to carry a substantially decreased burden of common disease-associated variants. The existence of rare protective genetic factors has been put forward as a potential explanation for these observations. Protective variants may attenuate the effects of deleterious risk alleles, allowing long-lived individuals to carry disease-associated variants without showing phenotypic effects. A literature search for longevity or health-related epistatic variants produced 18 variants in 15 genes. The variants were genotyped in DNA samples from 466 Super-Seniors, individuals aged 85 or older who had never diagnosed with cancer, cardiovascular disease, major pulmonary disease, diabetes, dementia, as well as in 421 middle-aged controls. Both groups were of European ancestry. GeneMANIA (Warde-Farley et al. , Nucleic Acids Res. 2010;38) was used to explore gene interactions. Entering the 15 genes from the literature search,GeneMANIA produced a network that connected the query genes and incorporated functionally similar genes. Lipid and cholesterol Gene Ontology (GO) terms were significantly over-represented in the network. When each of the 18 variants was tested separately for association with the Super-Senior phenotype, only rs429358 in *APOE* was associated, and only in women, an observation we have made previously (Tindale et al. , Neurobiol. Aging. 2014;35(3)). SNP-SNP interactions were analyzed pairwise in a codominant model using logistic regression analysis. Using Super-Senior status as the outcome variable, one pair of SNPs that had previously been associated with Alzheimer disease (Gusareva et al. , Neurobiol. Aging. 2014;35(11)), rs6455128 in *KHDRBS2* and rs7989332 in *CRYL1*, showed evidence of an interaction after adjusting for sex and *APOE*. Other pairs did not show genetic interactions in our population. GWAS and candidate gene approaches are not sufficient to study healthy aging and longevity. Understanding the genetics of longevity will likely involve further research into complex genetic interactions.

858T

Crispld2 plays a role in neural crest cell migration and cell viability during zebrafish craniofacial development. *Q. Yuan, L. Maili, E. C. Swindell, J. T. Hecht.* Department of Pediatrics, University of Texas Medical School at Houston, Houston, TX.

Variation in the cysteine-rich secretory protein LCCL domain containing 2 gene, *CRISPLD2*, is associated with nonsyndromic cleft lip and palate (NSCLP). Expression of Crispld2 is detected in the craniofacies of developing mice and in the oropharynx of zebrafish during craniofacial development, suggesting that CRISPLD2 may be involved in orofacial morphogenesis. We have previously reported that knockdown of zebrafish *crispld2* by an antisense morpholino (MO3) directed against the start site causes severe craniofacial abnormalities involving both the jaw and palate. MO3 injected embryos showed abnormal expression of several marker genes for neural crest cells (NCCs). We have used a *sox10:egfp* transgenic zebrafish model to examine the NCC migration in live embryos. Knockdown of *crispld2* in zebrafish altered NCC migration pattern at early and late time points during orofacial development. Moreover, we detected an increase in cell death after *crispld2* knockdown as well as an increase in apoptotic marker genes. These preliminary results observed in the morphants suggest that Crispld2 is involved in palate and jaw morphogenesis, and knockdown of *crispld2* specifically affects NCCs. To further confirm these phenotypic changes, *crispld2* mutant lines bearing a deletion in exon1, predicted to abolish the entire gene function, have been generated using CRISPR/Cas9 site-directed nucleases. These mutant lines will serve as a model system for morphant vs mutant phenotype correlation studies. Gene expression analysis will be performed in homozygous mutant lines, which will allow us to identify potential downstream effectors of *crispld2*. Importantly, our results demonstrate that genes involved in complex birth defects, such as NSCLP, can be identified in candidate gene studies and functionally assessed in zebrafish to define the biological processes that are the basis for these complex birth defects.

859F

Integrative analysis for identification of shared markers from various functional cells/tissues for rheumatoid arthritis. *S. Lei[1,2], W. Xia[1,2], J. Wu[3], L. WU[1,2], Y. Guo[3].* 1) Center for Genetic Epidemiolgy and Genomics, Soochow University, Suzhou, Jiangsu, China; 2) Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Diseases, Soochow University, Suzhou, Jiangsu. China; 3) Department of Rheumatology, the First Affiliated Hospital of Soochow University, Suzhou, Jiangsu, China.

Objectives: Rheumatoid arthritis (RA) is a systemic autoimmune disease that may affect many tissues and organs. Gene expression profiling of various cells and tissues has been extensively conducted in RA research. In this study, we conducted a multi-dataset analysis to identify potential shared genes from multiple RA related expression datasets from various cells or tissues, and then a follow-up integrative analysis to test functional relevance of the identified genes on RA. Methods: The selected microarray datasets were downloaded from the publically available National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO). We conducted multi-dataset statistical analysis containing both marginal analysis and joint analysis to identify shared genes. The performed integrative analyses included functional annotation clustering analysis, protein to protein interactions (PPI), gene-based association analysis, and ELISA. Results: Our analysis included 7 datasets, 2 from synovial tissue, 1 from synovial macrophages, 1 from synovial fluid mononuclear cells (SFMCs), 2 from peripheral blood mononuclear cells (PBMCs) and 1 from peripheral blood cells. Multi-dataset statistical analyses identified 18 shared candidate genes, which were mainly involved in the immune response, chemokine signaling pathway and several other related biological processes. PPI analysis indicated that PPBP, PF4, HLA-F, S100A8, RNASEH2A, P2RY6, JAG2 and PCBP1 were connected with the known RA genes and may play crucial roles in the pathogenesis or development of RA. HLA-F and PCBP1 are significant (p=1. 03E-31; p= 0. 013, respectively) in gene-based association analysis. Subsequently, ELISA analysis found that PCBP1 (P=0. 026) has significantly differential expression between RA patients and healthy controls. Conclusions: This study represented the first effort to identify shared genetic RA markers from different functional cells or tissues. The results taken together suggested that PCBP1 and HLA-F are likely to be two shared potential genetic RA biomarkers.

860W

Identification of Novel Metabolic Syndrome Loci by Combining Univariate GWAS Summary Statistics of Multiple Phenotypes. *R. M. Salem[1,2], Z. Liu[1,3], X. Lin[1,3], J. N Hirschhorn[1,2].* 1) Medical and Population Genetics, Broad Institute, Cambridge, MA; 2) Genetics and Endocrinology, Boston Children's Hospital, Boston, MA; 3) Dept of Biostatistics, Harvard School of Public Heath, Boston, MA.

Metabolic Syndrome (MetS) is defined as the clustering of interrelated risk factors for CVD and type 2 diabetes mellitus (T2D). MetS incorporates 5 characteristic phenotypes: raised blood pressure, elevated triglycerides, elevated low -density lipoprotein cholesterol, raised fasting glucose, and obesity (central adiposity). Individuals with abnormal values for 3 out of 5 factors are diagnosed of MetS. MetS increases risk of developing CVD 2-3 fold and risk of developing T2D by over 5 fold. The role of genetic variants, both acting alone and in concert with obesity and lifestyle factors remains largely unknown. Analysis of syndromic phenotypes and multiple phenotypes can be challenging due to issues of phenotype modeling, analysis methods, and pragmatic issues of having complete phenotype traits measured in study samples. To overcome this limitations, we leveraged GWAS summary statistics for the 5 phenotypes (BMI, LDL, Triglycerides, Fasting Glucose, and SBP) used in the MetS definition to identify pleiotropic loci that show an association with combined traits and associate with MetS. Analyses were performed using a recently developed mixed fisher statistical procedure that calculates a joint p-value across the set of phenotypes, while correcting for correlation between the traits (Liu et al. , in preparation). The MetS analysis yielded 422 genome wide significant loci ($p \leq 5e\text{-}8$), of which 267 are novel loci and only observed via the joint analysis, none of the individual traits was genome wide significant alone. DEPICT analysis were performed on the novel and full set loci to prioritize loci and identify enriched pathways and tissues. We show that analysis of a complex multiple phenotype disease (Metabolic Syndrome) can be achieved by leveraging GWAS univariate summary statistics. In conclusion, these results have the potential yield novel insights into the genetic architecture and understanding of Metabolic Syndrome.

861T

Effect of RANK/RANKL/OPG gene polymorphisms on Left ventricular diastolic dysfunction in thalassemia major patients. *M. M. Singh[1], S. Tiwari[2], S. Agarwal[1].* 1) Medical Genetics, Sanjay Gandhi Post Graduate Institute Of Medical Sciences, Lucknow, Uttar Pradesh, India; 2) Department of Cardiology, Sanjay Gandhi Post Graduate Institute Of Medical Sciences, Lucknow, Uttar Pradesh, India.

In thalassemia major most common cardiovascular complication is left ventricular dysfunction, the inability of ventricles to pump blood, which progressively leads to heart failure. RANK (TNFRSF11A)/RANKL (TNFSF11)/OPG (TNFRSF11B) pathway has been shown to play role in the left ventricular dysfunction leading to heart failure. We therefore aimed to analyze their association with left ventricular diastolic dysfunction in thalassemia major patients. 70 asymptomatic patients (mean age 14. 3±5. 5 years) of thalassemia major were recruited, for echocardiography. PCR-RFLP analysis was performed in all to identify five polymorphisms in the RANK/RANKL/OPG pathway system [RANK (rs1805034, rs12458117, +35966insdelC), RANKL (rs2277438) and OPG (rs2073617)]. For echocardiography out of 70 asymptomatic patients of thalassemia major we observed that 25% of these had diastolic dysfunction. The major differences in the LV dimensions both in systole (p-value 0. 002) and diastole (p-value <0. 01), left atrium (p-value 0. 018), isovolumetric relaxation time (IVRT) (p-value 0. 009), left ventricular mass (p-value<. 01) and LV mass index (LVMI) (p-value 0. 005) was observed. RANK and RANKL polymorphisms shows insignificant association (wild type and heterozygous+mutant type) However, for LA diameter and LV mass, association was found with rs1805034, LVESD with rs12458117, IVRT with RANK ins/del, and LVESD with rs2277438. OPG polymorphism (rs2073617) shows significant association in all echocardiography parameter LVESD (p-value 0. 02), LA (p-value 0. 003), IVS (p-value 0. 001), PW(p-value 0. 04), and LVMI (p-value 0. 02). Present study concludes OPG (rs2073617) polymorphism significantly contributes for left ventricular dysfunction in thalassemia major patients thus can be used as cardiac biomarker for assessing left ventricular dysfunctions in Thalassemics.

862F

Whole exome sequencing in families at high risk for Hodgkin lymphoma: identification of a predisposing mutation in the KDR gene. *M. Rotunno[1], M. L. McMaster[1], M. Yeager[1], B. Hicks[1], L. Burdette[1], A. M. Goldstein[1], J. Boland[1], S. Ravichandran[1], B. T. Luke[1], L. Fontaine[2], S. J. Chanock[1], M. Tucker[1], N. C. Caporaso[1], L. R. Goldin[1], NCI DCEG Cancer Sequencing Working Group, NCI DCEG Cancer Genomics Research Laboratory.* 1) NIH, NCI, Rockville, MD; 2) Westat, Rockville, MD.

Hodgkin lymphoma (HL) shows strong familial aggregation but no major susceptibility genes for HL have been identified to date. Studies based on exome sequencing are promising for identifying disease susceptibility rare genetic variants. The goal of this study was to identify high-penetrance variants in HL-prone families. We conducted whole exome sequencing in 17 HL-prone families with three or more affected cases or obligate carriers followed by targeted sequencing in an additional 48 smaller HL families. Reads were aligned using Novoalign and variants were called using GATK software. Possible technical sequencing artifacts were eliminated by filtering out variants found in more than 1% of samples from other studies identically processed and sequenced in our laboratory. The exome called variants were analyzed, annotated and prioritized using ANNOVAR, Ingenuity Variant Analysis, and custom R scripts. Only dominantly segregating, rare ($\leq$ 1% frequency in European populations), coding or potentially functional variants were kept (2,699 variants total) and prioritized based on damage prediction, conservation, literature links to cancer or immune related processes. The top 23 genes were followed up using custom Ampliseq panel targeted sequencing. Only one variant in KDR (kinase insert domain receptor, aka VEGFR2) was replicated in two independent HL families. This variant shows strong evidence as a predisposing mutation for familial HL as it is highly conserved, rare (European population frequency <0. 001), located in the activation loop, and predicted to be damaging. KDR is a type III receptor tyrosine kinase, the main mediator of VEGF-induced endothelial proliferation, survival, migration, and its activity is associated with several diseases including lymphoma. Functional experiments have shown that the identified KDR mutation can promote constitutive autophosphorylation on tyrosine in the absence of VEGF and that the kinase activity was abrogated after exposure to kinase inhibitors. In the largest sequenced cohort of HL families to date, we identified a causal mutation in the KDR gene which could increase downstream tumor cell proliferation activity and be targeted for therapy in HL and NHL patients carrying this mutation.

863W

Sequencing reveals novel, germline variants in *BRIP1*: an analysis of 1300 women diagnosed with breast cancer before age 40. *C. Moyer[1], J. Ivanovich[2], D. C. Koboldt[3], K. L. Kanchi[3], E. R. Mardis[3], P. J. Goodfellow[1].* 1) The James Comprehensive Cancer Center, The Ohio State University, Columbus, OH; 2) Siteman Cancer Center, Washington University School of Medicine, St Louis, MO; 3) McDonnell Genome Institute, Washington University, St Louis, MO.

This year, an estimated 23,000 American women will develop breast cancer before the age of 40. This early-onset is before the recommended age for annual screening, often resulting in late detection, advanced stage, higher medical costs, and decreased survival rate. Many of these young women have a family history of breast cancer consistent with an inherited risk factor. However, nearly 80% of early-onset breast cancer patients lack mutations in known susceptibility genes, indicating there are genetic risk factors that have not yet been discovered. To identify additional breast cancer associated alleles, targeted custom amplicon sequencing of candidate DNA repair genes (*FANCI*, *FANCE*, *ATR*, *BRIP1* and *ERCC2)* was performed on germline DNA of 604 early-onset breast cancer patients negative for mutations in *BRCA1/2.* In combination with whole exome data from 26 additional early-onset patients, 37 women (6. 1% of cohort) were shown to harbor 24 different, rare and or novel missense variants in the BRCA1 interacting protein C-terminal helicase 1 gene (*BRIP1)*. Five previously unreported missense variants identified in our screen cluster within the c-terminal helicase domain of *BRIP1,* all of which are predicted alter protein function. A D791V variant, identified in single patient, was subsequently shown to be shared by a sister with breast cancer. The family pedigrees for several patients with other *BRIP1* variants include multiple members affected with breast and prostate cancers, suggesting the inherited *BRIP1* variants may be associated with cancer risk. To expand and validate our findings, *BRIP1* variant calls from an additional 742 patients in the cohort will be reported. Although a germline, nonsense variant (R798X) has been described as a low-penetrance breast cancer allele, little is known about BRIP1's function in breast cancer. We hypothesize that inherited variants in *BRIP1* account for a small percentage of early-onset breast cancer in individuals who lack mutations in known breast cancer susceptibility genes. Our sequencing studies suggest that novel, missense variants within the helicase domain of *BRIP1* confer risk for breast cancer. Results from functional testing for the effect of mutations on helicase and ATPase activity will be reported.

864T

Novel rare and common genetic variants associated with red blood cell traits identified in > 130,000 individuals genotyped on an exome array. *N. Chami[1,2], M. H Chen[3,4], A. Slater[5,8], N. Pankratz[6], L. Warren[3,4], S. Ganesh[7], G. Lettre[1,2] on behalf of the Blood Cell Consortium.* 1) Faculty of Medicine, Université de Montréal, Montréal, QC, Canada; 2) Montreal Heart Institute, Montréal, QC, Canada; 3) Framingham Heart Study, Population Sciences Branch, Division of Intramural Research National Heart Lung and Blood Institute, National Institutes of Health Framingham, Boston, USA; 4) Department of Neurology, Boston University School of Medicine, Boston, USA; 5) PAREXEL International, Durham, NC; 6) Department of Laboratory Medicine and Pathology, University of Minnesota Medical School, Minneapolis; 7) Cardiovascular Medicine, Department of Internal Medicine and Department of Human Genetics University of Michigan, Ann Arbor; 8) AS previously employed by GlaxoSmithKline (GSK), Research Triangle Park, North Carolina ; work performed on this publication done while employed by GSK.

Introduction: Red blood cell (RBC) traits are important heritable biomarkers used for the diagnosis of hematologic disorders such as chronic anemia and blood cancers. Genome-wide association studies (GWAS) for RBC traits have identified dozens of loci that together explain <10% of the inter-individual variation. **Materials and Methods:** To further explore the genetic factors that influence RBC traits, we meta-analyzed ExomeChip results for 7 RBC-related phenotypes available in up to 130,273 individuals from 29 studies. We used the RareMetalWorker/RVTests/RareMetals software suite to test for genetic associations and to meta-analyze results. For gene-based tests, we applied the SKAT and VT algorithms, considering only coding variants with a minor allele frequency (MAF) <1%. All analyses were performed on each of Europeans, African Americans, and combined ancestry which also comprised East Asians, and Hispanics. **Results:** We replicated many known GWAS SNPs (*e. g*. *HBS1L-MYB*, *TMPRSS6*, and *HFE*) and rare coding variants (*e. g. EPO*, *HBB*). We have additionally performed genome-wide conditional analyses to identify novel independent associations. In European-ancestry individuals, a novel SNP located near *TRIB1* was independently associated with RDW (P=1. 5e-7). The same variant is also associated with total cholesterol levels. Moreover, we found two SNPs in genes involved in lymphoblastic and myeloid leukemia: rs236985 in *AFF1* was independently associated with RBC count (P=4e-8) and HCT (P=4e-10), and similarly rs4911241 in *NOL4L* with RDW (P=6e-8). Further, we identified novel rare variants: a low-frequency missense variant in *HNF4A* (rs1800961, MAF=2. 4%, HCT P=1. 5e-8) previously associated with C-reactive protein and cholesterol levels and a rare missense variant in *ANK1* associated with MCHC (MAF=2. 9% P=1. 6e-16). The *ANK1* rare variant remained significant after conditioning on a GWAS common variant previously associated with MCHC (P=2e-15). Gene-based analysis identified *PKLR* that is associated with HCT in SKAT (P=8e-7). Interestingly, Mendelian mutations in *PKLR* are a cause of a chronic hereditary anemia. **Conclusions:** Our results highlight exciting overlaps between RBC-associated variants and genes implicated in cholesterol metabolism, emphasizing the tight equilibrium between RBC membrane composition and circulating cholesterol levels. We are completing discovery analyses and replicating novel genetic associations in independent studies.

865F

Heritability of Normal Human Facial Shape. *J. B. Cole[1], M. Manyama[2], W. Mio[3], S. A. Santorico[1], B. Hallgrimsson[2], R. A. Spritz[1].* 1) Human Medical Genetics and Genomics Program, University of Colorado Anschutz Medical Campus, Aurora, CO; 2) Department of Cell Biology & Anatomy, University of Calgary, Calgary, AB; 3) Department of Mathematics, Florida State University, Tallahassee, FL.

Facial shape exhibits perhaps the greatest variation of any normal human trait, yet at the same time represents the most recognizable human characteristic. Facial shape comprises multiple complex multifactorial traits with clear genetic components. However, little is known about normal human facial development. Previous GWAS have sought to identify genetic determinants of facial shape with little, if any, reliable replication. Traditional facial shape phenotypes used in GWAS derive from both principal components analysis (PCA) and direct linear distance measurements using sparse morphometric landmarks on a 3D image of the face. Using over 15 million genomewide markers with MAF>. 01 on 3700 normal African Bantu children from Tanzania, we estimated narrow-sense heritability and total heritability explained by the genotyped SNPs of the first 5 PCs from a PCA of 29 landmarks across the face and 24 inter-landmark distances. Taking advantage of close and distant relatedness, we fit a linear mixed model adjusted for age, sex, and centroid size to jointly estimate two variance components corresponding to heritability explained by genotyped and non-genotyped SNPs, the sum representing total narrow sense heritability. Our estimates for narrow sense heritability range from 5% - 64%, with 28/29 being >24%. Our estimates for total heritability explained by the SNPs range from 3% - 56%, with 28/29 being >19%. The most heritable phenotypes include cranial base width (64%); PC4 representing width of the nose and midface and mandible height (55%); outer canthal width (53%); nasal bridge length (51%); and PC5 representing nose shape, displacement of the mouth relative to the nose, and rotation of the eyes (51%). Using 15 million common variants we can explain >90% of heritability for 16/29 phenotypes and at least 45% for all phenotypes. Interestingly, PC1 representing upper face height and midface width has the highest genetic correlation with linear distances representing midface depth (. 82) and labial fissure width (. 60), both distances relating to midface width and not face height. PC2 representing overall and lower face height and mid and lower face prognathism and PC4 representing width of the nose and midface and mandible height have high genetic correlation with nasal bridge length (. 69 and . 78, respectively) and nasal height (. 60 and . 65, respectively). PC4 additionally is highly correlated with labial fissure width (. 77), nasal width (. 69), and subnasal width (. 67).

866W

Genetic Sharing and Heritability of Pediatric Age of Onset Autoimmune Diseases. *Y. R. Li[1,2], S. D. Zhao[3], J. Li[1], J. A. Ellis[4], R. N. Baldassano[5,6], E. T. L. Prak[7], H. Li[8], B. J. Keating[5], H. Hakonarson[1,5,9].* 1) The Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Medical Scientist Training Program, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL USA. sdzhao@illinois. edu; 4) Genes, Environment and Complex Disease, Murdoch Children's Research Institute, Parkville, Victoria, Australia;; 5) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA;; 6) Division of Gastroenterology, The Children's Hospital of Philadelphia, Philadelphia, PA, USA;; 7) Department of Pathology and Lab Medicine, Perelman School of Medicine University of Pennsylvania, Philadelphia, PA, USA;; 8) Department of Biostatistics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA;; 9) Division of Pulmonary Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA, USA;.

Autoimmune diseases (AIDs), affecting over 7-10% of the population in the Western Hemisphere, representing a group of common polygenic diseases with few effective therapies. To quantify the contribution of common genomic variations to the risk of pediatric AI diseases (pAIDs), including JIA, SLE, CEL, T1D, UC, CD, PS, AS and CVID, we calculated the heritability attributable to genetic differences (SNP-h2) using genome-wide SNP-genotypes for each pAID and pair of pAIDs. SNP-h2 estimates were significant across multiple pAIDs, including T1D (0. 863 +/- SE 0. 07) and JIA (0. 727 +/- SE 0. 037), but were more modest in UC (0. 386 +/- SE 0. 04) and CD (0. 454 +/- 0. 025). The MHC strongly contributed to SNP-h2 in T1D and JIA, but did not contribute strongly to SNP-h2 in other pAIDs (e. g. UC or CD); the results correlate with how strongly the MHC is associated with pAIDs based on GWAS results, are consistent with population reported h2, and in some cases are greater than SNP-h2 calculated based adult GWAS (i. e. , T1D, UC, JIA/RA). Pairwise genetic correlation (rG) between UC and CD and between JIA and CVID were the highest, consistent with estimated pAID genetic sharing obtained through independent statistical measures. The MHC did not play a significant role in rG. Together, our results partition contributions of shared versus disease-specific genomic variations to pAID heritability, identifying pAIDs with unexpected risk sharing and for which assessments of heritability is highly valuable. Future work utilizing these data may help establish a clinical tool that can be useful in populations that are at high risk for developing pAIDs given family history or to identify patients with a secondary pAID.

867T

Height-specific polygenic score using common alleles predicts extreme height rank. *R. H. Miller, L. M. McKinnon, K. Crofts, M. T. W. Ebbert, J. S. Kauwe.* Biology, Brigham Young University, Provo, UT.

Human height is a complex trait affected by hundreds of contributing genetic loci1 and a significant environmental component2. In documented history, just a small number of individuals have reached a height of 90 inches or more without exhibiting significant pathology related to pituitary function. Here we present evidence that calculating a height-specific polygenic score for such an individual can predict his/her extreme height rank when compared to a score distribution from a population-based control group with the same European-American ancestry. We have collected DNA from an individual who is 90 inches tall with no significant health issues or other physiological abnormalities to test the hypothesis that the subject is exceptionally tall due to a unique and extreme combination of common alleles for increased height. We tested this hypothesis by calculating a height-specific polygenic score3 for the subject and a population-based reference group of 189 individuals, and comparing the subject's score to the population distribution. To calculate the height-specific polygenic score, we extracted 22,539 height-associated SNPs with genome-wide significance ($p < 5e-8$) from a recent meta-analysis of 2. 5 million SNPs1. Data from the meta-analysis include each SNP's effect (b) on height. Of the significant height-associated SNPs, we were able to extract 3696 SNPs for the subject and all individuals within the control group. Using these SNPs, we calculated the height-specific polygenic scores for all individuals. The subject's score is in the 99th percentile (p=0. 005) supporting our initial hypothesis that this subject's height is the result of a unique combination of common alleles of small effect. Ongoing analyses will broaden both the number of alleles evaluated and the number of samples in the reference population. References:1. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46,** 1173–1186 (2014). 2. Silventoinen, K. *et al.* Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries. *Twin Res. Hum. Genet.* **6,** 399–408 (2003). 3. Dudbridge, F. Power and Predictive Accuracy of Polygenic Risk Scores. *PLoS Genet.* **9,** (2013).

868F

Novel method to estimate regional genetic associations improves genetic scores performance. *G. Pare[1], S. Mao[1], W. Deng[2].* 1) Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada; 2) Statistical Sciences, University of Toronto, Toronto, Canada.

Despite considerable efforts, known genetic associations only explain a small fraction of predicted heritability. We have previously shown that large regions joint associations, where multiple contiguous genetic variants are included in regression models, can improve the variance explained by established association loci as compared to genome-wide significant SNPs alone (Pare et al. PLoS Gen 2015). However, such regional joint associations are not easily amenable to estimation using summary data because of sensitivity to linkage disequilibrium (LD). Since only large genetic meta-analyses are likely to be powered to identify weak associations, we propose a novel method to estimate regional variance explained using summary GWAS data, accounting for LD while remaining robust to misspecification. **Methods:** We first divided the genome into SNP blocks of approximately 1 Mb minimizing inter-block LD. We then derived a method to estimate regional genetic effects using summary data, and showed it is asymptotically equivalent to multiple regression models when genetic effects are strictly additive (i. e. no interaction nor haplotype effect). We next extended our method to provide corresponding gene scores (GS) incorporating all genotypes within a given region. **Results:** We first compared our method to multiple regression on height and BMI using individual-level (N=7,776) data from the Health Retirement Study (HRS). While strong correlations were noted ($R2 \sim 0.25$), we estimated the loss of variance explained due to the additivity assumption at 0. 15 and 0. 26, respectively. Using summary GWAS statistics from GIANT for height (N=253,288) and BMI (N=236,231), we estimated variance explained by each SNP block and derived corresponding GS. We found that only the top 20% of blocks contributed to height variance in HRS, with GS prediction R2 of 0. 055 and 0. 021 before and after adjustment for genome-wide SNPs. Indeed, polygenic contribution of the 20% top blocks by REML (using GCTA) was 0. 63 (SD=0. 03) in HRS, while it was 0. 00 (SD=0. 04) in the remaining 80%. In contrast, all blocks contributed to BMI variance, with GS prediction R2 of 0. 050 and 0. 044 before and after adjustment for genome-wide SNPs. In summary, our results show that 1) strictly additive models do not fully capture complex traits associations, 2) traits differ in terms of the presence and distribution of regional associations, and 3) GS derived from regional associations can improve trait prediction.

869W

Dissecting genetic and environmental contributors to asthma and allergy in two founder populations. *M. Stein[1], C. L. Hrusch[2], C. Igartua[1], R. Anderson[1], N. Metwali[3], M. Holbreich[4], P. S. Thorne[3], E. von Mutius[5], D. Vercelli[6,7], A. I. Sperling[2], C. Ober[1].* 1) Human Genetics, University of Chicago, Chicago, IL. 60637, USA; 2) Department of Medicine, Section of Pulmonary and Critical Care Medicine, and the Committee on Immunology, University of Chicago, Chicago, IL 60637, USA; 3) Department of Occupational and Environmental Health, University of Iowa, Iowa City, IA 52242, USA; 4) Allergy and Asthma Consultants, Indianapolis, IN, USA; 5) Dr. von Hauner Children's Hospital, Ludwig Maximilians University Munich, Munich, Germany; 6) Arizona Respiratory Center and Bio5 Institute, Tucson, AZ 85721, USA; 7) Department of Cellular and Molecular Medicine, University of Arizona, Tucson, AZ 85724, USA.

Although asthma and allergic diseases are heritable, studies in central Europe have shown protection from these diseases in children raised on farms compared to their non-farming neighbors. This observation is recapitulated in two U. S. farming populations: the Indiana Amish, of Swiss-German descent, and the South Dakota Hutterites, who originated in the South Tyrol. These populations share similar lifestyles, although Amish live on single-family traditional dairy farms and use horses for field work and transportation whereas the Hutterites live on large, communal farms that are highly industrialized. Remarkably, the reported prevalence of asthma and allergic disease is 3- and 5-fold higher, respectively, in Hutterite compared to Amish children. To begin to unravel the factors accounting for these differences in disease prevalence, we studied 30 Amish and 30 Hutterite schoolchildren (ages 6-14). None of the Amish and 6 Hutterite children had asthma, and 2 Amish and 9 Hutterite children had high levels of specific IgE against common allergens (p=0. 04). There were similar and low levels of airborne allergens (dog, cat, house dust mite, cockroach) in dust collected in 10 Amish and 10 Hutterite homes, but dust from Amish homes had significantly higher levels of endotoxin (median 4399 vs 648 EU/m2; p=3. 3x10-6), reflecting higher levels of microbial exposure in Amish homes. Principal component analysis of ~72,000 genome-wide SNPs showed overlapping Amish and Hutterite clusters that were distinct from other European populations, reflecting similar ancestry and making it unlikely that the striking differences in disease prevalence are due to genetic differences between these populations. In contrast, gene expression levels in peripheral blood leukocytes were markedly different between the two groups, with 2,809 differentially expressed genes (FDR of 1%) between Hutterite and Amish school children. These differences in gene expression were entirely explained by differences in cell composition: after regressing out proportions of T cells, B cells, monocytes, eosinophils, and neutrophils, no genes were differentially expressed at a FDR of 1%. Moreover, neutorphils, monocytes, and T regs were phenotypically different between Hutterites and Amish. Overall, our study provides a first step toward identifying environmental correlates of asthma protection, and suggest the specific immune cells involved in risk or protection from asthma and allergic disease. (R01 HL085197).

870T

Maternal filaggrin mutations increase the risk of atopic dermatitis in children: an effect independent of mutation inheritance. *J. Esparza Gordillo[1,2], A. Matanovic[1,2], I. Marenholz[1,2], A. Bauerfeind[2], K. Rohde[2], K. Nemat[3], M. A. Lee-Kirsch[3], M. Nordenskjöld[4], M. C. G. Winge[4], T. Keil[5], R. Krüger[6], S. Lau[6], K. Beyer[6], B. Kalb[6], V. Niggemann[6], N. Hübner[2], H. J. Cordell[7], M. Bradley[4,8], Y. A. Lee[1,2].* 1) Experimental and Clinical Research Center, Charite Univ Med, Berlin, Germany; 2) Max-Delbrück-Centrum (MDC) for Molecular Medicine, Berlin, Germany; 3) Klinik fur Kinder- und Jugendmedizin, Technical University Dresden, Dresden, Germany; 4) Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; 5) Institute for Social Medicine, Epidemiology and Health Economics, Charité Universitàtsmedizin Berlin, Berlin, Germany; 6) Pediatric Pneumology and Immunology, Charité Universitàtsmedizin Berlin, Berlin, Germany; 7) Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, United Kingdom; 8) Dermatology Unit, Department of Medicine, Solna Karolinska University Hospital, Stockholm, Solna, Sweden.

Epidemiological studies indicate that maternal allergy is a stronger risk factor for the offspring than paternal allergy, suggesting a preferential maternal transmission of disease risk. However, the molecular mechanism underlying this parent of origin effect is currently unknown. Common loss-of-function mutations in the filaggrin gene (*FLG*) cause skin barrier deficiency and strongly predispose to atopic dermatitis (AD). We analyzed the 4 most prevalent European *FLG* mutations (c. 2282del4, p. R501X, p. R2447X,and p. S3247X) in two family-based samples (759 and 450 AD families) and applied the PREMIM/EMIM tool to test for parent-of-origin effects. As expected, children carrying a *FLG* mutation had a 2. 4 fold increased disease risk (R1meta-analysis = 2. 4, P = 1. 0 x 10–36). Strikingly, we also observed a strong maternal *FLG* genotype effect indicating that children of *FLG*-carrier mothers had a 1. 5-fold increased AD risk (S1meta-analysis = 1. 50, P= 8. 4 x 10–8). This maternal effect was consistent in both sets of families and for all 4 mutations analyzed. Our results point to two independent scenarios where *FLG* mutations increase AD risk: i) carrying a mutation and ii) having a mutation carrier mother. This maternal effect is independent of mutation inheritance and can be seen as a non-genetic transmission of a genetic effect. Interestingly, the *FLG* maternal effect was observed only when mothers had IgE antibodies against environmental allergens, suggesting that *FLG*-induced changes in the maternal immunity shape the child's immune system through feto-maternal cross-talk and increase the child's risk for AD. Our study supports the concept of "developmental origins of disease" and indicates that maternal *FLG* mutations act as strong environmental risk factors for the child. Additionally, it highlights the potential of family-based studies in uncovering novel disease mechanisms in the field of complex disease.

871F

A meta-analysis of atopic dermatitis reveals novel loci. *L. Vazquez[1], B. Almoguera[1], J. Connolly[1], P. Sleiman[1], F. Mentch[1], J. Linneman[2], M. Brilliant[2], M. Williams[3], K. Borthwick[3], A. Sundaresan[3], H. Hakonarson[1,4].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Marshfield Clinic, Marshfield, WI; 3) Geisinger Health System, Danville, PA; 4) Department of Pediatrics, The Perelman Schoolof Medicine, University of Pennsylvania, Philadelphia, PA.

Introduction: Atopic Dermatitis is a chronic inflammatory skin disease with a prevalence rate up to 20% in children and 3% in adults. The condition is characterized by intense pruritus and a course marked by exacerbations and remissions. There are 19 GWA loci reported to date. **Methods:** All cases and controls were defined by an electronic health record (EHR)-based algorithm, which mines health EHRs for pertinent diagnostic (i. e. ICD9 codes) and medication information. The algorithm was run and validated internally at the Children's Hospital of Philadelphia, with external validation at an independent site revealing positive predictive values of 92%-96%. A total of 6,100 children and 5,421 adults of European ancestry were included in the analysis. Imputation of non-observed genotypes was carried out using SHAPEIT and IMPUTE2. Population ancestry was determined by principal component (PC) analysis using Eigenstrat 3. 0 and 10 PCs were generated and used as covariates in each individual association analysis, which was conducted using SNPtest. Metal was used for the meta-analysis of the two cohorts. **Results:** Three loci reached genome-wide significance in the meta-analysis of all 11,521 subjects: a novel locus on 17q21. 33 (p=8. 61e-09), where the carbonic anhydrase 10 gene (*CA10)* is located; *SLC22A8* in 3q21, which is known psoriasis susceptibility gene; and *CDH26* at 20q13. 33, a known susceptibility locus for atopic dermatitis. **Conclusion:** We have identified two novel loci that help contribute to our understanding of the missing heritability of atopic dermatitis.

872W

Chromatin regulatory circuitry defines inherited disease risk. *O. Corradin[1], A. J. Cohen[1], J. M. Lupinno[1], I. M. Bayles[1], P. C. Scacheri[1,2].* 1) Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH; 2) Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH.

SNPs associated with susceptibility to common diseases through Genome Wide Association Studies (GWAS) often lie in clusters of active gene enhancer elements including super or stretch enhancers. At these loci, the individual constituents of the enhancer clusters interact and cooperate to regulate target gene levels, and these constituents are distributed both within and beyond the boundaries of the linkage disequilibrium blocks containing the lead risk SNPs. These observations led us to hypothesize that SNPs surrounding, but not in LD with lead SNPs can target elements of a gene's regulatory circuitry and impact not only the effect of the risk SNP on gene expression, but also the clinical risk to disease. We tested this hypothesis using a novel strategy to integrate GWAS SNPs associated with six autoimmune diseases with haplotype maps, epigenomic landscapes, Hi-C chromatin interaction maps, and transcriptomic and proteomic datasets. We identified thousands of novel variants that are inherited independently of the risk locus (r2 < 0. 5, D' < 0. 9, LOD < 2), but nonetheless interact with the risk SNPs as part of the target gene's regulatory circuitry. These interacting SNPs, which we call "outside variants," were found to functionally modify the effects of the risk SNP on target transcript levels for over one-third of all evaluated GWAS loci (n = 64/186). A subset of these outside variants were functionally validated at the protein level, and using luciferase reporter assays. To determine if these effects contribute to disease pathogenesis, we evaluated their impact on clinical risk. Outside variants were found to alter the clinical risk to disease for over two-thirds of all evaluated GWAS loci (n = 23/33, P < 0. 05). Our findings indicate that the DNA sequence underlying all of a gene's regulatory elements, regardless of linkage disequilibrium with the GWAS SNPs, need to be taken into account when quantifying the effects of GWAS risk SNPs on target gene levels and calculating the true clinical risk of the associated locus. The findings also suggest that interactions defined by 3D chromatin architecture can be used to uncover novel variants that can explain additional heritability of common disease.

873T

The Inheritance of human lifespan in 20th Century Scotland. *P. K. Joshi[1], H. Tochet[1,2,3], P. Navarro[2], C. S. Haley[2], J. F. Wilson[1,2].* 1) Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, United Kingdom; 2) MRC Human Genetics Unit, University of Edinburgh, Western General Hospital, UK; 3) Wellcome Trust Centre for Human Genetics, Roosevelt Dr. Oxford, OX3 7BN.

Longevity is the ultimate phenotype and is a consequence of many intermediate complex disease phenotypes and other genetic and environmental factors. The genetic basis of human longevity is thus even less straightforward than complex trait diseases. Furthermore, the phenotype takes a lifetime to emerge and large sample sizes are needed for useful analysis. Whilst, a number of frequently cited studies have estimated the heritability of human longevity, for the populations concerned, in the range of 0. 20-0. 33, many of these studies have been twin based, or used isolated or otherwise unusual populations. It is therefore of interest to establish further estimates of the heritability of longevity with narrow confidence intervals in social settings more typical of Western European life today. Our study estimated the narrow-sense heritability (h) of human longevity conditioned on living beyond the age (42), prior to which accident or war was likely to be a greater factor. Our data was drawn from two sources. Firstly, we gathered records for offspring born in Eastern Scotland around 1900 and their parents, from Scottish public records using rare family names to facilitate efficient searches. Variables available for analysis were thus defined by the contents of death certificates. We examined the regression of child age at death on parental age at death using covariates reflecting social class, industry and location as well as year of birth and parental age at birth effects. Secondly we extracted parent-offspring trio data for offspring born around 1900, from a pedigree already gathered for the ORCADES complex disease study. Using the correlation between parent and offspring, we estimated h2 for human longevity (+/- 1 SE) as 0. 15 (+/- . 025) and found suggestions of differences between the sexes. At the same time, we found similar correlations between spouses and between a single parent and offspring, raising questions as to whether common environment has resulted in an overestimate of the heritability of lifespan by ourselves and others. Thus our results suggest the heritability of longevity may be somewhat lower in regular Western European settings than suggested by other studies but also raise tantalizing questions about the differences in the genetic basis of lifespan and the confounding effect of common environment, in particular mothering environmental influences on longevity outcomes.

874F

Genome-wide heritability estimates of sporadic brain arteriovenous malformations. *H. Kim[1, 2, 3], J. Nelson[1], S. M. Weinsheimer[4], N. Bendjilali[5], C. E. McCulloch[2], L. Pawlikowska[1, 3].* 1) Dept. of Anesthesia, University of California San Francisco, San Francisco, CA; 2) Dept. of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA; 3) Institute for Human Genetics, University of California San Francisco, San Francisco, CA; 4) Institute of Biological Psychiatry, Copenhagen, Denmark; 5) Dept. of Mathematics, Rowan University, Glassboro, NJ.

Introduction. Brain arteriovenous malformations (BAVM) are abnormal tangles of blood vessels that can lead to hemorrhagic strokes, seizures, headaches, and other neurological deficits. Most BAVMs are "sporadic" lesions of unknown genetic etiology. The prevalence of sporadic BAVM is typically estimated as 1d1. 8 in 10,000, but may be as high as 5 in 10,000 based on a recent meta-analysis of incidental findings on MRI. The heritability of sporadic BAVM has never been formally estimated. Thus, we calculated and tested whether sporadic BAVM was a heritable trait using genome-wide complex trait analysis. **Methods**. The study included 371 BAVM cases and 563 healthy controls, all Caucasian, and genotyped on the Affymetrix Genome-Wide Human SNP Array 6. 0 in the same laboratory. We applied stringent QC filters as described by de Candia et al (Am J Hum Genet. 2013;93(3):463-70), and estimated the heritability of sporadic BAVM after adjusting for age and sex with an expectation maximization algorithm implemented in GCTA software. We varied the disease prevalence and evaluated statistical significance by permutation test. **Results**. After QC filters, our sample comprised 297 cases and 468 controls, and 484,737 common SNPs. Age and sex adjusted heritability of BAVM was 17. 6% (SE=8. 9%) for a prevalence of 1 in 10,000, and 26. 9% (SE=13. 6%) for a prevalence of 5 in 10,000. The permutation p-value was 0. 015, suggesting that heritability is significantly greater than zero. **Conclusions**. Genomewide estimates of heritability suggest that sporadic BAVM has a significant, albeit modest, additive genetic contribution. However, our heritability estimates should be viewed with caution since the true prevalence of BAVM is unknown, disease prevalence strongly affects heritability estimates, and simulations have shown that traits with both low heritability and low prevalence can have upwardly biased estimates.

875W

Defining endophenotypes of Age-related Macular Degeneration in the Amish. *R. J. Sardell[1], L. D. Adams[1], R. Laux[2], D. Fuzzell[2], L. Reinhart-Mercer[1], L. Caywood[1], D. Dana[3], A. Bowman[3], M. G. Nittala[4], S. R. Sadda[4], W. K. Scott[1], D. Stambolian[3], J. L. Haines[2], M. A. Pericak-Vance[1].* 1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 3) Departments of Ophthalmology and Genetics, University of Pennsylvania, Philadelphia, PA; 4) Doheny Eye Institute, Los Angeles, CA.

Known common and rare genetic variants explain up to 60% of the heritability of AMD. Unexplained heritability may partly reflect challenges defining the complex phenotype; the current classification system uses broad-scale grades that may not adequately represent the underlying biological basis of the disease. Furthermore, most studies compare advanced cases with controls. We hypothesized that parsing the AMD phenotype into heritable sub-phenotypes, each with a distinct genetic basis ("endophenotypes"), will improve prediction of risk and increase our understanding of the genetic architecture of AMD. We used Ocular Coherence Tomography (OCT) imaging to quantify fine-scale retinal features, compare these traits to the traditional Age-Related Eye Disease Study (AREDS) scale and assessed their potential as endophenotypes. We sampled 566 related individuals from Amish families with early/intermediate AMD cases in Pennsylvania, Ohio and Indiana. The Amish provide an excellent opportunity to analyze the heritability of complex traits given their large nuclear families and relatively homogeneous environmental and genetic backgrounds. Approximately 69% of individuals were controls (grade 0/1), 11% had early AMD (grade 2), 12% had intermediate AMD (grade 3) and 8% advanced AMD (grades 4/5). AREDS grade (rs=0. 86) and choroidal thickness (r=0. 84) were highly correlated between left and right eyes. Choroidal thickness was moderately correlated with AREDS grade in both eyes (rs=-0. 31). A generalized linear mixed model, accounting for age, gender, clinic and repeated measures (left and right eyes) confirmed that choroidal thickness was highly repeatable within individuals (0. 78) and moderately heritable (additive genetic variance h2=0. 30, 95% CI=0. 13-0. 50). Choroidal thickness was also negatively correlated with age (r=-0. 48), but not gender. This study is the first reported estimate of the heritability of choroidal thickness, highlighting a heritable, quantitative OCT trait that is both measurable in all individuals and correlated with AMD status. Choroidal thickness may therefore be an informative biomarker of AMD risk. Moderate correlation between choroidal thickness and AREDS grade suggested that this OCT trait may capture phenotypic variation not captured by the traditional AREDS system. Choroidal thickness will therefore be tested for association with genetic variants to determine whether this trait has a distinct genetic etiology and defines an AMD endophenotype.

876T

Missing heritability may be explained by the common household environment and its interaction with genetic variation. *N. Wang[1,2,3], R. M. Watanabe[1,2,3].* 1) Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA; 2) Physiology and Biophysics, Keck School of Medicine of USC, Los Angeles, CA; 3) Diabetes and Obesity Research Institute, Keck School of Medicine of USC, Los Angeles, CA.

Genome-wide association studies have identified thousands of common genetic variants associated with complex diseases and disease-related quantitative traits (QTs). However, the majority of the heritable variance (narrow-sense,*h2*) remains unexplained. Although numerous plausible explanations for the inability to fully account for*h2*have been put forth,*e. g.* , undetected rare variants, structural variation, gene-gene and gene-environment interactions, none fully explains this phenomenon. We formulated an alternative explanation that*h2*estimates may be upwardly biased due to model misspecification. In particular, we hypothesize that*h2*estimates have not accounted for the common household environment (CHE) and/or its interaction with genetic variation. We tested our hypothesis by simulating trios with a QT as the sum of 4 components: (1) additive effect of*m*common SNPs with joint variancea2, (2) CHE distributed as*N*(0,dH2), (3) a multiplicative interaction between CHE and one SNP (xint) with effect sizebint, and (4) the residual error distributed as*N*(0,e2). Genotypes were simulated assuming linkage and Hardy-Weinberg equilibria. Trio members were individually assigned the same sampled CHE. Simulated data were analyzed by variance components with maximum likelihood to estimate*h2*. Different parameter combinations were tested with 1000 sets of 200 trios analyzed for each setting. We first examined the case where no interaction was in the true model, and onlya2ande2were in the estimation model. We theoretically derived equations based on Huber (1967) and showed the estimateda2was inflated and the degree of inflation increased with an increasing CHEdH2, consistent with simulation studies. For example, whena2=dH2=e2=100, estimateda2is 228; a 2. 3 folds inflation. We then examined the case where all 4 components were in the true model, and onlya2ande2were in the estimation model with or without specifying CHE. Simulation studies showed the estimateda2was inflated, even with specifying CHE, and the degree of inflation increased with an increasingbint. For example, whena2=dH2=e2=100and the effect allele frequency of the SNP xintbeing 0. 2, the inflation of a2 ranges from 1. 3 to 8. 7 folds with |bint| in the range of 1 to 5. In conclusion,*h2*estimates may be greatly biased upward due to model misspecification. This inflation can be caused by not taking into account CHE and its interaction with genetic variation.

877F

Bayesian analysis of polygenic effects on BMI. *J. H. Zhao, J. A. Luan, S. J. Sharp, C. Langenberg, R. Scott, N. J. Wareham.* MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge Biomedical Campus, Cambridge, CB2 0QQ, United Kingdom.

Background: In study of obesity-related traits, there has been considerable interest in polygenic effects on body mass index (BMI) which usually involves maximum likelihood methods with uncertainty in parameter estimation. We investigated this through Bayesian analysis of whole genome data in population-based studies. **Methods:** We analysed data from two population-based studies of European origin: EPIC-Norfolk (N=21,044) and Fenland (N=8,994), genotyped with Axiom BioBank genotyping arrays, and compared with results from earlier data from the same cohorts using Affymetrix 500K (N=2,514) and Affymetrix 6. 0 (N=1,357) genechips. We conducted our analyses using Bayesian linear regression implemented in R instead of the computer program GCTA. **Results:** Using scaled inverse chi-square priors on the polygenic and residual variance components and MCMC procedures involving with 15,000 burn-ins and 50,000 iterations, the point estimates (95% HPD interval) for heritability, the total additive genetic variance as proportion of the total phenoyptic variance, were 0. 33 (0. 28-0. 38) and 0. 45 (0. 36-0. 54) for EPIC-Norfolk and Fenland data compared to heritability estimates (95%CI) of 0. 13 (0-0. 35) and 0. 35 (0-0. 76), respectively using earlier data on the two cohorts. In particular, the Fenland estimate agrees with those reported from family studies such as the Framingham study and is slightly higher than EPIC-Norfolk possibly due to younger age of the participants (mean (SD) age 48. 9 (7. 4) and 59. 2 (9. 3), respectively). In addition, we obtained the posterior distribution of polygenic effects. **Conclusion:** Our work corroborates earlier evidence on polygenic effects on BMI from family studies and shows that Bayesian approach applied to large sample and whole genome data can give better account of the sampling variations in polygenic effects. We are examining performance of prediction with single nucleotide polymorphisms (SNPs) across the genome with the same framework.

878W

Comparison of Predictive Value of Genetic and Transcriptional Risk Scores. *U. Marigorta, G. Gibson.* School of Biology, Georgia Institute of Technology, Atlanta, GA.

Most variants discovered by genome-wide association studies (GWAS) lie in regulatory regions, and a sizable fraction act as alleles that alter gene expression (eQTL). These observations suggest that a large proportion of susceptibility to complex disease may be due to detrimental changes in gene regulation. Consequently, transcript abundance levels could be used as biomarkers for prediction of individual risk of disease. Since these presumably integrate both the genetic and environmental influences, it might be expected that transcript abundance is a better predictor of disease than genotypes alone. Given this expectation, we propose that the average relative abundance of transcripts influenced by disease-associated eQTL could be used to calculate transcriptional risk scores (TRS). We here report results of simulations designed to evaluate the relative performance of GRS and TRS as predictors of disease risk. For a wide range of parameter space consistent with current knowledge of the genetic architecture of disease, we observe that transcriptional risk scores are indeed likely to be more sensitive markers of susceptibility to disease. Moreover, we develop a number of alternative scores that make different assumptions about the weights of extreme values of gene expression, or the role of the gene embedded in a pathway. We argue that this parameterization might be used to constrain models of the relationship between eQTL, GWAS signals, and actual disease etiology. For example, if it is assumed that only extreme gene expression is causal in disease, common variant genetic effects should not be seen as average effects, but rather as markers of heterogeneous common variants of rare effect. Testing of these concepts requires development of more datasets where gene expression is measured in relevant tissues from patients with disease, in addition to healthy people samples that currently dominate the gene expression studies.

879T

Identifying genes associated with maternal nondisjunction of chromosome 21. *J. M. Chernus[1], E. G. Allen[2], Z. Zeng[3], C. Trevino[2], E. Feingold[1,3], S. L. Sherman[2].* 1) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Human Genetics, Emory University, Atlanta . GA; 3) Department of Biostatistics, University of Pittsburgh, Pittsburgh PA.

We have conducted the first GWAS to identify genes associated with nondisjunction of chromosome 21 in oocytes. A total of 2,186 study subjects were genotyped on the HumanOmniExpressExome-8v1-2 array. These subjects included over 700 live birth offspring with standard trisomy 21 derived from an error in the oocyte and over 1400 parents. Genotypes for chromosome 21 in children were called using methods we previously developed. Genotypes for parents and child were then used to establish parent of origin, stage (meiosis I or meiosis II) and meiotic recombination patterns for each child. We used two different designs for conducting the GWAS. In the first, we take advantage of the fact that the etiologies of meiosis I trisomy 21 and meiosis II trisomy 21 appear to be quite different, and perform a GWAS of mothers of meiosis I cases vs. mothers of meiosis II cases. This approach cannot find variants that are common risk factors for both meiosis I and meiosis II errors, but is a well-controlled design for finding variants that are unique to one or the other. The second approach uses the fathers as controls. In that approach, we may also discover genes that affect the likelihood of a fetus with trisomy 21 surviving to term.

880F

Genome-wide association study of 41 circulating cytokines. *A. V. Ahola-Olli[1], J. Kettunen[2], P. Würtz[2], N. Pitkänen[3], K. Aalto[1], M. Salmi[1], A. Havulinna[1], V. Salomaa[4], T. Lehtimäki[5], S. Jalkanen[1], O. T. Raitakari[1].* 1) University of Turku, Turku, Finland; 2) University of Oulu, Oulu, Finland; 3) University of Eastern Finland, Kuopio, Finland; 4) Finland National Institute of Health and Welfare, Helsinki, Finland; 5) University of Tampere, Tampere, Finland.

Introduction: Cytokines have essential roles in regulation of immune response. Dysregulation of immune system have been linked to various autoimmune diseases such as inflammatory bowel diseases and rheumatoid arthritis. The purpose of this study is to shed light to yet unknown biology behind immune system regulation and pathogenetic mechanisms behind autoimmune diseases. To achieve this, we performed genome-wide association study of healthy Finnish subjects randomly chosen from national registers. To dissect biological mechanisms behind the associated loci, we performed expression quantitative trait loci analyses for lead SNPs in whole blood and searched existing databases for additional information from the loci. Materials and Methods: Both the genotype and cytokine data were available from three Finnish population based cohorts: The Cardiovascular Risk in Young Finns Study, Finrisk97 and Finrisk02. Total of 8153 subjects were included in this study. Cytokine measurements have been performed with Bio-Rad's premixed Bio-Plex Pro Human Cytokine 27-plex Assay and 21-plex Assay. The amount of bound cytokine was detected using streptavidin-phycoerythrin conjugate. 1000 genomes imputation panel was used to impute genotypes. Snptest software was used to conduct genome-wide scans. Meta-analyses were performed with METAL software. Following filters were applied to data: info > 0. 7, minor allele count > 10 and association test info > 0. 7. Cis eQTL analyses were performed for lead SNPs. Results: We identified 14 new loci previously not associated with cytokines, hsCRP or white blood cell count. Six of these loci are located near a SNP previously associated with an autoimmune disease. Total of 27 loci had genome-wide significant association with the concentration of at least one cytokine. Conclusions: This study have provided important insights to development of autoimmune diseases and identified cytokines that probably have important role in development of these diseases.

881W

Analysis of pathways associated with Body Mass Index in individuals with multiple sclerosis. *M. G. Durrant[1], J. C. Denny[2], M. F. Davis[1].* 1) Department of Microbiology and Molecular Biology,, Brigham Young University, Orem, UT; 2) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN.

Current evidence suggests that a high body mass index (BMI) earlier in life is associated with multiple sclerosis (MS) later in life. The nature of this association is not fully understood. Previous research indicates that MS patients may also experience weight gain through the course of their disease, with the hypothesized cause being a combination of pain, fatigue, and medications such as glatiramer acetate.

The electronic medical records of 1,809 MS cases and 34,997 controls were analyzed to determine significant differences in BMI between the first recorded clinic visit and the most recent clinic visit. Results indicate that MS patients have a significantly higher BMI at the first and last clinic visits. Then, three linear regression and pathway-based association studies were conducted for 1,003 MS patients' BMI phenotype data from the first and last clinic visits, as well as the slope of BMI change between the first and last visits. Pathway analyses were conducted using the PARIS algorithm. The resulting significant KEGG pathways were compared to other BMI pathway analysis studies to identify unique pathways associated with MS patients.

No individual SNPs met genome-wide significance in any of the linear regression analyses. Among the top 20 most significant SNPs in each analysis, several SNPs occurred in groups around specific genes, including BANK1, RNLS, SLC14A2, TDH and C8orf12. After removing pathways already known to be associated with BMI, PARIS results indicated an association between a high BMI at first clinic visit and cerebellar long-term depression ($p = 0.014$), as well as an association between the overall BMI slope-change between visits and the systemic lupus erythematosus (SLE; $p = 0.001$) and RIG-I-like receptor signaling ($p < 0.001$) pathways.

We found that BMI was significantly elevated in MS patients both at the first and last clinic visit when compared to controls, providing some evidence for a link between MS and elevated BMI. The BANK1 gene has previously been linked to other autoimmune diseases such as type 1 diabetes (T1D) and SLE, but not directly with MS. RNLS has also been previously linked to T1D. Additionally, the relationship between the slope of BMI change between visits and other autoimmune disease pathways suggests that the nature of the MS itself and susceptibility to other autoimmune diseases such as SLE and T1D may contribute to high BMI in MS patients.

882T

A comprehensive framework for fine-mapping disease loci across autoimmune diseases. *A. Hadjixenofontos[1], M. Mitrovic[1], C. Cotsapas[1,2].* 1) Department of Neurology and Immunobiology, Yale University, New Haven, CT; 2) Broad Institute of MIT and Harvard, Boston, MA.

Epidemiological, genetic and clinical evidence suggests the co-morbidity between autoimmune diseases (AID) is due to shared pathogenic mechanisms. Understanding if risk is driven by the same alterations would allow treatment re-purposing across diseases and the emergence of a molecular nosology to guide both therapeutic intervention and development. Genome-wide association studies have identified tens of loci mediating risk to multiple AID; however, inferring a shared pathogenic mechanism requires demonstrating that the same genetic effect mediates risk to multiple diseases. To address this challenge, we have collated genotype-level data from 79,638 cases with inflammatory bowel disease, multiple sclerosis, systemic lupus erythematosus, type I diabetes, or celiac disease and ~85,000 matched healthy controls. We have created a robust analysis framework for cross-disease analysis where we can distinguish between heterogeneous but co-localizing effects and those that are truly shared across traits. After applying a unified quality control pipeline, we use a flexible population-matching scaffold that we have developed based on principal component projection. This scaffold allows us to select optimal sets of samples for analysis of each locus harboring risk variants for any subset of diseases. We have developed methods for assessing if associations in a genomic locus are driven by the same risk variant across independent phenotypes. We then use these methods in all case samples, identify cases in which the effect appears to be shared thereby increasing effective sample size in the region and thus increasing the resolution of fine-mapping. In simulations of the NFKB1/MANBA locus on chromosome 4, where an extended haplotype of 102 markers shows strong association to risk of multiple sclerosis, we find that we are able to reduce the size of the credible interval by 37% by increasing sample size in our framework. In addition, the simulated causal variant is included in the credible interval 96% of the time, compared to 71% of the time using only multiple sclerosis samples. We are currently fine-mapping 185 loci robustly associated to at least one AID in this fashion, including deriving expectations through simulation.

883F

Genome-wide Study of Late-Onset Myasthenia Gravis: Confirmation of *TNFRSF11A*, and Identification of *ZBTB10* and Three Distinct HLA Associations. *M. F. Seldin[1], O. K. Alkhairy[2], A. T. Lee[3], J. A. Lamb[4], J. Sussman[5], R. Pirskanen-Matell[6], F. Piehl[6], J. J. G. M. Verschuuren[7], A. Kostera-Pruszczyk[8], P. Szczudlik[8], A. Maniaol[9], H. F. Harbo[10], B. A. Lie[11], A. Melms[12], H-J. Garchon[13], N. Willcox[14], P. K. Gregersen[3], L. Hammarstrom[2].* 1) Biochemistry and Molecular Medicine, Sch Med, Univ California, Davis, Davis, CA; 2) Division of Clinical Immunology, Karolinska Institutet at Karolinska University Hospital Huddinge, Stockholm, Sweden; 3) The Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, North Shore LIJ Health System, Manhasset, NY; 4) Centre for Integrated Genomic Medical Research, Manchester Academic Health Science Centre, University of Manchester, Manchester, UK; 5) Department of Neurology, Greater Manchester Neuroscience Centre, Manchester UK; 6) Department of Neurology, Karolinska University Hospital Solna, Stockholm, Sweden; 7) Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands; 8) Department of Neurology, Medical University of Warsaw, Warsaw, Poland; 9) Department of Neurology, Oslo University Hospital, Ullevål, Oslo, Norway; 10) Department of Neurology, Oslo University Hospital and University of Oslo, Oslo, Norway; 11) Department of Medical Genetics, University of Oslo and Oslo University Hospital, Oslo, Norway; 12) Department of Neurology, Tübingen University Medical Center, Tübingen, Germany and Neurologische Klinik, Universitétsklinikum Erlangen, Erlangen, Germany; 13) Inserm U1173, University of Versailles,Campus, Paris-Saclay, France; 14) Department of Clinical Neurology, Weatherall Institute for Molecular Medicine, University of Oxford, Oxford, United Kingdom.

To define the genetics of late-onset myasthenia gravis (LOMG), we conducted a genome-wide association study including imputation of >6 million SNPs in 532 LOMG cases (anti-cholinesterase receptor antibody positive, onset age >50 years) and 2128 controls matched for sex and population substructure. The data confirm reported *TNFRSF11A* associations [rs4574025, p = 3. 9x10-07, odds ratio (OR) = 1. 42] and identify a novel candidate gene, *ZBTB10*, achieving genome-wide significance (rs6998967, p = 8. 9x10-10, OR = 0. 53). Several other SNPs show suggestive significance including rs2476601 (p = 6. 5x10-06, OR =1. 62) encoding the *PTPN22* R620W variant associated with early-onset MG (EOMG) and other autoimmune diseases. In contrast, *TNIP1* SNPs associated with EOMG show no association with LOMG, nor did other loci suggested for EOMG. SNPs within HLA show strong associations in LOMG, but with smaller effect sizes than in EOMG (highest OR ~2 vs. ~6). Moreover, the specific LOMG associated determinants show a different pattern from EOMG. Indeed, the strongest LOMG HLA determinant is *DQA1*0501* (p = 5. 9x10-12, OR = 0. 54). Association and conditioning studies for the HLA region show three distinct and largely independent association peaks for LOMG corresponding to i) HLA class II (highest attenuation when conditioning on DQA1), ii) HLA-A and iii) HLA class III (equivalent results with SNPs within *DDX39B-ATP6V1G2*, *NFKBIL1*, and *LST1*). Conditioning studies of HLA amino acid residues also suggest potential functional correlates. Together these data emphasize the need to subgroup MG for clinical and basic investigative studies, and imply distinct predisposing mechanisms in LOMG.

884W

Multi-level Genome-wide Association Study on Bone Geometry and Microstructures. *X. Fu, H. Shen, LJ. Zhao, Q. Tian, HW. Deng.* Biostatistics and Bioinformatics, Tulane University, New Orleans, LA.

Osteoporosis isan increasingly prevalent public health problem worldwide. Previousgenetic studiesfor osteoporosis were mainly focused on areal bone mineral density(aBMD). However,bone geometry and microstructuresare alsoplaying critical roles inindetermining bonestrength and risk of fracture. To identify novel osteoporosis susceptibility genes, weperformedamulti-level genome-wide association study (GWAS)onbone geometry and microstructuralcharacteristics. A total of3,209subjects from theFramingham Studywereincluded in the analysis. Eachof subjectswas measured for trabecular volumetricBMD(TBMD), integral vertebral body volumetricBMD(IBMD) and cross-sectional-areal (CSA) byquantitative computed tomography (QCT),as well asaBMD,at lumbar spine. Genotypesfor ~9. 9millionSNPswereimputedand passedquality control. Genome-wide Efficient Mixed Model Association(GEMMA)approach, which takesadvantageofuni-andmulti-variate linear mixed model, was used for the multi-level GWASanalyses. By univariate analyses, we identifiednovel significant associations for SNPs(rs143042601,p=2. 4e-09)in an intergenic region of2q23. 3withTBMD, andSNPs(rs114032442,p=4. 0e-08)proximal tothespalt-like transcription factor 1 (*SALL1*)genewith CSA. Through bi- and multi-variate analyses, we further identified a number of loci that are significantly associated with multipletraits. For instance,severalSNPs mapped tothelatecornifiedenvelope(*LCE*)gene familyon 1q21. 3showed significant association(p= 2. 6e-10~8. 8e-09)withTBMD and IBMD, and SNPs ingene*ZNF678*(rs149363602,p=1. 3e-08)weresignificantlyassociated withall fourphenotypes. In summary,we conducted a multi-level GWAS forbone geometry and microstructuresof lumbar spine and identified a number of novel osteoporosis susceptibility genes. Particularly, some of the identified loci were previously implicatedinregulating bone strength in animal models, but werefor this timeassociated with human bone phenotypes, highlighting the power of bi-/multi-variant analyses for human complex disorders/traits.

885T

Identification of 10 novel uterine leiomyomata susceptibility loci by genome-wide association analysis in population-based conventional and direct-to-consumer cohorts. *C. S. Gallagher[1,2], D. Velez-Edwards[3], R. M. Cantor[4], T. L. Edwards[5], M. Hayden[6], D. A. Hinds[7], J. Jeff[8], Y. Kamatani[9], M. Kubo[9], P. A. Lind[10], S. Low[9], N. G. Martin[11], S. E. Medland[10], G. Montgomery[12], A. Morris[13], Z. Ordulu[6], J. N. Painter[10], J. Perry[14], A. Takahashi[9], J. Y. Tung[7], K. Zondervan[15], D. I. Chasman[16], C. C. Morton[6,17,18,19].* 1) Department of Genetics, Harvard Medical School, Boston, MA, USA; 2) Submitting on behalf of the FibroGENE Consortium; 3) Vanderbilt Epidemiology Center, Department of Obstetrics and Gynecology, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA; 4) Department of Human Genetics, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, CA, USA; 5) Vanderbilt Epidemiology Center, Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA; 6) Department of Obstetrics, Gynecology and Reproductive Biology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 7) 23andMe, Mountain View, CA, USA; 8) Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 9) RIKEN Yokohama Institute, Japan; 10) Quantitative Genetics, QIMR Berghofer Medical Research Institute, Herston, Australia; 11) Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Herston, Australia; 12) Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Herston, Australia; 13) Department of Biostatistics, University of Liverpool, Liverpool, England; 14) Addenbrooke's Hospital, Cambridge, England; 15) Nuffield Department of Obstetrics & Gynaecology, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, England; 16) Department of Medicine, Division of Preventive Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 17) Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 18) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 19) School of Psychological Sciences, University of Manchester, Manchester, England.

Uterine leiomyomata (UL), also known as fibroids, are the most common tumor of the female reproductive tract and have been shown to have heritable risk. Within FibroGENE, a consortium for genome-wide association study (GWAS) of UL, we report a two-stage, large-scale GWAS combining data from a population-based research cohort (Women's Genome Health Study, WGHS) and a direct-to-consumer genotyping service (23andMe) with the goal of discovering novel UL susceptibility loci in women of European ancestry. We first investigated 3,375 self-reported cases and 9,465 controls from WGHS along with subgroup analyses among 2,717 severe cases and 6,359 controls. Two loci, in *WNT4* at 1p36. 12 (rs7412010; *P* = 4. 0x10-8; OR [95% CI] = 1. 24 [1. 15-1. 33]) and *FSHB* at 11p14. 1 (rs74485684; *P* = 6. 1x10-8; OR [95%CI] = 1. 29 [1. 18-1. 42]), revealed strong candidate associations with UL predisposition in the broad and subgroup analyses, respectively. In an independent cohort of 15,068 self-reported cases and 43,587 controls from the 23andMe research participant database, both candidate loci have genome-wide significance: 1p36. 12 (rs7412010; P = 9. 0x10-11; OR [95% CI] = 1. 13 [1. 09-1. 17]) and 11p14. 1 (rs74485684; *P* = 5. 9x10-8; OR [95%CI] = 1. 11 [1. 07-1. 15]). In a genome-wide meta-analysis of the two cohorts, associations at the initial loci were reinforced (rs7412010, *P*=1. 4x10-16; rs74485684, *P*=3. 4x10-10) and we observed additional genome-wide significant associations for 546 SNPs at 11 separate regions, including a region previously reported for UL in a cohort of Japanese women: rs1265164 on 10q24. 33 (*P* = 4. 3x10-10; OR [95%CI] = 1. 12 [1. 09-1. 16]). Strong candidate associations were also observed in a second region reported in the Japanese cohort: rs143521471 (P = 4. 1x10-7; OR [95%CI] = 1. 09 [1. 05-1. 12]). Candidate genes at the 10 novel UL susceptibility regions were previously associated with endometriosis (*GREB1*, *WNT4*) and biological processes relevant to UL: estrogen receptor signaling (*ESR1*), metabolism (*CYP17, SLC38A*), cancer susceptibility, cell growth, or development (*ESR1, TP53, GREB1, PDCD6, FRMD7*), and regulation of sex hormones (*ESR1, WNT4, FSHB*). Our findings provide novel insights into the pathogenesis of UL and suggest a number of potential therapeutic avenues.

886F

Polymorphisms in SLC6A3 and DBH genes of the dopaminergic pathway associated with resistance to M. tuberculosis infection. *K. R. Fluegge[1], R. P. Igo[1], N. B. Hall[1], L. L. Malone[3], H. Mayanja-Kizza[4], W. Henry Boom[2, 3], C. M. Stein[1, 3].* 1) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Department of Medicine, Case Western Reserve University, Cleveland, OH; 3) Case Western Reserve University and Makerere University Research Collaboration, Cleveland, OH; 4) Makerere University School of Medicine and Mulago Hospital, Kampala, Uganda.

Tuberculosis (TB), caused by Mycobacterium tuberculosis (Mtb), is a major public health threat globally, with a high burden in sub-Saharan Africa. According to the World Health Organization in 2013, Uganda's TB incidence rate was 166 per 100,000 people. Human host genetic factors have been demonstrated to be involved in the pathogenesis of TB. Furthermore, the immunology literature also hints that allelic variations within genes relevant to the dopaminergic system may impact susceptibility to infections, including TB. Exposure to Mtb is a necessary but insufficient condition for infection to occur. Of those ultimately infected, only about 10% of individuals develop active TB disease. The two main phenotypes of interest in TB pathogenesis are active TB disease and remaining persistently TST negative (PTST-) despite close and prolonged exposure to Mtb. In this analysis, we used data collected through a large ongoing household contact study in Kampala, Uganda, to examine genetic association with these phenotypes. We focused on seven genes within the dopaminergic system, analyzing 2266 single nucleotide polymorphisms (SNPs) within these genes and within 100 kilobasepairs of these genes that were genotyped on the Illumina Omni5 chip. The 483 subjects from 299 households included in this analysis were enrolled between 2002-2010. Analysis was conducted using generalized estimating equations, clustering the data by family. For each outcome, TB (42%) and PTST- (7%), both the dominant and the additive models were run. We found 11 SNPs in the solute carrier family 6 (neurotransmitter transporter), member 3 (SLC6A3, or dopamine transporter (DAT) gene) gene, 3 SNPs in the dopamine beta-hydroxylase (DBH) gene and 6 SNPs in the DOPA decarboxylase (DDC) gene which were associated with PTST-phenotype ($p < 0.05$). Among the most significant SNPs associated with PTST- were SLC6A3 rs11564771 (6. 96E-05) and DBH rs3025373 (1. 29E-05). However, only the SNP in the DBH gene was significant after multiple testing correction. This particular finding extends previous work on the role of the DBH gene in susceptibility to TB infection in a murine model. We found 9 SNPs in the SLC6A3 gene, 6 SNPs in the DBH gene and 3 SNPs in the DDC gene which were associated with the TB phenotype ($0.01 < p < 0.05$), none of which were significant after multiple testing correction. These findings suggest that the dopaminergic system may play a role in resistance to Mtb infection.

887W

UK BiLEVE: Novel loci and genetic architecture of FEV_1 extremes in the first genetic study in UK Biobank. *R. J. Allen[1], LV. Wain[1], N. Shrine[1], S. Miller[2], VE. Jackson[1], I. Ntalla[1], M. Soler Artigas[1], JP. Cook[1], AP. Morris[3], E. Zeggini[4], J. Marchini[5,6], P. Deloukas[7], A. Hansell[8], R. Hubbard[9], I. Pavord[10], NC. Thomson[11], DP. Strachan[12], IP. Hall[2], MD. Tobin[1,13], UK BiLEVE Consortium.* 1) Department of Health Sciences, University of Leicester, Leicester, United Kingdom; 2) Division of Respiratory Medicine, University of Nottingham, Queen's Medical Centre, Nottingham, United Kingdom; 3) Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom; 4) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom; 5) Department of Statistics, University of Oxford, Oxford, United Kingdom; 6) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 7) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University London, London, United Kingdom; 8) Faculty of Medicine, School of Public Health, Imperial College London, London, United Kingdom; 9) Faculty of Medicine and Health Sciences, School of Medicine, University of Nottingham, Nottingham, United Kingdom; 10) Respiratory Medicine, University of Oxford, Oxford, United Kingdom; 11) Institute of Infection, Immunity & Inflammation, University of Glasgow, United Kingdom; 12) Population Health Research Institute, St George's University of London, London, United Kingdom; 13) National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, United Kingdom.

Airflow obstruction occurs in chronic obstructive pulmonary disease (COPD) which is the 3rd leading cause of death globally. Forced Expiratory Volume (FEV1) is reduced in airflow obstruction. In the UK BiLEVE (UK Biobank Exome Lung Evaluation) study we aimed to investigate the genetic architecture of FEV1 extremes using a nested case-control design. A total of 50,008 individuals of European ancestry with high quality spirometry measures were sampled from the 502,682 individuals in UK Biobank. Among never smokers we sampled 10,000 individuals with the lowest percentage predicted FEV1, 10,000 from the middle of the percent predicted FEV1 distribution and the 5,000 with the highest percentage predicted FEV1. The same sampling method was used in heavy smokers. Using a custom design array (that included 130,000 rare missense and loss of function variants) a case-control analysis of high versus low percentage predicted FEV1 was performed separately in never and heavy smokers. Shared polygenic aetiology of phenotypes defined by FEV1 extremes was analysed using polygenic risk scores, gene expression levels were investigated and a genome-wide pathways analysis was performed. This study identified six novel variants that were genome-wide significant when comparing high versus low percentage predicted FEV1. Loci identified as significant in either heavy or never smokers were at 4q24 on the gene *NPNT* (p=9. 62x10-16), another independent signal at 4q24 in *TET2* (p=1. 31x10-8), at 17q21. 31 covering the *KANSL1*, *MAPT* and *CRHR1* genes (p=1. 66x10-10), on chromosome 6 within the HLA region (p=1. 26x10-10), at 17q25. 1 within *TSEN54* (p=1. 18x10-8) and an intergenic signal near *RBM19* and *TBX5* on chromosome 12 (p=1. 16x10-8). *TET2*, *TBX5* and 10 genes at 17q21. 31 showed differential expression during foetal lung development. The SNP rs2532349 in 17q21. 31 was also found to be associated with mRNA expression levels of 15 genes in the lungs and blood. There was evidence of shared genetic aetiology between low FEV1 and high FEV1 (p=1. 64x10-22). We also showed there is shared genetic aetiology for low FEV1 between heavy and never smokers (p=2. 29x10-16) and for low FEV1 between those with and without doctor-diagnosed asthma (p=6. 06x10-11). Genome-wide pathway analysis identified the histone subset of the chromatin packaging and remodelling process gene set, implicating a role for epigenetic regulation in lung health. This research has been conducted using the UK Biobank Resource.

888T

The International Psoriasis Council Exome Chip Project: Exome arrays reveal known and novel coding variant associations. *N. Dand[1], S. Mucha[2], L. C. Tsoi[3], J. T. Elder[4,5], A. Franke[2], M. A. Simpson[1] on behalf of the International Psoriasis Council Exome Chip Consortium.* 1) Division of Genetics and Molecular Medicine, King's College London, London, UK; 2) Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany; 3) Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 4) Department of Dermatology, University of Michigan Medical School, Ann Arbor, MI; 5) Ann Arbor Veterans Affairs Hospital, Ann Arbor, MI.

Genetic association studies, including genome-wide association studies and the recent international immunochip collaboration, have identified more than 50 common loci associated with psoriasis susceptibility in European populations. However, a substantial fraction of the genetic contribution to psoriasis risk remains undefined. Direct investigation of functional genetic variation has the potential to further elucidate the molecular genetic contributants to the heritability of psoriasis. Focus on protein-altering variation, and particularly low-frequency and rare coding variants, will inform the fine mapping of established common susceptibility loci and has the potential to identify low-frequency and rare susceptibility alleles overlooked by previous studies of common variation. We examined the role of functional protein coding variation in psoriasis susceptibility in a large case-control study incorporating 11,456 cases and 21,355 controls. Genotyping was performed at three centres (Michigan, USA; Kiel, Germany; and London, UK) using Illumina and Affymetrix Exome arrays that are designed to interrogate variants predicted to alter protein sequence. The two platforms share 231,130 variants including 28,181 common (minor allele frequency 5-50%), 21,578 low-frequency (0. 5-5%) and 184,371 rare (< 0. 5%) variants. We undertook a standard error weighted meta-analysis of 72,260 SNPs, outside of the MHC region, which are polymorphic in each of the three datasets. We observed genome-wide association (p < 5x10-8) of common missense alleles within established psoriasis susceptibility loci, including those in *IL13*, *STAT2*, *TRAF3IP2*, *TYK2*, and *ERAP1*. Aggregation tests for low-frequency and rare variation also revealed exome-wide association (p < 2. 5x10-6) for several genes, including *IFIH1*. These observations provide a basis for a more detailed analysis to determine the functional alleles responsible for these association signals.

889F

A genome-wide association study of uterine fibroids confirmed by imaging in eMERGE. *T. L. Edwards[1], M. J. Bray[10], K. S. Tsosie[1], K. E. Hartmann[2], J. C. Denny[3], J. Jeff[4], E. Kenny[4], P. L. Peissig[5], C. A. McCarty[5], M. S. Williams[6], R. P. Gogoi[6], H. S. Kuivaniemi[6], G. C. Tromp[6], A. Kho[7], J. A. Pacheco[7], K. L. Jackson[7], I. J. Kullo[8], J. Pathak[8], E. A. Stewart[8], D. M. Roden[9], D. R. Velez Edwards[10].* 1) Vanderbilt Epidemiology Center, Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University; 2) Department of Obstetrics and Gynecology, Institute for Medicine and Public Health, Vanderbilt University; 3) Department of Biomedical Informatics, Department of Medicine, Vanderbilt University; 4) Mount Sinai School of Medicine; 5) Marshfield Clinic; 6) Geisinger Health System; 7) Northwestern University; 8) Mayo Clinic; 9) Division of Cardiology, Department of Medicine, Department of Pharmacology, Vanderbilt University; 10) Department of Obstetrics and Gynecology, Vanderbilt Genetics Institute, Vanderbilt University.

Up to 77% of women have a uterine fibroid (UF) by menopause. UF are the leading indication for hysterectomy and account for $9B to $39B in annual healthcare costs. Risk of UFs is heritable with large disparities in UF risk, symptoms, and severity among ancestral groups. Recent African ancestry (AA) populations have a high burden of UF, with 2-3 fold higher prevalence, more severe symptoms, and earlier onset than Europeans (EA). Because up to 51% of UF cases are unaware of fibroid status and are misclassified as controls in surveys, association studies of UF risk without UF status confirmed by imaging suffer from significant information bias. We conducted an association study of rare and common genetic determinants of UF risk confirmed by pelvic imaging from electronic medical records (EMR) from 6 eMERGE Network sites (N=10,227; AA:1,354 cases and 1,419 controls; EA:2,843 cases and 4,611 controls). All subjects were classified using a validated EMR algorithm (PPV 98%, NPV 96%). All samples were genotyped by GWAS arrays and a subset was also genotyped by exome arrays (AA: 580 cases, 804 controls; EA: 1195 cases, 1164 controls). Genotypes were imputed using the 1KG reference and UF risk was modeled as a function of PCs of ancestry and genotype by logistic regression, followed by fixed effects meta-analysis of results (meta-analysis λGC=1. 02). The most statistically significant associations from transethnic analyses with meta-data from all cohorts were the intergenic SNPs rs2983218 at chr6q27 (OR=1. 23, p=8x10-10) and rs13103279 at chr4q34. 3 (OR=0. 81, p=2x10-9) near genes involved in angiogenic pathways, with consistent effects across samples (Cochran's Q p>0. 5). Statistically significant associations were observed in the subset of African ancestry cohorts for the rare (MAF<0. 01) variants rs190615976 at chr17 near *IGF2BP1* (OR=3. 82, p=2x10-10), chr2 at rs77950674 near *RUFY4* (OR=4. 15, p=7x10-10) and rs116003406 in the 3'UTR of *SLC9A4* (OR=3. 57, p=3x10-9), and chr4 at rs115077882 in *ARFIP1* (OR=1. 98, p=5x10-9). In the European subset the rare variant rs148308239 near *NEBL* on chr10 was significantly associated with UF risk (OR=0. 44, p=5x10-8). Ongoing genotyping experiments and additional cohorts will provide 4,864 imaging-confirmed samples for multistage replications of these associations. These results suggest UF risk is influenced by common and rare variants, and ancestral genetic heterogeneity in UF risk is likely.

890W

Clinical and genetic characteristics of novel emphysema distribution subtypes identified by unsupervised learning analysis. *A. El Boueiz[1,2], M. H. Cho[1,2], G. R. Washko[2], R. San José Estépar[3], D. A. Lynch[4], E. K. Silverman[1,2], D. L. DeMeo[1,2], P. J. Castaldi[1,5].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States; 2) Pulmonary and Critical Care Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States; 3) Surgical Planning Laboratory, Laboratory of Mathematics in Imaging, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States; 4) Division of Radiology, National Jewish Health, Denver, Colorado, United States; 5) Division of General Medicine and Primary Care, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, United States.

Background: Emphysema, irreversible destruction of lung parenchyma, is an important component of chronic obstructive pulmonary disease (COPD) with considerable variability in distribution and severity. Recent evidence suggests that a significant amount of this variability is due to genetic factors. To assess the clinical impact and determine the genetic influences of emphysema distribution, we examined lobar emphysema in the COPDGene cohort. **Methods**: 9213 NHW and AA smokers were included. Genotyping was performed on the Illumina Omni Express platform with additional markers imputed using the 1000 Genomes reference. Emphysema percentage (i. e. CT densitometry low-attenuation area less than -950 Hounsfield units) was measured in each lung lobe. Lobar emphysema variables were regressed on total emphysema and lobe-specific residuals were extracted. Unsupervised learning with random forest predictors was then used to cluster these residuals. GWASs were performed in the identified clusters per racial group to look for emphysema distribution risk loci. These clusters were also tested for associations with clinical outcomes and genetic variants previously associated with COPD. **Results**: We identified one "unclustered" group (n=6197) and three clusters that can be characterized as 1) Controls with no or minimal emphysema (n=1312), 2) Upper lobe predominant emphysema (n=905), and 3) Lower lobe predominant emphysema (n=799). GWASs, conducted in each lobar predominant cluster compared to the control cluster, identified a novel genome wide significant locus (9q31 near *ZNF462*; *P=E-8*) in the NHW upper lobe group. The other top hits were 14q21 near *DLGAP5* (AA upper lobe; *P=3E-7*), 3p14 near *FHIT* (NHW lower lobe; *P=6E-7*), and 8p12 near *TTC23* (AA lower lobe; *P=2E-7*). In the candidate analyses, strong associations were seen with rs13141641 near *HHIP (P=3E-6)* and rs2070600 near *AGER (P=E-5) respectively* in NHW upper and lower lobe groups vs control. Clinically, upper lobe predominant subjects had a more rapid disease progression over 5-year follow up. **Conclusion**: This clustering of densitometric lobar emphysema identifies subgroups of smokers with different emphysema distribution patterns, clinical features, and genetic signatures. These findings suggest that emphysema distribution is a clinically and biologically important COPD phenotype that can be used fordisease subtyping to ultimately improve and personalize therapy. **Grant support:** NIH R01s HL089897, HL089856, HL124233.

891T

Y chromosome variation and complex traits: the Ygen consortium. *I. Gandin[1,2], P. K. Joshi[2], P. Gasparini[1,3,4], T. Esko[5,6,7,8], J. F. Wilson[2,9], The Ygen consortium.* 1) Department of Medical Sciences, University of Trieste, Strada di Fiume 447 - Osp. di Cattinara, Trieste, 34149, Italy; 2) Centre for Global Health Research, Usher Institute for Population Health Sciences and Informatics, University of Edinburgh, Teviot Place, Edinburgh, EH8 9AG, Scotland; 3) Sidra Medical and Research Center, Doha, Qatar; 4) Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", via dell'Istria 65, Trieste, 34137, Italy; 5) Estonian Genome Center, University of Tartu, Riia 23b, 51010, Tartu, Estonia; 6) Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Cambridge, 02141, MA, USA; 7) Program in Medical and Population Genetics, Broad Institute, Cambridge Center 7, Cambridge, 02242, MA, USA; 8) Department of Genetics, Harvard Medical School, 25 Shattuck St, Boston, 02115, MA, USA; 9) MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, Scotland.

The genetics of complex traits have been studied in great detail using Genome-Wide Association Studies (GWAS), leading to substantive insights regarding the genetic determinants of quantitative traits and disease mechanisms. However, there remain significant parts of the genome which are not routinely assayed in GWA studies, the largest of which is the Y Chromosome. Only carried by men and constitutively haploid, the Y Chromosome totals about 25 Mb of sequenceable DNA containing 27 genes or gene families, including genes known to be implicated in Mendelian conditions such as infertility and deafness. Nevertheless, the Y Chromosome has been almost completed neglected in the context of complex traits and diseases, until a recent study that associated Y Chromosome variants and risk of coronary artery disease. Here we present Ygen, an international consortium that gathers together a large number of collaborating population-based cohorts. Ygen allows us to investigate the role of the Y chromosome in complex traits for the first time in an adequately powered setting. In the project more than 30 diverse traits of public health or evolutionary interest are considered, including anthropometric measures, haematological counts, cardiovascular risk factors, metabolic markers, inflammatory response and fertility. The study uses 70 SNPs included in the Illumina HumanExome Beadchip that were specifically designed to represent Y Chromosome diversity across the world. Genotypes are used to define haplogroups which in turn provide information on Y Chromosome lineages, thus increasing the number of SNPs tested. These are analysed using a single-SNP association approach and results combined in a meta-analysis. We will present results of the ongoing study which already includes more than 10,000 men. This will include an overview of quality control, focusing on genealogical consistency and SNP performance in haplogroup estimation, and the results of the association study.

892F

Electronic health records empower the first genome wide association study of developmental dysplasia of the hip. *K. Hatzikotoulas[1], M. J. Clark[2], D. M. Eastwood[3], J. M. Wilkinson[2], E. Zeggini[1].* 1) Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; 2) Academic Unit of Bone Metabolism, Northern General Hospital, University of Sheffield, Sheffield, United Kingdom; 3) Consultant Paediatric Orthopaedic Surgeon The Catterall Unit, Royal National Orthopaedic Hospital, Stanmore Middlesex HA7 4LP.

Developmental dysplasia of hip (DDH) is a disorder characterised by incomplete development of hip joint. Although DDH is one of the most common skeletal dysplasias, the incidence of DDH is approximately 1. 5 cases per 1. 000 births, and it remains slow and costly to recruit large-scale patient cohorts for powerful genetic association studies. The UK has an established and expanding e-health record (EHR) infrastructure that provides a comprehensive platform upon which to develop genomic research of clinical relevance and impact. The National Joint Registry for England, Wales, and Northern Ireland (NJR) is the largest registry of joint replacement in the world, with 1. 9 million recorded procedures, adding new procedures at a rate of 200. 000 each year. In this work we have successfully used the NJR as a platform to generate a DDH biobank of 903 individuals. Following genotyping on the Illumina CoreExome platform, we carried out the first ever genome-wide association study (GWAS) for DDH in 840 patients and 3364 controls. Following stringent quality control, ~450. 000 variants were tested for association with DDH. We identified 34 independent variants with p<1x10-5 (binomial p<2. 2x10-16). Among these, rs143384 reached genome-wide significance (allele A, allele frequency 0. 60, OR[95% CI] 1. 58[1. 41-1. 77], p=8. 5x10-16). Rs143384 is a 5' UTR variant in *GDF5*, which plays a central role in cell growth and differentiation in both embryonic and adult tissues. Rs143384 has previously been associated with height, chondrodysplasia and osteoarthritis, and has also been associated with DDH in a candidate gene study with p=0. 002. We are currently following up further GWAS signals in replication cohorts. Here, we establish the first genome-wide significant locus for DDH, discovered through linking EHRs with genomics as a proof of principle in enabling powerful genetic association studies of relatively rare but complex diseases.

893W

Assessing the Shared Genetic Basis of Benign Prostate Hyperplasia and Prostate Cancer; The Kaiser Permanente Genetic Epidemiology Research on Adult Health and Aging (GERA) Cohort. *J. D. Hoffman[1], C. G. Tai[1], T. J. Hoffmann[1,2], E. Jorgenson[4], C. P. Quesenberry[4], J. Shan[4], D. Aaronson[4], J. Presti[4], L. A. Habel[4], C. R. Chao[4], N. R. Ghai[4], D. K. Ranatunga[4], C. Schaefer[4], N. J. Risch[1,2], J. S. Witte[1,2,3], S. K. Van Den Eeden[3,4].* 1) Department of Epidemiology and Biostatistics, UC San Francisco, San Francisco, CA; 2) Institute for Human Genetics, UC San Francisco, San Francisco, CA; 3) Department of Urology, UC San Francisco, San Francisco, CA; 4) Kaiser Permanente, Division of Research, Oakland CA.

Benign prostatic hyperplasia (BPH) is a non-cancerous highly prevalent form of prostate enlargement observed within aging males; prostate cancer is one of the most common forms of cancer in the elderly male population. Although these two diseases frequently co-occur, little is known on whether they have a shared genetic basis. Using the Kaiser Permanente GERA cohort, we set out to examine whether known risk SNPs for prostate cancer were predictive of BPH. Our analysis included published results from the literature for prostate cancer and 11,090 men with BPH and 20,371 men without this condition, all of European descent. A genetic risk score incorporated 105 known prostate cancer variants, weighting each variant observed among BPH cases and controls by their respective log-odds ratios for prostate cancer as reported in the literature. Risk scores were adjusted for the first ten principal components and age at exam within a multivariate logistic regression model. Our results show that the genetic risk score was not significant in predicting BPH risk (p = 0. 67). When the 105 prostate cancer risk variants were analyzed individually, we observed a significant genetic association with rs1270884 and BPH (p = 1. 98*10-4). This variant is an intergenic SNP located at 12q24. An additional ten variants out of the 105 were nominally associated with BPH (p < 0. 05). In summary, although previous applications of the same risk score in the GERA cohort have shown highly significant associations with prostate cancer and PSA levels, we do not observe that known prostate cancer variants are predictive of BPH risk. Further analyses exploring co-heritability of BPH and prostate cancer, as well as the genetics of BPH independently, at the genome-wide level are warranted.

894T

GWAS of Urinary Metabolites within the Framingham Heart Study. *J. E. Huffman[1], S. -J. Hwang[1], G. M. McMahon[1,2], C. B. Clish[3], R. E. Gerszten[3,4], C. S. Fox[1,5].* 1) The Framingham Heart Study, NHLBI Population Sciences Branch, Framingham, MA; 2) Division of Nephrology, Brigham and Women's Hospital, Boston, MA; 3) Broad Institute, Cambridge, MA; 4) Cardiology Division, Massachusetts General Hospital, Boston, MA; 5) Division of Endocrinology and Metabolism, Brigham and Women's Hospital, Boston, MA.

The field of metabolomics provides a platform for the rapid analysis of many metabolites within a biological sample and is a rich source of potential biomarkers. Urinary metabolites allow insight into the health and function of the kidney and may serve as biomarkers for dysregulation and disease. In order to investigate the biological mechanisms associated with these urinary metabolites, we examined their associations with common genetic variation within the Framingham Offspring Cohort. Urinary metabolites (n=154) were measured among 384 exam 6 participants using liquid chromatography mass spectrometry. Each metabolite was natural log-transformed, adjusted for age and sex, then the genetic association tested with HapMap2 imputed genetic dosages (n=371 participants) after accounting for relatedness. SNPs were filtered to remove those which were poorly imputed or rare and deemed significant at a p-value < 5E-08. This was followed by a look-up of significant results in published findings from complementary studies and in GWASs of kidney function traits from the CKDGen consortium. Significant associations were found at 21 loci across 19 metabolites. SNPs within *AGXT2* were associated with levels of aminoisobutyric acid (rs37370, beta=1. 13, p=1. 92E-23) which replicates previous associations of this loci with both urine and serum concentrations. The *SLC7A9* region was associated with lysine levels (rs12460876, beta=0. 41, p=1. 05E-12) and has been associated with eGFRcrea levels (beta=+, p=5. 50E-09) and with urinary lysine and lysine-containing ratios. rs2286963 in *ACAD* was associated with concentrations of nonanoylcarnitine (beta=0. 36, p=1. 15E-11) and this locus was associated in a previous GWAS of urinary metabolites with an unidentified metabolite. The top SNP in all other significant regions had p-values > 7E-03 with kidney-related traits in the CKDGen consortium. They were also not reported to be associated with the corresponding urinary metabolite in previous studies but only the significant results are publically available and the metabolite in question may not have been measured. The protein products for several of the associated genes are involved in the metabolic reaction containing the associated metabolite. In this study we identified both known and novel loci associated with various urinary metabolites. Despite our small sample size, we achieved highly significant p-values which shows how amenable metabolomic data is to GWAS.

895F

Genome-wide association studies of upper airway microbial composition in a founder population. *C. Igartua[1], E. R. Davenport[1], Y. Gilad[1], C. Ober[1], J. M. Pinto[2].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Section of Otolaryngology-Head and Neck Surgery, Department of Surgery, University of Chicago, Chicago, IL.

Whether host genetic variation can shape the airway microbiome remains an open question that is relevant for a number of respiratory diseases. In this study, we mapped the relative abundance (RA) of bacterial taxa from the upper airway in the Hutterites, a founder population of European decent. As a result of their communal, farming lifestyle, individuals are exposed to remarkably similar environmental exposures. For these studies, we selected ~170k variants (MAF>10%, CR>90%, LD-r2<0. 5) from among all variants discovered in Hutterite genome sequences and examined associations with the RA of bacterial genera from adult Hutterites at two accessible sites, the nasopharynx (NP) and the nasal vestibule (NV), in two seasons, summer (NP N=88, NV N=87) and winter (NP N=77, NV=80). RAs of the 16S rRNA gene (V4) were calculated at the genus level after subsampling 250k sequence reads per individual. To identify variants associated with normalized bacterial abundance in genera present in at least 75% of individuals (summer NP=91 and VP=79, winter NP=59 and VP=52), we used a linear mixed model that included sex, age, and date of collection as fixed effects and kinship as a random effect to adjust for relatedness (GEMMA). In the summer samples, the RA of four bacterial taxa in the vestibule and three in the nasopharynx were associated with host genotypes at a FDR <5%. The strongest association in this season was between the genus *Dermacoccus* in the NV and an intergenic variant 8kb upstream of *TINCR* (rs62105955; *P*= 1. 52x10-8; FDR= 0. 006%), a long non-coding RNA gene that directly binds to PGLYRP3,a protein that binds to peptidoglycans of gram-positive bacteria. In the winter samples, we identified 2 bacterial genera in each of the two anatomic sites that were associated with host genotypes (FDR< 5%). For example, 6 SNPs were associated with the RA of genus *Aerococcus* in the NP, one of which is an exonic variant in *PGLYRP4* (rs3006448; *P*=1. 28x10-6; FDR=0. 03), another peptidoglycan recognition protein known to have bactericidal activity towards gram-positive bacteria. Our findings support the concept that host genetic variation can influence airway microbial composition and suggests that innate immunity genes may mediate such relationships. Studies of the interaction between host genetics, alterations in the microbiome, and airway biology may provide insights into design of novel therapies across a range of respiratory diseases. This work is supported by R01 HL085197.

896W

Identifying genetic associations with waist-to-hip ratio in Hispanic/Latino ancestral background groups: the HCHS/SOL Study. *A. Justice[1], K. Young[1,2], S. M. Gogarten[3], M. Graff[1,2], L. Sánchez-Johnsen[4], C. A. Laurie[3], C. R. Isasi[5], C. C. Laurie[3], K. E. North[1,2].* 1) Epidemiology, University of North Carolina, Chapel Hill, NC; 2) Carolina Population Center, University of North Carolina, Chapel Hill, NC; 3) Biostatistics, University of Washington, Seattle, WA; 4) Psychiatry, University of Illinois at Chicago, Chicago, IL; 5) Epidemiology & Population Health, Department of Pediatrics, Albert Einstein College of Medicine, Bronx, NY.

Central obesity is a leading public health concern and its prevalence has more than doubled since the 1980s, with the greatest burden carried by ethnic minority populations, and in particular Hispanics/Latinos. Emerging evidence suggests that genetic factors contribute to the obesity burden overall and to population-specific differences. We aim to 1) identify novel loci associated with waist-to-hip ratio adjusted for BMI (WHR), a relative comparison of central adiposity after accounting for overall body size, using the Hispanic Community Health Study/Study of Latinos (HCHS/SOL); 2) determine if differences in genetic associations differ by ancestral background (Caribbean versus mainland populations). Our analyses included 7,472 women and 5,200 men of mainland (Mexican, Central and South American) and Caribbean (Puerto Rican, Cuban, and Dominican) ancestry residing in the US, genotyped on the Illumina SOL custom Omni2. 5M array, imputed to the 1000 genomes Phase I reference panel. Due to known differences in genetic effects on WHR between males and females, we analyzed associations stratified by sex for WHR in addition to combined sexes using a linear mixed model regression, assuming an additive genetic model, adjusted for age, age2, study center, sample weights, population structure, relatedness, and sex in the combined analysis. For men, we identified one novel genome-wide significant (GWS) SNP near *PLEKHG4B* (P=2. 01E-8). No SNPs reached GWS for women-only or combined sexes in the primary analyses. However, one novel SNP near *NTM* reached GWS in the women-only mainland background group (P=5. 47E-8 in all women, P=1. 63E-8 in mainland women, I2=73). Many of the other loci also exhibit background-specific associations, possibly due to population history and especially admixture. We found notable differences in linkage disequilibrium and allele frequencies around our tag SNPs. For example, variants that exhibited association in Caribbean populations and not in mainland groups tended to be African-specific alleles or have higher frequencies in African descent populations. While differences in allele frequencies were not statistically significant, the differences in allele frequencies by background group were large enough to effect power in one group. These observations highlight the importance of large-scale genomic studies in ancestrally diverse Hispanic/Latino populations for identifying obesity-susceptibility that are ancestry or background-specific.

897T

Genome-wide association meta-analysis identifies novel variants associated with fasting plasma glucose in East Asians. *B. Kim, J. Hwang, B. Han.* Div. of structural and functional genomics, Korea National Institute of Health, Cheongju-si, Chungcheongbuk-do, South Korea.

Fasting plasma glucose (FPG) levels are tightly regulated as a part of metabolic homeostasis. Failure in blood glucose regulation can lead to elevated FPG levels, representing an independent risk factor for T2D and a predictor of cardiovascular disease. The fasting glucose level is a moderately heritable trait with the heritability around 30%. A considerable number of genetic determinants influencingfasting glucose levels have been identified from numerous genetic studies in the last few years. The total heritability of fasting glucose levels, however, is yet to be fully explained. To date, 39genetic loci harbouring variants associated with FPG have been identified from genome-wide association (GWA) studies and GWA meta-analyses that were conducted in populations of European ancestry. The genetic basis of glycemic regulation has not been fully explored in non-European populations with only one study in East Asians which identified a single locus associated with FPG (rs895636 at the *SIX2-SIX3* loci. Consideringdifferences in the allele frequencies and linkage disequilibrium structuresamong ethnic groups, large scale genetic studies in populations of non-European ancestries may increase the chance to detect additionalnovel genetic loci for FPG. Genome-wide association meta-analysis has an advantage to identify geneticvariantswith small effect size and low allele frequency that were hardly detected in a single GWA study. Therefore, in this study, we aimed to identify novel loci influencing fasting glucose variation by conducting GWA meta-analysis in East Asian populations. We conducteda two-stage association study, comprisinga discovery set (stage1) of 24,740 individualsfrom the Asian Genetic Epidemiology Network (AGEN) and follow-up*de novo*genotyping replication set (stage 2) of 21,345 individuals from independent East Asian populations. Three new FPG loci reach genome-wide significance in or near *PDK1-RAPGEF4, KANK1* and *IGF1R*. Our results could provide additional insight into the genetic variation implicated in fasting glucose regulation.

898F

Identification of four novel risk loci and functional causative elements in nonsyndromic cleft lip with or without cleft palate. *K. U. Ludwig[1,2], A. C. Boehmer[1,2], B. Nagarajan[3], H. Schuenke[1,2], J. Klamt[1,2], J. Hecker[4], K. Aldhorae[5], A. Rojas-Martinez[6], L. Goelz[7], A. Rada-Iglesias[8], H. Fier[4], J. L. Cotney[9], M. M. Noethen[1,2], M. Knapp[10], B. Odermatt[3], E. Mangold[1].* 1) Institute of Human Genetics, University of Bonn, Bonn, Germany; 2) Department of Genomics, Life&Brain Center, Bonn, Germany; 3) Institute of Anatomy, University of Bonn, Bonn, Germany; 4) Institute for Genomic Mathematics, Bonn, Germany; 5) Department of Orthodontics, College of Dentistry, Thamar University, Thamar, Yemen; 6) Department of Biochemistry and Molecular Medicine, School of Medicine, Universidad Autonoma de Nuevo Leon, Monterrey, Mexico; 7) Department of Orthodontics, University of Bonn, Bonn, Germany; 8) Center for Molecular Medicine, University of Cologne, Cologne, Germany; 9) Deptartment of Genetics and Genome Sciences, University of Connecticut School of Medicine, Farmington, CT 06030, USA; 10) Institute of Medical Biometry and Informatics and Epidemiology, University of Bonn, Bonn, Germany.

Nonsyndromic cleft lip with or without cleft palate (nsCL/P) is a common congenital malformation with multifactorial etiology. Here, we imputed genotype data from a recent meta-analysis (Ludwig et al. 2012, Nat. Genet.) in order to identify novel risk loci. These data were then also used to delineate causative elements at known risk loci. Genome-wide imputation was performed for 399 cases, 1,318 controls, and 666 trios of Central European descent. One previously unreported variant showed genome-wide significance (rs6740960 at 2p21, P=1. 6x10-08). This variant maps 5bp adjacent to a predicted Myc binding site, which is located upstream of the *PKDCC* gene. In a previous report, this gene was downregulated in mice carrying a deletion of the region homologous to the major nsCL/P risk locus at 8q24 (Uslu et al. 2014). Genotyping of 35 selected SNPs with P<10-03 in an independent nsCL/P sample (609 cases, 1,745 controls) revealed three additional genome-wide significant loci in the combined sample (2p24. 2, 14q22. 1, 15q13. 3). Notably, a significant positional overlap of association signals was observed in functional datasets of relevance to craniofacial development. Also we discovered that 30% of the genetic heritability could be attributed to common variants within this dataset, with around 24% being attributable to the 20 established risk loci identified through the present study and previous research. This novel data set was then used to finemap known susceptibility regions and identify candidate variants for functional follow-up. For example, the top-associated genotyped variant rs8001641 at 13q31 maps to a non-coding region upstream of *SPRY2*. In the imputation data, rs1854110 showed a larger odds ratio than rs8001641, and this result was supported by independent replication. Notably, rs1854110 was located in a putative regulatory enhancer element identified in craniofacial tissue. After cloning this element into a reporter vector based on Tol2/E1b-mediated transgenesis, GFP expression was observed in cephalic regions in zebrafish embryos injected with the plasmid, thus demonstrating the functional effect of the regulatory element. In summary, investigation of a high-density dataset identified four novel risk loci for nsCL/P, and provided evidence for causality of regulatory effects. The present study demonstrates that the integration of genomic and functional datasets is a powerful approach to the investigation of craniofacial development and nsCL/P etiology.

899W

Copy number variation in *FCGBP* gene is associated with endometriosis development. *F. A. Mafra[1,2], D. M. Christofolini[2], B. Bianco[2], C. P. Barbosa[2], R. Pellegrino[1], H. Hakonarson[1].* 1) Center for Applied Genomics. The Children's Hospital of Philadelphia. 3615 Civic Center Boulevard, Abramson Research Center, Suite 1014. Philadelphia, PA, USA; 2) Center of Human Reproduction and Genetics. Faculdade de Medicina do ABC. Av. Principe de Gales, 821, CEPES Lab 101. Santo Andre, SP, Brazil.

Introduction:Endometriosis is a steroid-dependent condition in which a tissue that is histologically similar to the endometrium with glands and stroma grows outside the uterine cavity. This condition play an important role in infertility, a very common gynecologic problem that affects approximately 15-20% of couples who attempt pregnancy. Gene interactions, pointing to associations between the development of endometriosis and genetic polymorphisms have been described. The Copy Number Variation (CNVs) investigation approach can be an important step in discerning a possible pathophysiology of endometriosis. **Objective**: Identification of genomic regions that might contribute to the development of endometriosis by identifying CNVs (primary identified by microarrays) that differentiate cases from healthy controls. **Material and Methods:** Case-control study comprising 564 infertile women with endometriosis and 652 fertile woman of the control group. A total of 150 samples were selected and analyzed according to DNA pooling methodology. Experiments were performed using Beadchip HumanOmni2. 5 (Illumina). Real-time qPCR was performed in all subjects to validate the candidate CNVs in the *FCGBP*, *NR5A1*, *PTGES2*, *SLC25A25* and *CXXC5* gene regions. Approval Research Ethics Committee (CEP FMABC n. 310. 094). **Results:** We found 49 CNVs loci present in patients with endometriosis and absent in the control group. Of these, 33 were kept and 16 were excluded from the study; six due to size smaller than 1 kb and 10 for not present the minimum number of probes required to be considered as a reliable CNV. 48% (16/33) of the CNVs were located at subtelomeric regions, while 17 (52%) were intrachromosomal. Four loci were selected for validation based on the gene functions present in each locus, and its possible correlation with the development and progression of endometriosis. The 4,678 bp duplication involving *FCGBP* gene was observed in 8 patients and was absent in controls, thus conferring significantly increased endometriosis risk. **Conclusion:** This is the first report (to date) of CNV data in a Brazilian population with endometriosis. We have identified one locus involving the *FCGBP* gene on chromosome 19q13. 1 possibly associated with disease risk. The discovery of how these CNVs act can be an important step in discovering the pathophysiology of endometriosis. **Support:** This work was supported by FAPESP #2011/01363-7 #2012/22394-8 #2014/07136-8.

900T

Exome array-based genome-wide association study of fasting glucose level in Korean population. *s. Moon, M. Hwang, B. Kim.* Div. of structural and functional genomics, Korea National Intsitute of Health, Cheongju-si, Chungcheongbuk-do, South Korea.

Elevated blood glucose is associated with increased risk of type 2 diabetes (T2D). Until now, more than 50 loci associated with fasting glucose (FG) and fasting insulin (FI) level have been identified by genome-wide association studies (GWAS). However, a substantial proportion of the genetic contribution remains unclear. Here, we explore the role of coding variants on FG by exome array-based genome-wide association study in Korean population. A total of 14,952 samples as a part of the Korean Genome Epidemilogy Study (KOGES) were genotyped using Illumina HumanExomeBeadChip. Subsequenctly, through quality control processes such as exclusion of relatives, samples with sex inconstant, variants with missing, and low call rate, 14,032 samples remained. The we performed GWAS of FG using EPACTS software. From the results, a SNP nearby MTNR1B satisfied genome-wide significance (6. 5x10-7) in the discovery stage. After the replication study, we discovered 15 additional variants which have suggestive combined P-values. Most of significant variants were overlapped with previously known SNPs on several genes such as HECTD4, MYL2, KCNQ1, SLC30A8 and GCKR. Particularly, several loci were enriched on 12q24. Moreover, we newly identified an intron SNP on ACAD10 and two nonsynonymous SNPs on ALDH2 and BRAP, respectively. These findings provide evidences of genetic influence related to FG.

901F

Replication study of whole genome association for caries in a Guatemalan population. *N. Mukhopadhyay[1], M. Govil[1], C. A. Sanchez[1], F. W. -B. Deleyiannis[2], K. Neiswanger[1], J. M. Resick[1], A. R. Vieira[1], R. M. Silva[3], A. M. Letra[3], M. L. Marazita[1].* 1) Oral Biology, Univ Pittsburgh/Sch Dental Med, Pittsburgh, PA; 2) Department of Surgery, Plastic and Reconstructive Surgery, University of Colorado School of Medicine, Denver, Colorado; 3) Department of Endodontics and Craniofacial Research Center, The University of Texas Health Science Center School of Dentistry at Houston, Houston, Texas.

Dental caries is one the most common chronic diseases world wide, posing a high health burden. Both genetic and environmental factors affect the severity of caries, and several studies have been conducted to identify these factors in Caucasian and African-American cohorts from rural Appalachia, Pittsburgh and Iowa, and in the Philippines and Turkey. Findings have implicated genes for enamel formation, taste preferences, and host immune response. In this study we attempted to replicate prior association findings in a Guatemalan cohort ascertained for non-syndromic oral clefts. Genome-wide association (GWA) analysis was carried out using the PLINK program for a novel age-adjusted quantitative caries phenotype based on counts of decayed primary and permanent teeth excluding precavitated lesions, and missing or filled teeth. We used 379 subjects aged 2-60 years with phenotype information, including 52 individuals with no caries and 327 affected individuals whose caries indices ranged from 1 to 28 (mean=7. 32, SD=6. 36). Caries indices of the 379 phenotyped subjects were adjusted for age using LOESS fitting with a second-degree polynomial. Genotyping was performed with the Illumina HumanHap 550 SNP panel. Guatemalans have a mixed Caucasian/Native American ancestry; therefore, we corrected for population substructure using ancestry principal components derived from a larger set of 1,559 genotyped individuals. A genome-wide significant association was observed on chromosome 21 (p-value 8. 2E-8) near DSCAM. Suggestive associations (7. 9E-5 ≥ p-value ≥ 3. 4E-6) was observed on chromosomes 1 (HCRTR1, CSRP1, KMO), 6 (SMAP1), 12 (CAPS2) and 19 (DUXA). DSCAM is a candidate for Down syndrome, HCRTR1 regulates feeding behavior and glucose uptake, CAPS2 encodes a calcium-binding protein, and DUXA is involved in early embryonic development. These genes have not been reported in previous GWA studies, possibly due to the fact that Guatemala samples come from a different ancestral background, and/or our use of a novel quantitative phenotype. Work is underway to leverage existing familial relationships in a quantitative association framework. This work was supported by NIH-NIDCR R00-DE018085, U01-DE018903, R21-DE016930, R01-DE016148. Visits to Guatemala were conducted in collaboration with Children of the Americas. All analyses were carried out on a Centos 276-node cluster, Indy, supported by the School of Dental Medicine at the University of Pittsburgh.

902W

Identification of three additional susceptibility loci for ulcerative colitis through an Immunochip analysis in Koreans. *H. Oh[1], M. Hong[1], H. Choi[1], J. Baek[1], T. Haritunians[2], B. D. Ye[3], S. H. Park[3], J. Liu[4], D. P. B. McGovern[2], S. Yang[3], K. Song[1].* 1) Biochemistry and Molecular Biology, University of Ulsan College of Medicine, Seoul, South Korea; 2) Cedars-Sinai Medical Center, Los Angeles, USA; 3) Department of Gastroenterology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; 4) Genome Institute of Singapore, Singapore.

We have performed an immunochip study in 1,685 Korean patients with ulcerative colitis and 3,872 controls. We confirmed 3 reported loci at genome-wide significance: *IL23R* at 1p31 (rs76418789; OR = 1. 75, *P* = 1. 25 x 10-8), *IRF5* at 7q32 (rs4728142; OR = 1. 40, *P* = 3. 17 x 10-8), *JAK2* at 9p24 (rs1830610; OR = 1. 35, *P* = 2. 28 x 10-9). Additionally, eight European IBD loci showed significant associations with UC, increasing the number of known UC risk loci to 11 in Koreans. The percentage of phenotype variance explained by the 11 risk loci was 5. 26 % in Koreans (on the liability scale, population prevalence = 0. 0308%). Our finding of overlapping UC susceptibility loci between Korean and European populations strengthen the hypothesis that genetic associations for UC tend to overlap more among different ethnic groups than for CD.

903T

Investigation of genetic risk factors of very low birth weight infants within the German Neonatal Network. *M. Preuss[1], A. Ziegler[1], E. Herting[2], W. Göpel[2].* 1) Institute of Medical Biometry and Statistics, University of Luebeck, Luebeck, Schleswig-Holstein, Germany; 2) Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Germany.

Very Low Birth Weight (VLBW) infants have substantially increased mortality and morbidity rates, but the factors influencing long-term development are not well understood. The German Neonatal Network (GNN) was founded in 2009 to identify genetic, clinical and social factors influencing etiology and long-term development of VLBW. Clinical information includes oxygen demand, administration of surfactant, catecholamine, steroid hormones and bronchopulmonary dysplasia (BPD), brain haemorrhage (IVH), sepsis and death among others. The cohort size is 20,000, and the recruitment includes more than one quarter of all German VLBW per year. DNA samples from more than 9000 VLBW as well as buccal swabs from mothers have been collected from a total of 54 participating German hospitals. Approximately 2600 VLBW from GNN were genotyped on the Axiom™ Genome-Wide CEU 1 Array, and replication was performed in another 4400 GNN VLBW. Results of the initial genome-wide association study revealed genome-wide significance (p <5E-08) for several traits. An interesting finding is for the use of surfactant during hospital stay with an association to *LINGO2* (lead SNP rs4878404, initial p=5E-06, replication one-sided p=2. 3E-03). These results demonstrate that GNN is a unique resource for genetic and pharmacogenetic studies in VLBW.

904F

Novel genetic loci and Alzheimer's disease genes influence retinal nerve fibre layer thinning: an European ancestry meta-analysis of 10,502 individuals. *C. Venturini[1,2], H. Springelkamp[3], J. Lamparter[4], R. Höhn[4,5], A. Nag[2], E. Yonova-Doing[2], K. M. Williams[2], A. I. Iglesias[3], T. Zeller[6,7], A. Mirshahi[4,8], S. Nickels[4], P. Hysi[2], C. W. Klaver[3], C. J. Hammond[2], N. Pfeiffer[4], C. M. van Duijn[3], A. C. Viswanathan[1,9].* 1) Department of Genetics, Institute of Ophthalmology, UCL, London, United Kingdom; 2) Twins Research & Genetic Epidemiology, King's College London, London, United Kingdom; 3) Department of Epidemiology and Clinical Genetics, Erasmus University Medical Centre, Rotterdam, Netherlands; 4) Department of Ophthalmology, University Medical Center Mainz, Germany; 5) Department of Ophthalmology, Inselspital, Bern, Switzerland; 6) Clinic of General and Interventional Cardiology, University Heart Center Hamburg, Germany; 7) German Center for Cardiovascular Research (DZHK), Partner Site Hamburg/Kiel/Lübeck, Germany; 8) Dardenne Eye Hospital, Bonn-Bad Godesberg, Germany; 9) NIHR Biomedical Research Centre for Ophthalmology at Moorfields Eye Hospital NHS Foundation Trust, London, United Kingdom.

The retinal nerve fibre layer (RNFL) thickness is a complex trait used in the diagnosis of glaucoma and the assessment of its progression and stability. It has also been suggested as a possible biomarker for neurodegenerative conditions such as Alzheimer's disease (AD). Despite the known high heritability of RNFL thickness, the genes affecting its variation have not been extensively investigated. The purpose of this study was to investigate genetic factors underlying RNFL thickness and whether genes for AD also play a role in RNFL thinning in the general population. We conducted a meta-analysis of genome-wide association studies imputed to HapMap2 in 10,502 individuals of European ancestry to identify genes associated with RNFL thickness variation. We identified 5 novel genetic regions associated with RNFL thickness: on chromosome 1 nearby *C1orf100* and *ZBTB18/ZNF238* genes (rs7554059 p=1. 17E-08 and rs6429459 p=1. 35E-08) and *BRINP3* (rs12087686 p=1. 92E-08), on chromosome 14 near to *DLK1* transcript (rs11625568 p= 4. 72E-08), on chromosome 10 within *CACNB2* gene (rs1277751 p= 5. 04E-08 and rs1277752 p=5. 54E-08) and on chromosome 17 within *PITPNM3* gene (rs1966778 p=6. 63E-08). We next created a multi-SNP genotypic risk score to test the association of AD risk alleles on RNFL thickness using meta-analysis data. We used 16 independent SNPs from ADGC meta-analysis for AD that also showed nominal significance in the RNFL meta-analysis. This risk score was associated with the decrease of RNFL thickness (beta=-0. 0047, se=0. 001, p=0. 005). To the best of our knowledge this is the most extensive study investigating RNFL thickness genetics in an European-ancestry population. These findings suggest that the genetic factors underlying RNFL thickness are mainly involved in neuronal development (*ZBTB18/ZNF238* and *BRINP3*) and calcium channels and cell adhesion (*CACNAB2* and *PITPNM3*). The study also shows that known Alzheimer's genes, such as *APOE*, *PVRL2*, *PICALM* and *CLU*, might also play a role in RNFL thinning. Our findings suggest common genetic background of RNFL thickness and AD and thus supporting its role as a potential biomarker for neurodegenerative diseases.

905W

Discovery of genetic associations underlying metabolically healthy obesity (MHO) and metabolically obese normal-weight (MONW). *N. Y. Yang[1,2], U. M. Schick[1,2], C. Schurmann[1,2], T. O. Kilpeläinen[3], R. J. F. Loos[1,2,4].* 1) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA; 2) The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA; 3) The Novo Nordisk Foundation Center for Basic Metabolic Research, Section of Metabolic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; 4) The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA.

Background: Obesity is associated with increased risk of metabolic complications including cardiovascular disease and type 2 diabetes. However, 17-45% of obese individuals remain metabolically healthy, also called the metabolically healthy obese (MHO). Furthermore 7-30% of normal weight individuals demonstrate metabolic complications typically seen in the obese, also called the metabolically obese normal weight (MONW). The physiological mechanisms that underlie MHO and MONW are poorly understood. While many known adiposity loci show cross-phenotype (CP) associations with metabolic complications that are concordant with the phenotypic correlations, some loci display discordant associations (i. e. adiposity increasing alleles associate with reduced risk of complication). In this study, we performed a genome-wide search to systematically identify loci with discordant CP associations between adiposity and cardiometabolic traits, which may underlie MHO and MONW. **Methods:** We analyzed summary statistics from nine publicly available genome-wide datasets with ~2,000,000 SNPs using CPASSOC, a CP association method that integrates summary statistics from multiple correlated traits. We evaluated all pairwise combinations of adiposity measures (BF%, BMI, WHRadjBMI) and cardiometabolic phenotypes (fasting glucose, insulin, HDL- and LDL-cholesterol, triglycerides, systolic blood pressure) to identify discordant CP associations between traits. **Results:** We identified 239 SNPs that clustered in 65 loci showing significant (P<5E-8) discordant CP associations for pairs of adiposity and cardiometabolic phenotypes. We identified known (e. g. *IRS1*, *PPARG, GCKR*) and novel (e. g. *TRIB1, SLC22A2, CCDC92)* loci that showed association with a wide range of phenotypes, whereas others demonstrated very specific association signatures (e. g. *ABO* with LDL-cholesterol/WHRadjBMI). SNPs from 38 of the loci are *cis*-eQTLs for a proximal gene (FDR<0. 05) in whole blood, adipose, or muscle. Prioritization with DEPICT identified enrichment of gene sets related to lipid localization (GO:0010876, P=4. 6E-7) and gluconeogenesis (GO:006094, p=3. 28E-6), among other metabolic pathways. **Conclusion:** By integrating data across multiple traits, we identified novel loci with distinct CP association signatures that provide more refined insights into biological pathways that may underlie MHO and MONW. These signatures may help pinpoint the causal gene within each locus to prioritize for further functional follow-up.

906T

Reanalysis of 15 Genome-wide Association Data Sets Revealed Novel Inflammatory Bowel Disease Loci. *Y. Zhang[1], L. Tian[2], P. Sleiman[3], H. Hakonarson[2,3,4].* 1) Department of Statistics, Pennsylvania State Univ, State College, PA; 2) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA.

Crohn's disease and Ulcerative Colitis are the two principal types of Inflammatory Bowel Disease (IBD) that are common in people of European origins. Genome-wide association studies and more recently imputation and meta-analyses have achieved multiple milestones in identifying the genetic architecture of IBD. The massive collaborative efforts of the International IBD Genetics Consortium (IIBDGC) has combined more than 75000 cases and controls from 15 genome-wide association studies (GWAS) and identified 71 new IBD loci, making a total of 163 IBD susceptibility loci reported to date. Here, we reanalyzed the Immunochip IBD data sets from IIBDGC using a novel Bayesian method, called BEAM3, that combines single and group based SNP testing into an unified framework. Our BEAM3 method removes indirect associations caused by the linkage disequilibrium among genetic variants, thereby producing much increased resolution at the disease loci. We identified 35 novel IBD loci reaching genome-wide significance that were not included in the current list of 163 IBD loci (permutation p-value <2e-7, corresponding to Bonferroni adjusted p-value <0. 025). Using two validation data sets independent of the Immunochip samples, we were able to replicate 8 new loci in both validation data sets and an additional 8 new loci in one validation data set. Many of the replicated new IBD loci implicated functionally interesting candidate genes, including *DNMT3A, THADA, UBE2E3*, *CCR2*/*CCR5*, *OSMR*, *CIT* and *CCL20* (locus-wise replication p-values are between 1e-4 and 1e-2). Our reanalysis of the currently largest IBD data sets added new insights into the IBD genetics and also highlighted the importance of advanced computational methods in disease association studies.

907F

Copy Number Variations in *CTNNA3* and *RBFOX1* Associate with Pediatric Food Allergy. *M. Bakay[1], J. Li[1], I. Fung[2], J. T. Glessner[1], R. Pandey[1], Z. Wei[5], F. D. Mentch[1], R. Pellegrino[1], T. Wang[1], C. Kim[1], C. Hou[1], F. Wang[1], R. M. Chiavacci[1], K. A. Thomas[2], J. M. Spergel[2,4,6], H. Hakonarson[1,3,4], P. Sleiman[1,3,4].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Allergy and Immunology, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 3) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; 4) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; 5) Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 08540, USA; 6) Center for Pediatric Eosinophilic Disorders, The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA.

Food allergy has become a significant public health concern, which is estimated to have a prevalence of 8% among children in the United States. Several susceptibility loci for food allergy have been examined through candidate gene studies, but there have been no studies systematically investigating copy number variations (CNV) which may contribute to the etiology of food allergy. To identify novel genetic factors for food allergy, we performed genome-wide SNP genotyping of pediatric food allergy cohorts and conducted CNV calling based on SNP intensity data. We then conducted CNV association analysis for a discovery cohort of 357 cases and 3,980 controls, and did a replication study on 167 cases and 1,573 controls. We found significant association between CNVs in gene αt-catenin (*CTNNA3*) and food allergy in both the discovery cohort and the replication cohort. CNVs in another gene RNA Binding Protein, Fox-1 Homolog1 (*RBFOX1*) showed significant association specifically in the Caucasian cohorts. We also validated these CNVs by quantitative polymerase chain reaction. In addition, we showed that knockdown of *CTNNA3* led to increased degranulation in mononuclear cells, comparing to control. In summary, we have identified two CNV loci that are significantly associated with food allergy among children and further functional study suggests that CTNNA3 may be involved in sensitization to allergen.

908W

Burden Analysis of Copy Number Variation in Gallstone Disease. *E. E. Perez-Palma[1,2], B. I. Bustos[1], D. Lal[2], L. Azócar[3], C. Moraga[4], J. Hampe[5], P. Nürnberg[2], R. A. Gutiérrez[4,6], J. F. Miquel[3,6], G. V. De Ferrari[1,6].* 1) Centro de Investigaciones Biomédicas, Facultad de Ciencias Biológicas, Universidad Andres Bello, Chile; 2) Cologne Center for Genomics, University of Cologne, Germany; 3) Depto. de Gastroenterología, Facultad de Medicina, Pontificia Universidad Católica de Chile; 4) Depto. de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile; 5) Medical Department I, University Hospital Dresden, TU Dresden, Germany; 6) FONDAP Center for Genome Regulation (CGR), Chile.

Cholesterol gallstones disease (GSD) constitutes a major burden to public health systems due to its high prevalence among occidental countries. Current evidence supports the hypothesis that GSD has a significant genetic component. However, to date the only common genetic risk factors found to be associated are two SNVs, one in *ABCG8* and the other in *UGT1A1* (only in men), discovered through genome-wide association studies performed in Germany and replicated consistently in a Chilean and other cohorts. Together these two variants explain only a minor part of the estimated heritability of GSD, suggesting the existence of additional genetic variants that remain to be discovered. In the present study we explored the contribution to GSD susceptibility that copy number variants (CNVs) might have by examining large case control cohorts composed of 1,076 Chilean and 1,957 German samples, in total 1,574 cases and 1,459 controls, genotyped on the Affymetrix GeneTitan platform using the Axiom LAT 1 and GW Hu SNP World arrays, respectively. The Affymetrix Axiom CNV tool was used to calculate allele intensity ratios and B allele frequencies. These data were then put into the Nexus Copy Number Software to call CNVs using the SNP-FASST2 Segmentation algorithm. We considered CNV calls with at least 50 probes involved and larger than 100kb. Further, we avoided calls overlapping telomeric and centromeric regions. A total of 1,414 events was detected (633 in Chileans; 781 in Germans). Burden analyses revealed enrichment of patients carrying at least one duplication covering genes expressed in disease related tissues, particularly in small intestine: suggestive for the Chilean cohort (p=4. 91x10-2) and statistically significant in the German dataset (p=4. 91x10-4). Analyzing the whole dataset together confirmed the finding (p=2. 10x10-4). Breaking the dataset down for gender effects revealed that the observation was almost exclusively male-specific. At this level the overall amount of CNVs detected was higher in patients than in controls (p=2. 34x10-4) and primarily affected genes expressed in the small intestine (7. 96x10-5) with significant participation of genes involved in lipid metabolism according to gene ontology (GO:0006629, p=1. 6x10-3). Our preliminary results provide novel insights in the genetic architecture of GSD and suggest that CNVs may explain a part of the missing heritability in GSD. **Funding:** FONDECYT #1130303-JFM, #1140353-GVD, FONDAP #1509000.

909T

The Contribution of Glycemic and Non-Glycemic Genetic Variants Identified Through a Trans-Ethnic Discovery Meta-Analysis of A1c to Type 2 Diabetes Prediction and Reclassification. *E. Wheeler on behalf of MAGIC, AAGILE, AGEN and TAICHI.* Wellcome Trust Sanger Institute, Cambridge, United Kingdom.

Glycated hemoglobin (A1c) is currently used as a diagnostic test for type 2 diabetes (T2D). Ethnic-specific meta-analyses in European and East Asian populations have identified 17 loci associated with A1c thought to act on A1c through glycemic or erythrocyte biology. We investigated the genetic determinants of A1c in up to 159,940 non-diabetic participants from cohorts of European, African American, East Asian, and South Asian ancestry with genome-wide or Metabochip data. Trans-ethnic meta-analysis in MANTRA and ethnic specific meta-analyses in METAL identified 37 novel loci associated with A1c: 35 loci in our trans-ethnic MANTRA analysis and 2 in the European-only meta-analysis. We also confirmed the 17 known loci and found independent secondary signals at 5 of these loci. Conditional analyses and cross trait look-ups classified the 60 A1c loci as glycemic (eg. *CDKAL1*, log10BF = 16. 52), erythrocytic (*G6PD*, log10BF = 50. 60), or unknown (eg. *ATAD2B*, log10BF=13. 30). We tested the hypothesis that glycemic A1c single nucleotide polymorphisms (SNPs) predicted incident T2D differently from erythrocytic SNPs in both the Framingham Offspring Study (N=1,705) and InterAct (N=16,687). T2D risk was modeled using 2 separate genetic scores calculated from the weighted sum of A1c -raising alleles at 19 glycemic SNPs (GS-G) and at 21 erythrocytic SNPs (GS-E). In age and sex adjusted models, higher GS-G was associated with increased risk of incident T2D (HR=1. 05 [1. 04-1. 06], P=5. 45x10-22 in InterAct). In contrast, higher GS-E was not associated with increased risk of incident T2D (P > 0. 1). However, higher GS-E decreased risk of incident T2D after adjustment for baseline A1c (binary: <5. 7 vs. ≥5. 7%) (HR=0. 96 [0. 94-0. 98], P = 4x10-7 in InterAct). Similarly, in a Net Reclassification Improvement (NRI) analysis we found that adjustment for GS-E over baseline A1c (binary: <5. 7 vs. ≥5. 7%) reduced the number of false positive cases of T2D in InterAct by 13% (P < 0. 0001), consistent with the idea that nonglycemic genetic factors contribute to the misclassification of T2D risk based on A1c in some individuals. In conclusion, the novel loci identified expand our understanding of genetic and biologic determinants of A1c variation. Examining whether non-glycemic variants partly explain inter-ethnic differences in A1c levels or their ability to accurately predict and or diagnose T2D in multi-ethnic populations may have relevance in clinical practice.

910F

Genome-wide association study for Gallstone disease in Latin Chilean population. *B. I. Bustos[1], E. Pérez-Palma[1], S. Buch[2], L. Azócar[3], C. Moraga[4], M. Toilat[5], J. Hampe[2], P. Nürnberg[5], R. A. Gutiérrez[4,6], G. V. De Ferrari[1,6], J. F. Miquel[3,6].* 1) Centro de Investigaciones Biomédicas, Facultad de Ciencias Biológicas, Universidad Andrés Bello, Santiago, Chile; 2) Medical Department I, University Hospital Dresden, TU Dresden, Dresden, Germany; 3) Cologne Center for Genomics, University of Cologne, Köln, Germany; 4) Depto. de Gastroenterología, Facultad de Medicina, Pontificia Universidad Católica de Chile, Santiago, Chile; 5) Depto. Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago, Chile; 6) FONDAP Center for Genome Regulation (CGR).

Cholesterol gallstones disease (GSD) is one of the most prevalent gastrointestinal disorders in western countries with a high genetic susceptibility. Only one genome-wide association study (GWAS) on GSD has been published in the past, reporting a lithogenic variant on the sterolin transporter ABCG8 (p. D19H). This signal was replicated in different populations worldwide, including Latin Chileans. However, it only explains a small portion of the population attributable risk. The Chilean population, with one of the highest prevalence of GSD in the world (36% and 17% in adult women and men, respectively), has been the target of only a few genetic studies. Here, we present results from the first GWAS performed on the Chilean population comprising 1,095 individuals (529 cases, 566 controls), aiming to uncover novel susceptibility risk factors for GSD. Subjects were recruited from Community medical Centers in Santiago city. Status of GSD (cases or controls) was diagnosed by abdominal ultrasonography. Genotyping was done using Affymetrix Axiom LAT 1 World array Plates. Genome-wide Imputation was performed with IMPUTE2 using the 1000 Genomes Project Phase 3 reference panel. Association analyses were done using SNPTEST. After QC and imputation, we analyzed ~9. 5 Million SNVs observing a number of clear signals surpassing the suggestive genome-wide significance threshold (<1x10-5). We confirm a signal in chromosome 2 inside the previously known ABCG5/8 locus (p= 7. 04x10-6). Novel signals were found on chromosome 7 (p= 5. 93 x10-7) and 14 (p= 5. 31x10-6), in regions that encompasses ELMO1 (gene involved in cell motility) and TRAF3 (involved in immune/inflammatory response), respectively. Since women compared with men are 3 times more likely to develop GSD, we performed a women specific analysis (489 cases and 525 controls). Results show an additional signal in chromosome 7 inside the GPR30 gene (p= 5. 78x10-6), a G-protein coupled estrogen receptor. Interestingly, knock-down mice for this gene have shown increased GSD susceptibility under a lithogenic diet. Our results show suggestive evidence of novel genetic variants associated with GSD providing new insights into the genetic susceptibility of this prevalent disease in high a risk Latin population. Replication and meta-analysis efforts with other independent cohorts are currently underway in order to have a more comprehensive view of the current results. **Funding:** FONDECYT #1130303-JFM, #1140353-GVD, FONDAP #1509000.

911W

AS3MT, Other Genes and Arsenic Metabolism Biomarkers in American Indians: the Strong Heart Family Study. *P. Balakrishnan[1-3], D. Vaidya[2,4], N. Franceschini[5], V. S. Voruganti[6], M. Gribble[7], K. Haack[8], S. Laston[9], J. Umans[10,11], K. Francesconi[12], W. Goessler[12], K. North[5], E. Lee[13], L. Best[14], J. MacCluer[8], J. Kent Jr[8], S. Cole[8], A. Navas-Acien[1-3,15].* 1) Department of Environmental Health Sciences, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD; 2) Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD; 3) The Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins Medical Institutions, Baltimore, MD; 4) Clinical and Translational Research, Johns Hopkins School of Medicine, Baltimore, MD; 5) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 6) Department of Nutrition and UNC Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis, NC; 7) Division of Biostatistics, Department of Preventative Medicine, University of Southern California, Los Angeles, CA; 8) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 9) South Texas Diabetes and Obesity Institute, University of Texas Health Science Center, San Antonio, TX; 10) MedStar Health Research Institute, Hyattsville, MD; 11) Georgetown and Howard Universities Center for Clinical and Translational Science, Washington, DC; 12) Institute of Chemistry-Analytical Chemistry, University of Graz, Austria; 13) Center for American Indian Health Research, College of Public Health, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 14) Missouri Breaks Industries Research, Inc. , Timber Lake, SD; 15) Department of Oncology, Johns Hopkins School of Medicine, Baltimore, MD.

Inorganic arsenic (iAs) is a naturally occurring toxicant and carcinogen in groundwater and food. Some rural and tribal U. S. communities are disproportionately exposed to iAs in drinking water. Inter-individual variability in iAs metabolism influences arsenic (As) toxicity. The genetic determinants of As metabolism are not well understood, especially among American Indians. Our objective was to evaluate genetic determinants of As metabolism in the Strong Heart Family Study, a large family-based study in American Indian communities from Arizona, Oklahoma and North/South Dakota. We evaluated 2,431 individuals from extended pedigrees with As and genotype data. Percent dimethylarsinate (DMA%), percent methylarsonate (MMA%), and iAs% in urine were calculated as the relative proportion of each As species to their sum. Percent As species were logit transformed. We created principal components (PC) of As species. PC1 reflected higher iAs%/MMA% and lower DMA%. PC2 reflected higher iAs% and lower MMA% at similar DMA%. Genotyping used the Illumina MetaboChip and single nucleotide polymorphism (SNPs) from candidate cardiometabolic genes. The alpha threshold was corrected for multiple testing using Bonferroni (MetaboChip 3. 07e-7; candidate SNPs 9. 33e-5). Additive SNP effect and minor allele frequency (MAF) cutoff of 0. 01 was used. Models were adjusted for age, sex, total As levels, and principal components for population stratification. Models were analyzed by study region using SOLAR and meta-analyzed using METAL. We also conducted conditional analysis and Bayesian quantitative trait nucleotide analysis with index SNPs. The index SNP for percent As species (rs12768205, MAF 0. 27) was in AS3MT at 10q24 (logit DMA% Z-score 10. 07, logit MMA% -7. 99, logit iAs% -5. 92). Index SNPs rs3740394 for PC1 and rs7098825 for PC2 were located at 10q24 in a region of strong LD (r2 > 0. 8). Among the candidate SNPs, a non-synonymous coding SNP in AS3MT (rs11191439, MAF 0. 18) was significantly associated with PC1 (1. 12e-36), PC2 (3. 96e-8), logit DMA% (2. 89e-19) and logit MMA% (2. 60e-7). A SNP upstream of USMG5 at 10q24 (rs7911488, MAF 0. 25) was significantly associated with PC1 (3. 57e-31), logit DMA% (7. 04e-18), logit MMA% (6. 54e-8) and logit iAs% (1. 04e-6). Our results suggest the importance of AS3MT in As metabolism and potentially in As toxicity. Significant associations among other cardiometabolic genes in 10q24 may also be involved in As metabolism, especially between iAs% and MMA%.

912T

Mega-analysis versus Meta-Analysis: On the gain of jointly imputing and analyzing 33,976 subjects compared to meta-analyzing study-specific aggregated genome-wide association statistics on the example of Age-related Macular Degeneration. *M. M. Gorski for the International AMD Genomics Consortium (IAMDGC).* Mathias M. Gorski, Department of Genetic Epidemiology, Institute of Epidemiology and Preventive Medicine, University of Regensburg, D-93053 Regensburg, Germany.

Background: Current meta-analyses of genome-wide association studies impute and analyze data study-specifically and combine study-specific statistics. We have evaluated the gain in substituting this common practice with imputing and analyzing all subjects jointly (mega-analysis) by examining effect sizes, p-values and imputation quality. We use age-related macular degeneration (AMD), a leading cause of blindness, as model disease as it exhibits a broad range of effect sizes for common and rare variants. **Methods:** For our methodological evaluation, we use a real data set that allows us to conduct a mega-analysis and mimic a meta-analysis to avoid artificial data sets from simulations. We used quality controlled genotypes from a custom-modified chip array of 33,976 subjects from 26 study centers from the International AMD Genomics Consortium. The genotypes were phased (shapeit. v2. r727. linux. x64) and imputed (minimac-omp_2013_7_17) with the Giant All Phase I v3 reference panel jointly (mega-imputation) and per study (meta-imputation). We then analyzed both data sets by the Firth-corrected logistic (epacts v3. 2. 6) regression jointly (mega-analysis) and per study (meta-analysis). Results are exemplified on chromosome 22, including the known genetic AMD loci *SYN3/TIMP3* and *SLC16A8.* **Results:** The imputation quality in the mega-imputation was higher compared to the meta-imputation in all and well imputed (RSQ>0. 4) variants (median difference of RSQ = 0. 035 and 0. 0016, respectively). We were also interested in the impact of these differences on association analysis results conducting a mega-analysis (with mega-imputed genotypes) compared to a meta-analysis (by study, with meta-imputed genotypes). We found, that the p-values in the meta-analysis were approximately two magnitudes higher in *SNY3/ TIMP3* (1. 17x10-22 vs. 9. 77x10-25) and one magnitude higher in *SLC16A8* (6. 86x10-10 vs. 6. 04x10-11). **Conclusion:** Our evaluation shows that imputation quality can be slightly, but not substantially improved by a mega-imputation compared to a meta-imputation. We were also able to show that the results of the association analysis would be slightly different with the p-values decreasing by one to two orders of magnitude when conducting a meta-imputation/analysis compared to a mega-imputation/analysis. The possibility of parallelization in the meta-analysis was a computational advantage compared to the mega-analysis.

913F

Epigenetic susceptibility and novel biomarkers for severe influenza A virus infection. *A. Wennerström[1,2], N. Pervjakova[1,3,4], I. Julkunen[1], P. Jousilahti[1], M. Perola[1,2,3,4].* 1) National Institute for Health and Welfare, University of Helsinki, Helsinki FI-00271, Finland; 2) The Institute for Molecular Medicine, University of Helsinki (FIMM);; 3) Department of Biotechnology, Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia; 4) Estonian Genome Center, University of Tartu, Tartu 51010, Estonia.

Influenza viruses represent one of the top challenges for public health. Highly pathogenic influenza infections are substantially important factors of morbidity and mortality, especially in an event of pandemic. Although influenza virus is studied actively, a comprehensive characterization of genetic and epigenetic variants associated with increased infection risk is still missing. Although influenza virus is studied actively, a comprehensive characterization of genetic and epigenetic variants associated with increased infection risk is still missing. Here, we aimed to discover novel associations with influenza virus in Finnish population. The phenotype was defined as a case when a person infected by influenza was admitted to an intensive care unit. We performed a GWAS analysis using 56 cases and 142 controls, and epigenome-wide association study (EWAS) for 56 cases and 520 controls. Genotype data were imputed to the 1000 Genome Phase 1 release 3 reference panel (March 2012). For the EWAS analysis DNA methylation was determined by Illumina Infinium 450K Human Methylation array. SWAN normalization was applied to the data. Probes with detectable methylation level < 5% of samples (detection p-value < 0. 01) were excluded. Also, CpG sites on X and Y chromosomes were discarded, which yielded to 485,513 probes for further analysis. Following discovery analyses, we sought follow-up for SNP and CpG sites reaching 5x10-8. The GWAS approach identified two loci in chromosome 5; one near genes *CCNG1*, *NUDCD2*, *HMMR1* and *MAT2B* and another near FBXL17 gene (best SNP: rs143267366, p = 2. 46x10(-8)). Epigenome-wide association analysis revealed 98 CpG variants. The most significant results were found for *WDR51B* (p= 7. 46x10-14), *MAT2A* (8. 90x10-13) and *S100A12* (1. 07x10-11). However, these results should be replicated in an independent material. The current influenza global trend can be characterized by two factors – increased variety of animal co-circulating viruses and exchanging genetic material that give a rise to novel strains that affect the ability of protection through vaccination. Our results provide new insights into the genetic background linked to viral infection, and may enable new strategies for prevention of influenza A virus infection and other consequences of this disease.

914W

Functional linear models extensions uncover pleiotropic effects of chronic pain phenotypes. *D. V. Zaykin[1], L. Qing[2], G. D. Slade[3], R. Dubner[4], R. B. Fillingim[5], J. D. Greenspan[4], R. Ohrbach[6], W. Maixner[3], L. B. Diatchenko[7,8], O. A. Vsevolozhskaya[2].* 1) Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA; 2) Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, MI, USA; 3) Regional Center for Neurosensory Disorders, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA; 4) Departments of Oral and Maxillofacial Surgery and Neural and Pain Sciences, and Brotman Facial Pain Center, University of Maryland School of Dentistry, Baltimore, MD, USA; 5) Department of Community Dentistry and Behavioral Science, University of Florida, College of Dentistry, and Pain Research and Intervention Center of Excellence, Gainesville, FL, USA; 6) Department of Oral Diagnostic Sciences, University at Buffalo, Buffalo, NY, USA; 7) The Alan Edwards Centre for Research on Pain, McGill University, Montreal, Quebec, Canada; 8) Department of Anesthesia, Faculty of Dentistry, McGill University, Montreal, Quebec, Canada.

Growing scientific evidence suggests that intricate interactions of genetic risk factors with environmental exposures play a major role in the development of chronic pain conditions. In studies of relative contribution of an individual's genetic composition to the perception of pain, the general characteristics of pain sensitivity are typically measured by a wide range of different, yet possibly etiologically related pain phenotypes. Testing each of these pain-perception traits individually is subject to problems of multiple testing and low statistical power. Furthermore, pain-related traits may share common etiology and comprise binary, categorical, and quantitative measurements. In the current study, we propose a novel statistical approach for simultaneous testing of multiple correlated phenotypes, including quantitative binary, categorical or a combination thereof, with the flexibility of adjusting for other covariates. We illustrate our approach with the association analysis between groups of multiple pleiotropic phenotypes and genetic variants of the *P2RX7* gene from the Orofacial Pain: Prospective Evaluation and Risk Assessment (OPPERA) study.

915T

A trans-ethnic meta-analysis of genome wide association studies reveals new loci associated with childhood obesity. *J. P. Bradfield[1], J. F. Felix[4,5,6], V. W. V. Jaddoe[4,5,6], S. F. A. Grant[1,2,3], Early Growth Genetics Consortium.* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Divisions of Human Genetics and Endocrinology, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA; 4) The Generation R Study Group, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands; 5) Department of Epidemiology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands; 6) Department of Paediatrics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, the Netherlands.

Childhood obesity has more than doubled in the last 30 years. Children who are obese are more likely to be obese as adults and are more at risk for several adult health problems such as type 2 diabetes and heart disease. To search for genetic variants that are associated with childhood obesity, we performed a trans-ethnic meta-analysis of 22 genome wide association studies with European, African, South American and East Asian ancestry. The meta-analysis discovery stage comprises 10,668 cases, defined as greater than the 95th percentile of body mass index (BMI) between 2 and 18 years old, and 12,647 controls, defined as less than the 50th percentile all throughout childhood. Using the MANTRA algorithm, we meta-analyzed approximately 19 million SNPs and found fourteen loci that attained genome wide significance ($P < 5x10-8$). We have identified 4 novel loci (*PHLDA1, METTL15, RBMS3* and *COTL1*) in addition to 10 loci (*FTO, TMEM18, FAIM2, MC4R, SEC16B, ADCY3, TNNI3K, GNPDA2, FHIT* and *TFAP2B*) that had been previously discovered. More than 100 loci were found to be suggestive of association ($P < 5x10-6$), seven of which (*HOXB3, ADCY9, NEGR1, PTPB2, OLFM4, LINGO2* and *SULT1A2*) had previously been found to be associated with BMI or obesity. A sex specific association analysis was also performed. This revealed three new genome wide significant loci (*RBM6, STARD4* and *LINC01524*) among the females. All of these loci will be carried over into our replication effort, further narrowing down the list of potential variants. In summary, we have discovered potential pediatric obesity loci that are being validated by large replication efforts.

916F

Genome-wide association study identifies *EFEMP1* as a new candidate biliary atresia susceptibility gene. *Y. Chen[1,2], C. M. Grochowski[2], M. Gilbert[2], R. Rajagopalan[2], J. Bailey-Wilson[3], D. Stambolian[1], H. C. Lin[1,2], K. M. Loomes[1,2], N. B. Spinner[1,2], M. Devoto[1,2].* 1) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) The Children's Hospital of Philadelphia, Philadelphia, PA; 3) National Human Genome Research Institute, National Institutes of Health, Baltimore, MD.

Biliary atresia (BA) is a rare liver disease presenting within the first months of life. It is characterized by obliteration of the extrahepatic biliary tree in a progressive, necroinflammatory manner, leading to cholestasis, fibrosis, cirrhosis, and chronic liver damage, and accounts for 50% of pediatric liver transplantations. The etiology of BA is not understood, and environmental, inflammatory, and genetic risk factors have been proposed. A GWAS in Chinese patients identified a signal on 10q25 upstream of *ADD3* and *XPNPEP1*, which we replicated in European-Americans. In this study, we performed a GWAS in 450 European-American non-syndromic BA patients collected through the Childhood Liver Disease Research Network (ChiLDReN) and 1981 genetically matched controls from the Age-Related Eye Disease Study (AREDS). Genotyping was done with the Illumina Omni2. 5 array. Adjusted logistic regression was carried out to test SNP association with BA. The most significant SNP was rs10865291 on 2p16 (p = 2. 7x10-7; OR=1. 6; 95% CI (1. 3 -1. 9)), in the fifth intron of *EFEMP1*. *EFEMP1* encodes the EGF-containing fibulin-like extracellular matrix protein Fibulin-3 which has been implicated in tissue regeneration and organogenesis. It has also been reported to activate Notch signaling *in vitro*, although with less efficiency than *JAG1.* ddPCR performed on cDNA from BA liver specimens collected at time of liver transplant, from healthy control liver, and from disease control liver (n = 5 each) revealed that *EFEMP1* expression was increased in BA liver samples by 9-fold compared to healthy control liver samples (P-value < 0. 002), although no significant expression difference was detected between the BA and disease control groups. Immunohistochemistry showed that Fibulin-3 was specifically expressed in tubular structures around portal tracts in BA liver. Follow-up co-staining with CK-19 and alpha smooth muscle actin will be performed to determine cell type-specific expression. Ongoing studies include replication of the GWAS in additional BA cohorts of European and African-American descent. In conclusion, we identified a locus on 2p16 within the *EFEMP1* gene as the second candidate region associated with susceptibility to BA using the GWAS approach. *EFEMP1* is highly upregulated in BA and disease control liver, and its protein Fibulin-3 is expressed in the portal regions. The identification of this new candidate susceptibility gene will facilitate characterizing the genetic basis of BA.

917W

A genome-wide investigation of food addiction. *M. C Cornelis[1], A. Flint[2], A. E. Field[2], P. Kraft[2], J. Han[3], R. M. van Dam[4], E. B. Rimm[2].* 1) Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 2) Harvard School of Public Health, Boston, MA; 3) Richard M. Fairbanks School of Public Health, Simon Cancer Center, Indiana University, Indianapolis, IN; 4) Saw Swee Hock School of Public Health and Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore.

Evidence of parallels between drug addiction and eating behavior continues to accumulate. Genetic studies of addictive substances have yielded a number of susceptibly loci that point to common higher-order genetic pathways underlying addiction. To elucidate the relationship between drug addiction and eating behaviors we conducted a genome-wide association study (GWAS) of food addiction, determined by the modified Yale Food Addiction Scale (mYFAS), among 9,314 U. S. women of European ancestry and examined results for enrichment of single-nucleotide polymorphisms (SNPs)(n=44), genes (n=238) and pathways (n=11) implicated in drug addiction. Two loci met GW-significance (P<2. 5×10-8) mapping to 17q21. 31 and 11q13. 4 that harbor genes with no obvious roles in eating behavior. GW results were significantly enriched for gene members of the MAPK signaling pathway (P=0. 02). No candidate SNP or gene for drug addiction was significantly associated with food addiction. A SNP previously linked to illicit drug use, and the candidate genes *LOC100130673*, *HOMER1*, *ZHX2*, *DRD2*, and *SURF6,* were nominally associated with food addiction. A global exploratory pathway analysis yielded significant enrichment for interleukin signaling genes (FDR=0. 003). In the first GWAS of mYFAS, we identified suggestive loci worthy of further follow-up but provide limited support for shared genetic underpinnings of food addiction and drug addiction.

918T

Novel rare and low frequency variants associated with lipoprotein subclasses and triglyceride measures. *J. P. Davis[1], J. R. Huyghe[2], A. U. Jackson[2], H. M. Stringham[2], T. M. Teslovich[2], R. P. Welch[2], C. Fuchsberger[2], A. E. Locke[2], N. Narisu[3], P. S. Chines[3], A. J. Kangas[4], P. Soininen[4,5], M. Ala-Korpela[4,5,6,7], J. Kuusisto[8], F. S. Collins[4], M. Laakso[8], M. Boehnke[2], K. L. Mohlke[1].* 1) Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 3) National Human Genome Research Institute, NIH, Bethesda, MD, USA; 4) Computational Medicine, Institute of Health Sciences, University of Oulu, 90014, Oulu, Finland; 5) Nuclear Magnetic Resonance Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, FIN-70211, Kuopio, Finland; 6) Oulu University Hospital, FIN-90220, Oulu, Finland; 7) Computational Medicine, School of Social and Community Medicine and the Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol, BS8 2BN, UK; 8) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, 70210 Kuopio, Finland.

Genome-wide association studies have identified common genetic variants associated with size and content of plasma lipoproteins and triglycerides, but the contributing role of rare and low frequency coding and noncoding variants is not fully understood. In the METabolic Syndrome In Men (METSIM) study, we examined high-throughput proton NMR-based measurements of 72 lipoprotein subclasses and triglyceride measures in 8,380 men aged 45-74 from Kuopio, Finland. Samples were genotyped with the Illumina Omni Express and ExomeChip Arrays (733,000 total markers) and then imputed with a 19 million variant European reference panel. We performed single-variant and gene-based tests of association with the 72 traits, adjusted for age and lipid-lowering medication use. We identified novel genome-wide associations with one or more trait at three loci, labeled by a nearby gene: (i) *CWC22* (MAF 0. 2%, *P*=6. 8x10-9) associated with phospholipids in medium HDL, (ii) *ANKRD17* (MAF 1. 7%, *P*=3. 3x10-22) associated with concentration of small LDL particles, and (iii) *HECW1* (MAF 0. 5%, *P*=6. 7x10-10) associated with total cholesterol in medium VLDL. In an additional analysis conditioned on known lipid and lipoprotein signals, nine variants located within 1 Mb of a known lipid locus remained significantly (*P*<5x10-8) associated with one or more trait. All of the novel secondary signals were noncoding except a rare missense variant in *LCAT* associated with HDL cholesterol (MAF 0. 5%, *P*=7. 6x10-10). Gene-based tests of association support a functional role for missense variants for *LIPC* (trait: triglycerides in very large HDL, *P*=9. 8x10-11) and *LIPG* (trait: phospholipids in medium HDL, *P*=2. 2x10-16). Genotyping to validate rare imputed variants is underway. At the four novel loci, functional annotation revealed multiple variants in LD (*r2*>0. 7) with lead variants overlapping epigenetic marks of regulatory activity in liver, blood, and adipose, providing several candidates to test in assays of gene function or transcriptional activity. These rare and low frequency coding and noncoding variants associated with lipoprotein subclasses and triglyceride measures may provide further insight to the molecular basis of dyslipidemia and the etiology of metabolic disorders.

919F

Genome-Wide Meta-Analysis on the Sense of Smell among US Older Adults. *J. Dong[1], J. Yang[2], G. Tranah[3], N. Franceschini[4], N. Parimi[3], G. Alkorta-Aranburu[5], Z. Xu[1], A. Alonso[6], S. Cummings[3], M. Fornage[7], X. Huang[8], S. Kritchevsky[9], Y. Liu[10], S. London[1], L. Niu[11], R. Wilson[2], P. De Jager[12], L. Yu[2], A. Singleton[13], T. Harris[14], T. Mosley Jr. [15], J. Pinto[16], D. Bennett[2], H. Chen[1].* 1) Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina;; 2) Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois; 3) California Pacific Medical Center Research Institute, San Francisco, California; 4) Department of Epidemiology, University of North Carolina Gillings School of Global Public Health, Chapel Hill, North Carolina; 5) Department of Human Genetics, The University of Chicago, Chicago, Illinois; 6) Division of Epidemiology & Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota; 7) Institute of Molecular Medicine and Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas; 8) Pennsylvania State University-Milton S. Hershey Medical Center, Hershey, Pennsylvania; 9) Sticht Center on Aging, Wake Forest School of Medicine, Winston-Salem, North Carolina; 10) Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina; 11) Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina; 12) Program in Translational Neuro Psychiatric Genomics, Institute for the Neurosciences, Departments of Neurology and Psychiatry, Brigham and Women's Hospital; Harvard Medical School; Program in Medical and Population Genetics, Broad Institute, Boston, Massachusetts; 13) Laboratory of Neurogenetics, National Institute on Aging, Bethesda, Maryland; 14) Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, Bethesda, Maryland; 15) Division of Geriatrics, Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi; 16) Section of Otolaryngology-Head and Neck Surgery, Department of Surgery, The University of Chicago Medicine and Biological Sciences, Chicago, Illinois.

Olfactory dysfunction is common among older adults and affects their safety, nutrition, quality of life, and mortality. More importantly, the decreased sense of smell is an early symptom of neurodegenerative diseases such as Parkinson's disease (PD) and Alzheimer's disease (AD). However, the genetic determinants for the sense of smell have been poorly investigated. We here performed the first genome-wide meta-analysis on the sense of smell among 6,252 older US adults of European descent from the Atherosclerosis Risk in Communities study, the Health, Aging, and Body Composition study, and the Rush University Religious Orders Study and Memory and Aging Project. We identified 13 loci with suggestive evidence for an association with the sense of smell ($Pmeta < 1 \times 10\text{-}5$) independent of *ApoE* ϵ4 status, a genetic variation associated with poor sense of smell. Two SNPs at chromosome 17q21. 31 (rs199443 in *NSF*, $P = 3.02 \times 10\text{-}6$; and rs2732614 in *KIAA1267-LRRC37A*, $P = 6.65 \times 10\text{-}6$) exhibited *cis* effects on the expression of microtubule-associated protein tau (*MAPT*, 17q21. 31) in 447 frontal-cortex samples obtained postmortem and profiled by RNA-seq ($P < 1 \times 10\text{-}15$). Gene-based and pathway-enrichment analyses further implicated *MAPT* in the sense of smell. Similar results were obtained after excluding participants who reported a physician-diagnosed PD or use of PD medications. This meta-analysis provides the first evidence that *MAPT*, a major susceptibility gene for PD and possibly for AD, may also play a role in regulating the sense of smell in older adults.

920W

Genome-wide association analysis for chronic vein insufficiency identifies two genes as the first susceptibility loci. *D. Ellinghaus[1], A. Fiebig[1], E. Ellinghaus[1], F. Lai[1], P. Krusche[2], N. Frings[2], C. Schreiber[3], S. Nikolaus[3], C. Gieger[4], W. Lieb[5], P. Rosenstiel[1], S. Schreiber[1,3], A. Franke[1].* 1) Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel, Kiel, Schleswig Holstein, Germany; 2) Capio Mosel-Eifel-Clinic, Bad Bertrich, Germany; 3) Department of General Medicine, University Hospital Schleswig-Holstein (UKSH), Campus Kiel, Germany; 4) Institute of Genetic Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 5) PopGen Biobank, Section Epidemiology, Christian-Albrechts-University of Kiel, Germany.

Advanced chronic venous insufficiency (CVI) is a multifactorial condition and presents one of the most vascular abnormalities, especially in Northern and Western Europe, affecting about 3% of the population. Despite the frequency of CVI, the underlying etiology and pathophysiology are still poorly understood. According to the CEAP guidelines for chronic venous disease, patients with dilated subcutaneous veins of at least 3 mm diameter and without signs of edema or skin changes, are classified as C2. Patients with CVI can be classified as either C3 (edema), C4 (skin changes without ulceration), C5 (healed venous ulcer) or C6 (active venous ulcer). A genetic component has been proposed for many years, strongly indicated by reports on familial clustering and twin studies. The heritability of CVI is estimated about 17%, thereby suggesting genetic risk factors in the etiology of CVI. So far, no gene has been identified as a susceptibility factor for CVI. To further our understanding of the genetic etiology of CVI, we undertook the hitherto first genome-wide association study (GWAS) for this multifactorial condition comprising 9961 individuals in total (discovery and two replication panels). SNP genotype imputation was carried out using HapMap2 and HapMap3 reference haplotypes to predict missing autosomal genotypes. For downstream analyses, we included only SNPs that were imputed with high confidence (estimated r2 between imputed and true genotypes > 0. 8) and had a MAF >= 1%. In total, 1,935,136 quality-controlled autosomal imputed SNP markers were available for association analysis. The combined analysis of discovery and replication stages yielded genome-wide significant evidence (P<5×10-8) for two loci as well as one additional suggestive locus with P<5×10-7. From these analyses we are able to draw insight into the biological underpinnings of CVI.

921T

Trans-ethnic GWAS of pelvic organ prolapse among European American, African American and Hispanic post-menopausal women from the Women's Health Initiative. *A. Giri[1], J. Wu[2], R. Ward[3], K. Hartmann[1,3,4], A. Park[5], K. North[6], M. Graff[6], R. Wallace[7], G. Gareh[8], L. Qi[9], M. Sullivan[10], A. Reiner[11], D. Velez Edwards[1,3,12], T. Edwards[1,4,12].* 1) Institute of Medicine and Public Health, Vanderbilt Epidemiology Center, Vanderbilt University, Nashville, TN; 2) Department of Obstetrics and Gynecology, Center for Women's Health Research, University of North Carolina, Chapel Hill, NC; 3) Department of Obstetrics and Gynecology, Vanderbilt University Medical Center, Nashville, TN; 4) Department of Medicine, Vanderbilt University Medical Center, Nashville, TN;; 5) Department of Obstetrics and Gynecology and Urology, Georgetown University Medical Center, Washington, DC; 6) Department of Epidemiology, University of North Carolina, Chapel Hill NC; 7) Department of Epidemiology, University of Iowa, Iowa City, IA; 8) Department of Obstetrics and Gynecology, Loma Linda University, Loma Linda, CA; 9) Department of Public Health Sciences, University of California, Davis, CA; 10) Department of Obstetrics and Gynecology, University of Miami, Miami, FL; 11) Department of Epidemiology, University of Washington, Seattle, WA; 12) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN.

Up to 40% of post-menopausal women are affected by pelvic organ prolapse (POP) and 1 in 5 women will undergo surgery for POP/urinary incontinence in their lifetime. A growing body of evidence supports genetic predisposition for POP. Using validated measurements of POP and existing GWAS data from within the Women's Health Initiative (WHI) Hormone Therapy (HT) trial, we performed a trans-ethnic GWAS study for POP in European American (EA), African American (AA) and Hispanic (HP) women. Women with ≥ 1 pelvic exam at baseline or during follow-up (evaluated for the presence and severity of rectocele, cystocele and uterine prolapse) and for whom GWAS data was available were eligible. Women with grade 0 prolapse were considered controls. Cases were defined as women with any POP (grades 1-3) or moderate/severe POP (grades 2-3). We first performed genotyping platform/race-specific multiple logistic regression analyses between SNPs imputed to the 1000 genomes (all reference panels) and POP (grade 0 vs. 1-3 and grade 0 vs. 2-3) while adjusting for key risk factors including age at POP ascertainment, body mass index, parity and continuous axes of genetic ancestry. We then performed random-effects meta-analyses on effect estimates from all four subsets (AA, HP and 2 subsets of EA women) within the WHI-HT. A total of 7523 any POP cases, 2078 moderate/severe POP cases and 4964 controls were included in our final analyses. While we did not find any genome-wide significant findings, in the grade 0 vs. 1-3 meta-analyses, a common intronic variant in the *ETV6* gene rs11834045 approached GWAS significance. Each unit increase in the T allele vs. the C allele was associated with increased risk for POP with odds ratio (OR) of 1. 17 (95% CI: 1. 10, 1. 23; P=7. 9x10-8); effect estimates were in the same direction in all four datasets. Similar results were observed for this SNP in the grade 0 vs. 2-3 analyses (OR: 1. 16, 95% CI: 1. 07, 1. 26; P: 5. 0x10-4). SNPs upstream of the *SIRPG* gene were nominally associated with POP (rs6034359 {C vs. T}: OR 0. 83; 95% CI: 0. 71, 0. 97; P=1. 3x10-7) when considering grade 0 vs. 1-3 POP. Estimates from the grade 0 vs. 2-3 analyses were also similar as in the grade 0 vs. 1-3 analyses (OR: 0. 85, 95% CI: 0. 76, 0. 94; P: 4. 5x10-4) for this SNP. Meta-analyses with additional cohorts are currently underway. Our results suggest there may beassociationsbetween common SNPs in two novel loci and POP in the first trans-ethnic meta-analysis in EA, HP and AA women.

922F

Genomewide Association of Polycystic Ovary Syndrome Implicates Alterations in Gonadotropin Secretion in European Ancestry Populations. *M. Hayes[1,2,3], M. Urbanek[1,2], D. A. Ehrmann[4], L. L. Armstrong[1], J. Y. Lee[1], R. Sisk[1], T. Karaderi[5], T. M. Barber[6], M. I. McCarthy[5,7,8], S. Franks[9], C. M. Lindgren[5,10], C. K. Welt[11], E. Diamanti-Kandarakis[12], D. Panidis[13], M. O. Goodarzi[14], R. Azziz[15], Y. Zhang[16,17], R. James[16,17], M. Olivier[18], A. H. Kissebah[16,17], . Reproductive Medicine Network[20], E. Stener-Victorin[19], R. S. Legro[20], A. Dunaif[1,2].* 1) Endocrinology Metabolism & Med, Northwestern Univ Sch Med, Chicago, IL; 2) Ctr Genetic Med, Northwestern Univ Sch Med, Chicago, IL; 3) Anthropology, Northwestern University, Evanston, IL; 4) Endocrinology, Diabetes & Metabolism, Univ Chicago, Chicago, IL; 5) Wellcome Trust Ctr Human Genetics, Univ Oxford, Oxford, UK; 6) Warwick Med Sch, Univ Warwick, UK; 7) Oxford Ctr Diabetes, Endocrinology & Metabolism, Univ Oxford, Oxford, UK; 8) Oxford NIHR Biomed Res Ctr, Churchill Hosp, Headington, UK; 9) Inst Reproductive & Develop Biol, Hammersmith Hospital, Imperial College London, London, UK; 10) Med & Population Genetics, Broad Inst Harvard & MIT, Cambridge, MA; 11) Endocrinology, Metabolism & Diabetes, Univ Utah, Salt Lake City, UT; 12) Endocrinology & Metabolism, Univ Athens Medical School, Athens, Greece; 13) Endocrinology & Human Reproduction, Obstetrics & Gynecology, Aristotle Univ of Thessaloniki, Greece; 14) Endocrinology, Diabetes and Metabolism, Cedars-Sinai Med Ctr, Los Angeles, CA; 15) Obstetrics & Gynecology Med Coll of Georgia, Georgia Regents University, Augusta, GA; 16) TOPS Obesity & Metabolic Res Ctr, Med Coll Wisconsin, Milwaukee, WI; 17) Human and Molecular Genetics Center, Medical Coll of Wisconsin, Milwaukee, WI; 18) Genetics, Texas Biomed Res Inst, San Antonio, TX; 19) Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden; 20) Obstetrics & Gynecology, Penn State Coll Med, Hershey, PA.

Polycystic ovary syndrome (PCOS) is a common, highly heritable complex disorder of unknown etiology characterized by hyperandrogenism, chronic anovulation and defects in glucose homeostasis. Increased luteinizing hormone relative to follicle stimulating hormone secretion, insulin resistance and developmental exposure to androgens are hypothesized to play a causal role in PCOS. Here we map common genetic susceptibility loci in European ancestry women for the 1990 National Institutes of Health (USA) PCOS phenotype of hyperandrogenism and anovulation, which confers the highest risk for metabolic morbidities (e. g. insulin resistance and dysglycemia), as well as quantitative reproductive hormone levels. We performed a discovery GWAS (Stage 1) of 984 PCOS case and 2,964 population control women of European ancestry genotyped on the Illumina Omni Express, followed by replication (Stage 2) in 1,799 PCOS case and 1,231 phenotyped reproductively normal control European ancestry women genotyped on the Metabochip. Both stages were imputed (IMPUTE v2) to the 1000 Genomes March 2012 (v3) reference panel, and the resultant genotype call probabilities were used in a logistic regression model between PCOS case-control status and the genotype probabilities under an additive model adjusting for BMI and the first three principal components of ancestry. The *b* and standard errors were combined across the GWAS and replication cohorts using meta-analysis (METAL) under a fixed effects model weighting each strata by sample size. We followed this with a second replication (Stage 3) of the top 24 associations after a meta-analysis of stage 1 and 2 in a UK cohort of 217 PCOS cases and 1,335 population controls. Three loci reached genomewide significance in the three-strata meta-analysis; two novel loci, chr 8p32. 1 *GATA4/NEIL2* (rs804279, A allele, stage-specific OR=0. 74-0. 82, Pmeta=8. 0x10-10), and chr 11p14. 1 *FSHB* (rs11031006, G allele, stage-specific OR=1. 12-1. 41, Pmeta=1. 9x10-8), and one previously found in Han Chinese PCOS GWAS, chr 9q22. 32 *c9orf3/FANCC* (rs10993397, C allele, stage-specific OR= 0. 72-0. 88, Pmeta=4. 6x10-13). The same chr 11p14. 1 SNP near *FSHB* also reached genomewide significance in a meta-analysis of linear regression results of the quantitative luteinizing hormone (LH) levels in the PCOS cases (rs11031006, G allele, stage-specific *b*=0. 60-3. 19 mIU/mL, Pmeta=2. 7x10-16). These findings implicate gonadotropins in the pathogenesis of PCOS.

923W

The Netherlands Twin Register Axiom Biobank platform for GWAS. *J. Hottenga[1,2], E. Ehli[3], A. Abdellaoui[1], I. Fedko[1], G. Willemsen[1], E. J. C. N. de Geus[1,2], C. Grieser[4], G. E. Davies[3], D. I. Boomsma[1].* 1) The Netherlands Twin Register, VU University, Amsterdam, NL; 2) EMGO institute for Health and Care Research, VU University and VU University Medical Center, Amsterdam, NL; 3) Avera Institute for Human Genetics, Sioux Falls, SD, USA; 4) Affymetrix, Santa Clara, CA, USA.

With the prices of SNP arrays rapidly dropping, it is now more cost efficient to generate genotype data by using genome-wide chips than for example by a 30 SNP based Iplex. The Affymetrix Axiom Biobank Array was developed for large genotyping projects, for example the UK Biobank and the Million Veterans Project. We developed a NTR Axiom Biobank Array for the Netherlands Twin Register to get data with respect to twin zygosity, ancestry, a good imputation backbone with enough extra SNP coverage for areas difficult to impute, like the HLA region, mitochondrial SNPs and the X chromosome. Sets of SNPs for known GWAS hits are also included, e. g. SNPs involved in psychiatric, cardiovascular, metabolomic, pharmacogenomic and other phenotypes. To fit the genetic data with earlier NTR genotyping, we constructed the imputation backbone of the chip to be backward compatible with the Affymetrix 6. 0 platform. We have tested the genome-wide coverage and imputation accuracy of the full NTR Axiom Biobank Array and of the combined Affymetrix 6. 0-Axiom backbone using the GoNL and 1000G reference sets. In this experiment two different marker sets were used for testing, i) markers from the combined backbone only (n= 555,029) and ii) a marker set specific to the full custom designed NTR Axiom Biobank Array (n= 618,889). These marker sets were extracted for 249 unrelated females from the GoNL reference set, and then imputed against the full 1000G reference. The concordance of the imputed SNPs was subsequently compared to the original GoNL genotyping in 9. 5M polymorphic GoNL SNPs. Results show that the genome-wide coverage of the full NTR Axiom Biobank Array is 53. 8% with >99% genotype accuracy and 88. 8% with >95% genotype accuracy. 99. 1% of the SNPs have a concordance of 80% or higher without any filtering on R2 imputation quality of SNPs or MAF. For the combined backbone these values are 52. 1%, 87. 6% and 98. 9%, respectively. Furthermore we show that when only the combined backbone is imputed, there is a significant difference in allele frequencies for several SNPs in regions of the genome ($p < 0.00001$); most noticeable on chromosome 1, 6-HLA, 11, 14, 17, 19 and X. This confirms the choice of these regions to be more densely genotyped to improve the overall imputation accuracy. In conclusion, we have developed an improved genome chip to quickly assess 619K SNPs within Dutch individuals, which can be employed for GWAS on biomedical and psychiatric traits. .

924T

Discoveries from platelet count genome-wide association study in the Hispanic Community Health Study: Study Of Latinos (HCHS/SOL). *D. Jain[1], UM. Schick[2,3], JV. Morrison[1], T. Sofer[1], CJ. Hodonsky[4], JP. Davis[4], S. Vadlamudi[4], KL. Mohlke[4], C. Shurmann[3], RJ. Loos[3], CC. Laurie[1], T. Thornton[1], KE. North[4], AP. Reiner[1,2].* 1) University of Washington, Seattle, WA; 2) Fred Hutchinson Cancer Research Center, Seattle,WA; 3) Mt. Sinai School of Medicine, New York, NY; 4) University of North Carolina, Chapel Hill, NC.

Hispanic Latinos are a diverse population in which the genetic variability of platelet count (PLT) has not been analyzed. We performed a genome-wide association study of PLT in 12,491 participants of HCHS/SOL. Participants were genotyped on the Illumina SOL Omni2. 5M array (plus ~150k custom content) and imputed to the 1000 Genomes Phase I reference panel. We implemented linear mixed-model regression adjusting for sex, age, study center, ancestry and sample weights, and including random effects for individual relatedness and sampling design factors. Ten independent loci were significantly (P<5x10-8) associated with PLT in HCHS/SOL, including three novel loci and several known PLT loci (e. g. *ARHGEF3, TPM4*). 49% of the previously identified European platelet-count GWAS loci generalized to HCHS/SOL with directional consistency (FDR 5%). Novel loci included a Native American variant in *ACTN1* (rs117672662, b=0. 61, P=8. 65x10-29), *GABBR1* (rs3131857, b = -0. 16, P = 3. 97x10-10) and a suggestive association near *ETV7* (rs9470264, b = -0. 19, P=5. 64x10-8). The novel associations were successfully replicated in 6,434 Hispanic-American samples from the Women's Health Initiative and BioMe. Conditional analyses identified secondary association signal near *HBS1L-MYB* (rs6934903, b = 0. 20, P = 2. 79x10-8). Additionally, we identified new population-specific variants at known PLT GWAS loci (*MEFC2, BAK1*). Functional annotation of the three novel loci using epigenetic data from megakaryocytes (platelet precursors) identified seven SNPs that localized in putative enhancers or promoters and were in high linkage disequilibrium with the index SNP. In particular, *ACTN1* variant rs117672662 localized in a megakaryocyte-specific putative enhancer in intron 1. Furthermore, transcriptional reporter and gel shift assays in THP-1 monocytes with rs117672662 showed allelic differences in transcriptional activity and protein-binding, supporting a regulatory role. Further investigations will be necessary to identify gene targets of the putative enhancer in which rs117672662 resides. A likely target is *ACTN1*, which harbors causal mutations for congenital macrothrombocytopenia. In conclusion, discovery of a putative functional variant of Native American ancestral background in previously identified familial thrombocytopenia locus (*ACTN1*) as well as two other novel loci (*ETV7* and *GABBR1)* underscore the value of using diverse populations in GWAS.

925F

Genome-Wide Association Study of HIV-Associated Neurocognitive Disorder (HAND): a CHARTER Group Study. *P. Jia[1], A. Kallianpur[2], T. Hulgan[3], W. Bush[4], D. Samuels[5], Z. Zhao[1,6], C. Bloss[7], R. Heaton[7], R. Ellis[8], N. Schork[9], C. Marra[10], A. C. Collier[11], D. Clifford[12], B. Gelman[13], N. Sacktor[14], S. Morgello[15], D. Simpson[15], J. A. McCutchan[16], S. Letendre[16], I. Grant[7], the CHARTER Study Group.* 1) Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN; 2) Department of Genomic Medicine, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH; 3) Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN; 4) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 5) Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, TN; 6) Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, TN; 7) Department of Psychiatry, University of California San Diego, San Diego, CA; 8) Department of Neurology, University of California San Diego, San Diego, CA; 9) Scripps Translational Science Institute, La Jolla, CA; 10) Department of Neurology, University of Washington, Seattle, WA; 11) Department of Medicine, University of Washington, Seattle, WA; 12) Department of Neurology, Washington University, St. Louis, MO; 13) Department of Pathology, University of Texas Medical Branch, Galveston, TX; 14) Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD; 15) Department of Neurology, Icahn School of Medicine of Mount. Sinai, New York, NY; 16) Department of Medicine, University of California San Diego, San Diego, CA.

HIV-associated neurocognitive disorder (HAND) complicates HIV infection even during combination antiretroviral therapy (cART) and may be influenced by host genomics. We performed a genome-wide association study (GWAS) of HAND in 1050 consenting participants in the CNS HIV Antiretroviral Therapy Effects Research (CHARTER) Cohort. All participants underwent standardized, comprehensive neurocognitive and neuromedical assessments to determine a Global Deficit Score (GDS) and were classified as having neurocognitive impairment (NCI) by GDS≥0. 5. To minimize confounding by conditions other than HIV, individuals with major neurocognitive confounds such as traumatic brain injury or developmental learning disorders were excluded from analysis. NCI was also categorized according to Frascati definitions [asymptomatic neurocognitive impairment (ANI; n=359), mild neurocognitive disorder (MND; n=92), and HIV-associated dementia (HAD; n=30)]. HAND outcomes included GDS-defined NCI (binary GDS, 366 cases with GDS≥0. 5 and 684 controls with GDS<0. 5), GDS as a continuous variable, dichotomized ANI, MND, or HAD, or combined ANI and MND, each *vs.* the neurocognitively normal referent category. Genotype data was obtained using the Affymetrix Human SNP Array 6. 0 platform. Standard quality control (QC) procedures were conducted. The post-QC dataset included 812 men and 238 women. Multivariable logistic or linear regression-based association tests were performed for each NCI phenotype. The genomic inflation factor implied no systematic inflation (αranging from 0. 99 to 1. 03). GWAS results did not reveal SNPs meeting the genome-wide significance threshold (5. 0×10-8) in any of the six NCI phenotypes. For binary GDS, the most significant SNPs were rs6542826 (p=8. 1×10-7) and rs11681615 (1. 2×10-6), both located in *SH3RF3*. For continuous GDS, the most significant SNP was rs11157436 (p=1. 3×10-7) on chromosome 14 in the T-cell-receptor *alpha* locus. For all HAND, and its milder forms (MND and ANI), the most significant SNP was located upstream of *PTPN3* (p=1. 4×10-7 and p=2. 2×10-7, respectively). Preliminary analysis also suggests enrichment of the functional pathway "essential iron-containing enzymes, flavoproteins or cytochromes" in HAND overall (p=0. 023) and in only milder forms of HAND (p=0. 007). This GWAS, conducted among cART-era participants from a single cohort with robust neurological phenotypes, suggests possible roles for several loci in HAND that require further exploration.

926W

Genome-wide meta-analysis of histological phenotypes provides insights into the genetic architecture of human skeletal muscle. *T. Karaderi[1], N. Oskolkov[2], K. Strom[2,5], S. Keildson[1], C. Ladenvall[2], A. Mahajan[1], L. Lind[3], E. Ingelsson[4], H. Holmberg[5], N. Fuku[6], G. Wang[6], Y. Pitsiladis[6], A. P. Morris[1], R. Scott[7], L. Groop[2], P. Franks[8], C. M. Lindgren[1], O. Hansson[2].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, Malmö, Sweden; 3) Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 4) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 5) Swedish Winter Sports Research Centre, Department of Health Sciences, Mid Sweden University, Östersund, Sweden; 6) Centre for Sport and Exercise Science and Medicine, University of Brighton, Brighton, UK; 7) MRC Epidemiology Unit, School of Clinical Medicine, University of Cambridge, Cambridge, UK; 8) Department of Clinical Sciences, Genetic and Molecular Epidemiology, Lund University, Malmö, Sweden.

Muscle fiber-type distribution affects basic processes such as movement and cellular respiration. Athletes in endurance and explosive sports have a higher proportion of slow-twitch type I and fast-twitch type II fibers, respectively. Fiber-type distribution has also been associated with obesity and type 2 diabetes. Insulin resistant individuals tend to have reduced muscle oxidative capacity with less oxidative 'red' type I and more glycolytic 'white' type IIX fibers. Muscle capillary density is also a key determinant of glucose uptake by skeletal muscle; it was observed that impaired glucose tolerance worsens with reduced capillary density. Twin studies showed that ~45% of phenotypic variance is due to a genetic component. Thus, we aimed to identify genetic factors and biological mechanisms related to skeletal muscle histological phenotypes. We conducted a genome-wide fixed-effects meta-analysis of muscle histological phenotypes including capillary density and fiber-type distribution in Swedish males from 3 cohorts (n=656). Genotype data were imputed to 1000 Genomes all ancestries panel (March 2012). In each cohort, phenotypes were inverse rank normalized and tested for association with each imputed variant under an additive model adjusting for age and body mass index. Muscle microarray expression data (n=77) were used for eQTL analysis of associated markers. Follow-up of capillary density was done in elite Swedish cross-country skiers (n=15) and Jamaican sprinters (n=119). Immunohistochemistry staining (IHC) was also performed in skeletal muscle. We identified 11 independent loci (*STEAP*, *NYAP2*, *ADRA1B*, *TNFSF11*, *FAM155A*, *SLC22A10*, *FASLG*, *RBFOX1*, *FOXJ2*, *KCNMA1* and *RAB3GAP2*) at p<5x10-8 associated with six phenotypes. G allele of rs115660502 (minor allele frequency (MAF)=0. 048) was associated with increased capillary density (p=2x10-8) and decreased *RAB3GAP2* expression (FDR=0. 007). G allele had a higher frequency in Swedish skiers (MAF=0. 10, p=0. 127) and a lower frequency in Jamaican sprinters (MAF=0. 004, p=0. 031) compared to respective population controls. RAB3GAP2 protein, the regulatory subunit of Rab3 GTPase-activating complex involved in exocytosis, was also localized to muscle capillaries in IHC staining. These results expand our understanding of muscle architecture, implicating multiple genetic factors. The rs115660502 association indicates that RAB3GAP2 regulates muscle capillary density and may thereby contribute to endurance performance in athletes.

927T

Genome-wide association study of spontaneous preterm birth. *M. Karjalainen[1,2], M. Ojaniemi[1,2], M. Rämet[1,2], M. Hallman[1,2].* 1) PEDEGO Research Center, and Medical Research Center Oulu, University of Oulu, Oulu, Finland; 2) Department of Children and Adolescents, Oulu University Hospital, Oulu, Finland.

Background and aims: Spontaneous preterm birth (SPTB) is a major factor associating with deaths and with lowered quality of life. According to family studies, SPTB has a significant genetic background with a potential role for both maternal and fetal genetic factors. However, little is known about the actual predisposing genes. The aim of the study was to discover genetic variants predisposing to SPTB using a genome-wide association study (GWAS). **Methods**: The study population consisted of mothers with SPTB (gestational age, GA <36 weeks) and SPTB infants, and mothers with exclusively term deliveries (GA 38-41 weeks) and infants born at term recruited in Oulu and Tampere University Hospitals, Finland. Only individuals of Finnish origin were included. DNA samples were genotyped with the Illumina HumanCoreExome BeadChip consisting of approximately 550,000 single-nucleotide polymorphisms (SNPs). The analyses were performed with PLINK, v. 1. 09. In addition to the study of SPTB as a dichotomous trait, GA was studied as a quantitative trait. After quality control, 230 case and 378 control mothers, and 250 case and 418 control infants remained for GWAS. **Results**: We detected potential association signals ($p < 10\text{-}4$) for several SNPs in GWAS. Many of these SNPs were located in the vicinity of genes that can be considered as plausible candidate genes for SPTB; these included e. g. the *IGSF21*, *IFI44*, and *COL23A1* genes encoding immunoglobin superfamily, member 21; interferon induced protein 44; and collagen, type XXIII, alpha 1, respectively. In the next step of the study, we will analyze the most promising SNPs in additional populations from Finland and other countries to assess which associations are replicated. **Conclusions**: In GWAS of spontaneous preterm birth, we detected several suggestive associations. These initial results will have to be replicated in additional populations. Furthermore, the role of the arising candidate genes in the onset of preterm delivery will have to be investigated using functional studies.

928F

Uncovering novel genes contributing to phenotypic heterogeneity in a common birth defect. *E. J. Leslie[1], J. C. Carlson[1,2], J. R. Shaffer[3], E. Feingold[2], C. Laurie[4], K. Doheny[5], L. Moreno[6], A. E. Czeizel[7], L. L. Field[8], F. Deleyiannis[9], G. Wehby[10], E. E. Castilla[11], A. R. Vieira[1,3], A. Butali[12], K. Christensen[13], L. Ma[14], C. J. Buxo-Martinez[15], J. T. Hecht[16], K. Neiswanger[1], A. C. Lidral[6], T. H. Beaty[17], J. C. Murray[18], S. M. Weinberg[1], M. L. Marazita[1].* 1) Center for Craniofacial and Dental Genetics, Department of Oral Biology, University of Pittsburgh, Pittsburgh, PA; 2) Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 3) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 4) Genetic Coordinating Center, Department of Biostatistics, University of Washington, Seattle, WA; 5) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 6) Department of Orthodontics, College of Dentistry, University of Iowa, Iowa City, IA; 7) Foundation for the Community Control of Hereditary Diseases, Budapest, Hungary; 8) Department of Medical Genetics, University of British Columbia, Vancouver, Canada; 9) Department of Surgery, Plastic and Reconstructive Surgery, University of Colorado School of Medicine, Denver, CO; 10) Department of Health Management and Policy, College of Public Health, University of Iowa, Iowa City, IA; 11) Center for Medical Education and Clinical Research, Buenos Aires, Argentina; 12) Dows Institute for Dental Research, College of Dentistry, University of Iowa, Iowa City, IA; 13) Department of Epidemiology, Institute of Public Health, University of Southern Denmark, Odense, Denmark; 14) Peking University, School of Stomatology, Beijing, China; 15) School of Dental Medicine, University of Puerto Rico; 16) Department of Pediatrics, University of Texas Health Science Center at Houston, Houston, TX; 17) Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD; 18) Department of Pediatrics, University of Iowa, Iowa City, IA.

Nonsyndromic cleft lip with or without cleft palate (CL/P) is one of the most common birth defects in humans, affecting approximately 1 in 700 newborns. Like many other common diseases, CL/P is phenotypically heterogeneous and has a complex etiology caused by genetic and environmental risk factors. The primary subphenotypes of CL/P are cleft lip (CL) and cleft lip with cleft palate (CLP). Historically CL and CLP have been considered variants of the same defect with a common genetic etiology (CL/P) and are typically combined for genetic analyses. Previous genome-wide association studies (GWAS) have identified at least 15 risk loci explaining approximately 20% of the overall heritability of CL/P. We conducted a GWAS in a multiethnic cohort of 1,319 case-parent trios with CL/P and combined these results with a previously published GWAS of 1,593 trios (Beaty et al. 2010 Nature Genetics) for a total of 2,912 trios. We identified eleven loci associated with CL/P at the genome-wide level (P<10-8) strengthening previous associations with *IRF6*, 8q24, *NTN1*, *ARHGAP29*, and *COL8A1*. Among the SNPs showing evidence of association with 10-8-5 were four other previously identified loci and novel loci near biologically relevant genes including *TP63*, *WNT9B*, *SMAD2*, and *MYCN*. Further, we hypothesized that CL and CLP could have unique risk factors whose signal would be diluted in the combined CL/P analysis. To address this, we stratified our trios by subphenotype (N= 731 CL trios, N=2,181 CLP trios). Some signals from the combined analysis were common to both groups (e. g. *IRF6*, *8q24*) while others were unique to CL or CLP. Specifically, we found three loci that were strongly associated with only CL (P<10-6), but not CLP (P>0. 3). One of these was on 1p36 near *SKI*. The *Ski* knockout mouse has a midline facial cleft, providing support for this unique association with CL. Five other loci had strong signals in only CLP (P<10-6), but not in CL (P>0. 2). Among these were gene deserts on chromosomes 11 and 16, suggesting causal variants may affect non-coding regulatory elements. Overall, these analyses have identified additional novel risk loci contributing to risk of CL/P. Furthermore, our results provide substantial genetic support for the epidemiological observation that CL and CLP could have some distinct risk factors. This work was supported by grants X01-HG00784 and R01-DE016148.

929W

Genome-wide association study of acute otitis media in children. *J. Li[1], G. van Ingen[2], A. Goedegebure[2], M. E. March[1], V. W. V. Jaddoe[3], F. D. Mentch[1], A. G. Uitterlinden[3], H. A. Moll[3], C. M. van Duijn[3], F. Rivadeneira[3], R. J. Baatenburg de Jong[2], M. P. van der Schroeff[2], H. Hakonarson[1,4,5].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Otolaryngology, Head and Neck Surgery, Erasmus University Medical Center, the Netherlands; 3) The Generation R Study, Department of Epidemiology, Erasmus University Medical Center, the Netherlands; 4) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 5) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Acute otitis media (AOM) is a common disease ofearly childhood, and the most common reason for antibiotic treatment in children. Environmental, host and genetic factors together account for the etiology of OM, with an estimated heritability of 22-74%. We performed a genome-wide association study (GWAS) investigating genetic susceptibility loci to AOM with age of onset less than 3 years old. Our study was conducted among children of European ancestry at the Children's hospital of Philadelphia (CHOP, n=8,751) and the Generation R study, Erasmus University Medical Center in the Netherlands (GenR, n=2,286). Discovery GWASs were performed at both centers, followed by meta-analysis. A locus at 6q25. 3 reached genome-wide significance (rs2932989, P=4. 4x10-8). We further replicated the results in independent cohorts at CHOP (n=1,840) and GenR (n=2,638). This locus also significantly correlates with methylation probe cg05678571 in gene FNDC1. Therefore, by GWAS, we identified the first genome-wide significant locus associated with this common pediatric disease.

930T

Findings from the Initial Phase of the GSCAN Exome Chip Project. *M. Liu[1], D. Liu[2], S. Bertelsen[18], Y. L. Chou[3], S. David[4], J. Faul[5], J. Gong[6], A. Hammerschlag[7], C. Hsu[8], D. Irons[9], A. Jackson[10], A. Loukola[11], M. Mangino[12], G. Pistis[13], R. Rhode[14], Y. Shao[14], M. Steri[13], H. Stringham[10], V. Turcot[15], L. Wetherill[17], W. Zhao[16], S. Vrieze[1], CHDExome+Cons. , Cons. for Genetics of Smoking Behaviour, GWAS&Sequencing Cons. of Alcohol & Nicotine.* 1) Institute for Behavioral Genetics, University of Colorado - Boulder, Boulder, CO; 2) College of Medicine, Penn State University, State College, PA; 3) Department of Psychiatry, Washington University, St. Louis, MO; 4) Stanford University Medical Center, Stanford University, Stanford, CA; 5) Institute for Social Research, University of Michigan, Ann Arbor, MI; 6) Fred Hutchinson Cancer Center, Seattle, WA; 7) Center for Neurogenomics & Cognitive Research, VU University of Amsterdam, Amsterdam, Netherlands; 8) Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA; 9) Psychology Department, University of Minnesota, Minneapolis, MN; 10) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 11) Department of Public Health, University of Helsinki, Helsinki, Finland; 12) Twin Research & Genetic Epidemiology Unit, King's College London, London, UK; 13) Institute of Neurogenetics and Neuropharmacology, Consiglio Nazionale delle Ricerche, Rome, Italy; 14) Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC; 15) Montreal Heart Institute, Montreal; 16) School of Public Health, University of Michigan, Ann Arbor, MI; 17) Department of Medical and Molecular Genetics, Indiana University-Purdue University at Indianapolis, Indianapolis, IN; 18) Department of Neuroscience, Icahn School of Medicine at Mount Sinai, Manhattan, NY.

Tobacco and alcohol use are complex behaviors known to be influenced by genetic factors from family studies. Genetic association studies have found common variants associated with tobacco and alcohol use (e. g. *CHRNA5*and*ALDH2*respectively). Recent developments in rare variant genotyping now allow testing association between these behaviors and rare nonsynonymous variants identified through exome sequencing. To this end, the GWAS & Sequencing Consortium of Alcohol and Nicotine use (GSCAN) was formed, which has aggregated exome array association results across 18 cohorts with 5 phenotypes including cigarettes per day (N=34,611), pack years (N=32,285), age of initiation of smoking (N=22,982), smoking initiation (N=51,331), and drinks per week (N=48,890). 17 cohorts genotyped using the Illumina HumanExome array (~250,000 low-frequency markers) and one cohort genotyped with the Illumina HumanCoreExome array (~250,000 low-frequency markers and 250 common tag SNPs). After extensive quality control, a 463,184 variants were found to segregate in the dataset. Among these variants, 158,636/3,731/1,473 are non-synonymous/loss-of-function/splice variants with MAF<1%. To aggregate information between cohorts and enlarge sample sizes, we generated summary association statistics in each cohort and performed single variant analyses on all variants and a variety of gene-based meta-analyses on low-frequency nonsynonymous variants with MAF<1%. We found associations at known loci and identified a novel signal associated with cigarettes per day at rs36015615 in*STARD3*(N=34,611, Beta=1. 3, p=9. 5e-9). This signal did not replicate in two independent cohorts, the Consortium for Genetics of Smoking Behavior (N=28,583, Beta=. 056,*p*=. 84) and the CHD+Exome consortium (N=17,789, Beta=-. 01,*p*=. 94). We will also report results from multivariate single variant and gene based tests. GSCAN continues to aggregate additional samples to increase power for rare variant associations.

931F

A cis-eQTL near the *ZNF682* gene is associated with risk for diabetic nephropathy in Pima Indians. *I. Masindova[1], Y. L. Muller[1], P. Piaggi[1], S. Kobes[1], P. Chen[1], V. Nair[2], M. Kretzler[2], W. C. Knowler[1], R. G. Nelson[1], R. L. Hanson[1], C. Bogardus[1], L. J. Baier[1].* 1) Phoenix Epidemiology and Clinical Research Branch, NIDDK/NIH, Phoenix, AZ; 2) Department of Internal Medicine and Computational Medicine, University of Michigan, Ann Arbor, MI.

Diabetic nephropathy (DN) is a serious chronic complication of type 2 diabetes (T2D) and the leading cause of end-stage renal disease (ESRD). Pima Indians are highly affected by T2D and DN; therefore, the aim of our study was to identify susceptibility loci for DN in this population. A custom Pima Indian Affymetrix Axiom Array has been previously analyzed for 548,206 markers that capture 92% of the common variation (MAF≥5%, redundancy defined as r2≥0. 85) detected in whole genome sequence data from 335 Pima Indians. Subjects genotyped on this array included 1475 full-heritage Pima Indians, of whom 344 had T2D and DN (urinary albumin-to-creatinine ratio ≥300mg/g or presence of ESRD) and 1131 had T2D but not DN. Replication of all variants preliminary associated with DN is ongoing in a second sample of 800 non-full heritage Pima subjects (on average ½ Pima) which included 140 subjects with T2D and DN, and 660 subjects with T2D but not DN. To date, replication of 27 SNPs has been assessed in this independent second sample. SNP rs10407776 associated with DN in the first sample [OR=1. 51 per copy of the G allele, p=2x10-5], replicated in the second sample [OR=1. 35, p=0. 047]) and combining both samples provided the strongest evidence for association with DN [OR=1. 44, p=9. 44x10-6 adjusted for sex, T2D duration and heritage]. This SNP maps 31kb upstream of the gene *ZNF682*. Although 23 additional tagging SNPs across the *ZNF682* genomic region (chr19:20015227-20250277, ~100kb flanking each size of the gene) also nominally associated with DN, conditional analyses suggested that these signals were dependent on rs10407776. Merging genotypic data with expression data of 185 adipose and 190 muscle biopsies of non-diabetic Pima Indians (Affymetrix Human Exon 1. 0 ST Array) and 55 kidney glomeruli from diabetic American Indians (Affymetrix U133 Array) provided evidence that rs10407776 may act as a cis-expression quantitative trait locus (cis-eQTL) to influence *ZNF682* expression while concomitantly influencing risk for DN. The DN risk allele (frequency of G=53%) correlated with lower expression of *ZNF682* in adipose [r=-0. 275, p=1. 60x10-4], muscle [r=-0. 269, p=1. 90x10-4] and kidney glomeruli [r=-0. 316, p=0. 022]. Although *ZNF682* has not been functionally implicated in development of DN, its protein product binds to ubiquitin C and this complex regulates several target genes involved in the pathogenesis of DN.

932W

Genome-wide association study of clinically-defined gout identifies multiple risk loci: a clue for future companion diagnostics of gout. *H. Matsuo[1], K. Yamamoto[2], H. Nakaoka[3], A. Nakayama[1], M. Sakiyama[1], A. Takahashi[4, 5], T. Nakamura[6], I. Danjoh[7], S. Shimizu[1], Y. Kawamura[1], S. Tatsukawa[1], Y. Sakurai[8], N. Hamajima[9], I. Inoue[3], M. Kubo[5], K. Ichida[10], H. Ooyama[11], T. Shimizu[12], N. Shinomiya[1].* 1) Dept Integrative Physiol, National Defense Med College, Tokorozawa, Japan; 2) Department of Medical Chemistry, Kurume University School of Medicine, Kurume, Fukuoka, Japan; 3) Department of Integrated Genetics, National Institute of Genetics, Mishima, Shizuoka, Japan; 4) Research Center, National Cerebral and Cardiovascular Center, Suita, Osaka, Japan; 5) Center for Integrative Medical Sciences, RIKEN, Yokohama, Kanagawa, Japan; 6) Laboratory for Mathematics, National Defense Medical College, Tokorozawa, Saitama, Japan; 7) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 8) Department of Preventive Medicine and Public Health, National Defense Medical College, Tokorozawa, Saitama, Japan; 9) Department of Healthcare Administration, Nagoya University Graduate School of Medicine, Nagoya, Japan; 10) Department of Pathophysiology, Tokyo University of Pharmacy and Life Sciences, Tokyo, Japan; 11) Ryougoku East Gate Clinic, Tokyo, Japan; 12) Midorigaoka Hospital, Osaka Japan.

Background and objectives Gout is a common disease which results from hyperuricemia. Recently, genome-wide association studies (GWASs) of gout have been reported; however, they included self-reported gout cases. Therefore, the relationship between genetic variation and clinical subtypes of gout remains to be clarified. Thus, we first performed a GWAS of clinically-defined gout cases only. **Design, setting, participants & measurements** In this study, GWAS was conducted with 945 male patients with clinically-defined gout cases and 1,213 male controls in a Japanese population. Additionally, a replication study of 1,048 clinically-defined cases and 1,334 controls was performed. **Results** Five gout susceptibility loci were identified at the genome-wide significance level (p< 5. 0 × 10-8), which contained well-known urate transporter genes (*ABCG2* and *SLC2A9*) and additional genes reportedly having relationships with metabolic pathways: rs1260326 (p = 1. 9 × 10-12; OR = 1. 36) of *GCKR* (a gene for glucose and lipid metabolism), rs2188380 (p = 1. 6 × 10-23; OR = 1. 75) of *MYL2-CUX2* (genes associated with cholesterol and diabetes mellitus) and rs4073582 (p = 6. 4 × 10-9; OR = 1. 66) of *CNIH-2* (a gene for regulation of glutamate signaling). The latter two are identified as novel gout loci. Moreover, among the identified SNPs, we demonstrated that the SNPs of *ABCG2* and *SLC2A9* were differentially associated with subtypes of gout and clinical parameters underlying specific subtypes (renal overload type and renal underexcretion type). The effect of the risk allele of each SNP on clinical parameters showed significant linear relationships with the ratio of the case-control ORs for two distinct types of gout (r=0. 96 [p=5. 0×10-4] for urinary urate excretion and r=0. 96 [p=4. 8×10-4] for urate clearance). **Discussion** Our findings provide clues to better understand the pathogenesis of gout and will be helpful to develop companion diagnostics.

933T

Genome-wide meta-analysis of polycystic ovary syndrome in women of European ancestry identifies new loci in hormone pathways. *C. Meun[1], T. Karaderi[2], R. Magi[3], A. W. Drong[2], L. Armstrong[4], L. Broer[5], M. R. Jones[6], F. R. Day[7], PCOS Genetics Consortium.* 1) Reproductive Medicine, Erasmus MC, Rotterdam, Zuid-Holland, Netherlands; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 3) Estonian Genome Center, University of Tartu, Tartu, Estonia; 4) Division of Endocrinology, Metabolism and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, United States of America; 5) Genetic Laboratory, Department of Internal Medicine, Erasmus MC, Rotterdam, Zuid-Holland, Rottedam; 6) Division of endocrinology, Diabetes and Metabolism, Department of Medicine, Cedar Sinai Medical Center, Los Angeles, California, United States of America; 7) MRC Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom.

Polycystic Ovary Syndrome (PCOS) is the most common heritable endocrine disorder in reproductive-aged women, and a major cause of infertility with a prevalence of ~6-10%. It is characterised by hyperandrogenism, menstrual irregularities and polycystic ovarian morphology, and associated with metabolic disturbances including type 2 diabetes, obesity and dyslipidemia. Previous genome-wide association studies (GWAS) identified 11 loci associated with PCOS in Han Chinese. Our aim is to identify novel loci associated with PCOS in a large collection of European descent populations using meta-analysis of GWAS data. We performed a genomic control corrected inverse-variance weighted fixed-effects meta-analysis in up to 4,890 cases and 20,405 controls of European descent from 6 cohorts (Iceland, the Netherlands, United Kingdom, Estonia and two from the United States). All cases fulfilled either the NIH and/or the Rotterdam diagnostic criteria. Analysis models were adjusted for (1) age only, and (2) age and body mass index (BMI). All data were imputed to the 1000 Genomes March 2012 all ancestries reference panel. EasyQC was used to filter markers with minor allele frequency <1% and imputation quality R2<0. 3 or info<0. 4 within each cohort. Markers present in at least 2 studies and in at least 50% of the effective sample size were included in the final results and ANNOVAR was used to annotate the results. We identified 5 novel loci (*GATA4/NEIL2*, OR=1. 23, PBMI=5. 97x10-11, *FSHB*, OR=1. 33, PBMI=3. 63x10-10; *TMEM135*, OR=1. 21, PBMI=3. 73x10-8; *ZBTB16*, OR=1. 27, PBMI=1. 44x10-8; *XG*, OR=1. 75, PBMI=8. 49x10-11) at genome-wide significance (P<5. 00x10-8). We did not observe anygenome-widesignificantassociations within a 100kb window of the SNPs reported in the Han Chinese GWAS. However, two loci (*DENND1A*, OR=1. 41, PBMI=9. 74x10-7; *YAP1*, OR=1. 39, PBMI=6. 76x10-8) had suggestive P-values (PBMI<1. 00x10-6) and an additional three loci (*C9orf3,* OR=1. 14, P=4. 88x10-6; *RAB5B*, OR=1. 12, P=7. 26x10-5; *TOX3,* OR=1. 16, P=4. 43x10-5) had evidence for association with nominally significant P-values (PBMI<1. 00x10-4). PCOS associations remained significant after adjustment for BMI and therefore, were not considered secondary to BMI association. Observed loci such as *FSHB* and *DENND1A* highlight the role of hormone pathways in PCOS. Our results provide evidence for new genetic loci associated with PCOS, extending our knowledge of the genetic architecture and etiology of this complex endocrine disorder.

934F

A common polymorphism in *HIBCH* is a cobalamin-independent determinant of methylmalonic acid concentrations in blood. *F. Pangilinan[1], A. M. Molloy[2], J. L. Mills[3], B. Shane[4], M. B. O'Neill[1], D. M. McGaughey[1], A. Velkova[1], H. O. Abaan[1], P. M. Ueland[5], H. McNulty[6], M. Ward[6], J. J. Strain[6], C. Cunningham[7], M. Casey[7], C. D. Cropp[8], Y. Kim[8], J. E. Bailey-Wilson[8], A. F. Wilson[8], L. C. Brody[1].* 1) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Department of Clinical Medicine, School of Medicine, Trinity College Dublin, Dublin, Ireland; 3) Division of Intramural Population Health Research, Eunice Kennedy Shriver NICHD, Bethesda, MD; 4) Department of Nutritional Sciences and Toxicology, University of California, Berkeley, CA; 5) Section of Pharmacology, Institute of Medicine, University of Bergen and Haukeland University Hospital, Bergen, Norway; 6) Northern Ireland Centre for Food and Health, University of Ulster, Coleraine, Northern Ireland; 7) St James Hospital, Dublin, Ireland; 8) Computational and Statistical Genomics Branch, National Human Genome Research Institute, Bethesda, MD.

Vitamin B12 (cobalamin) deficiency affects up to 20% of older adults, with severe untreated deficiency leading to irreversible neurological damage. Clinical diagnosis of deficiency can be challenging because low serum cobalamin concentrations correlate poorly with symptoms. Serum methylmalonic acid (MMA), a sensitive biomarker regarded as a 'gold standard', is used to confirm a diagnosis of vitamin B12 deficiency. However, the role of genetic factors, other than cobalamin related genes that influence MMA concentrations, is not known. To identify genetic variation that influences MMA levels in blood, we measured serum MMA in 2,210 healthy young Irish adults and carried out a genome-wide association study (GWAS). MMA levels were strongly associated with SNPs in 3-hydroxyisobutyryl-CoA hydrolase (*HIBCH*, p=4. 9x10-52). This hydrolase acts in the valine catabolism pathway that produces methylmalonyl-CoA, the substrate for cobalamin-dependent mutase. The association signal in the *HIBCH* gene accounted for 9. 9% of the variance in MMA concentration. The most highly associated SNP (*HIBCH* rs291466, *c*. 2T>C) codes for a missense change of the ancestral initiator methionine (MAF=0. 43) to a threonine (*p*. Met1Thr). Surprisingly, the threonine (C) allele was associated with increased steady state levels of *HIBCH* mRNA and protein. Threonine (CC) homozygotes had, on average, 46% higher serum MMA than methionine (TT) homozygotes. The association between MMA and *HIBCH* rs291466 is also highly significant in two replication cohorts consisting of 189 pregnant Irish women and 1,481 older Irish adults. Despite pregnancy-related increases in MMA and the cobalamin deficiency seen in up to 20% of older adults, both replication cohorts exhibited significant association of *HIBCH* rs291466 with MMA levels (pregnancy cohort, p<0. 001; older adult cohort, p=2. 5x10-09). Observing such a robust association in the elderly is unexpected as MMA concentrations can increase with age due to impaired renal function and environmental factors unrelated to HIBCH activity. HIBCH is unique to valine catabolism. We suggest that flux through this pathway be further evaluated in situations where protein catabolism is elevated, *e. g*. , the elderly. Additionally, accounting for *HIBCH* rs291466 should increase the utility of MMA in evaluating vitamin B12 status. For example, homozygosity for the threonine (C) allele may contribute to elevated MMA in individuals whose vitamin B12 levels are difficult to interpret.

935W

Genome-wide association study to identify variants predisposing to severe bronchiolitis. *A. M. S. Pasanen[1], M. K. Karjalainen[1], M. Ruotsalainen[2], E. Piippo-Savolainen[2], E. Goksör[3], G. Wennergren[3], M. Hallman[1], M. Rämet[1], M. Korppi[4].* 1) PEDEGO Research Center and Medical Research Center Oulu, University of Oulu, Department of Children and adolescents, Oulu University Hospital, Oulu, Finland; 2) Kuopio University Hospital, Pediatrics, University of Eastern Finland, Kuopio, Finland; 3) Queen Silvia Children's Hospital, Pediatrics, University of Gothenburg, Gothenburg, Sweden; 4) Pediatric Research Center, Tampere University and Tampere University Hospital, Tampere, Finland.

Background and objective: Bronchiolitis is a common lower respiratory tract infection mainly affecting infants under 2 years of age. It is a leading cause of hospitalization for children of that age and considered to be a risk factor for developing asthma. Although candidate gene studies have found some associated loci, the genetic background of bronchiolitis is not well known. The aim of this study is to identify genetic polymorphisms predisposing to severe bronchiolitis and subsequent asthma by genome-wide association study (GWAS). **Methods:** Case-control populations of 217 cases hospitalized for bronchiolitis in infancy and 778 controls were genotyped with genome-wide bead chips of 300,000-700,000 variants. The study populations have been collected in Kuopio and Tampere, Finland and in Gothenburg, Sweden. Cases and some of the controls are followed-up in regard to respiratory health. 684 controls are Finnish population controls from Nordic Control Allele Frequency and Genotype Database (NordiCDB). Genotype imputation has been carried out and statistical analyses have been conducted with plink 1. 9. The most significantly associating variants will be analyzed in independent populations collected in the Netherlands and Finland. **Results**: Potential association signals (p-value ≤10-5) were seen for several single-nucleotide polymorphisms (SNPs). Many of these resided at intergenic regions, suggesting regulatory functions. Some of the best associating SNPs were in or adjacent to genes related to e. g. immune responses or metabolism. Preliminary candidate genes for bronchiolitis included e. g. *TRA* locus, encoding T cell receptor alpha chains, and *VCAN (VERSICAN)*, which interacts with *CCL8* encoding chemokine (C-C motif) ligand 8. In the next stage, the most promising variants are analyzed in an independent population to examine which associations are replicated. **Conclusions**: In GWAS of bronchiolitis, several preliminarily association signals were detected. These include for example variants in genes related to immune responses and metabolism. Many of the best associating variants are at intergenic regions, suggesting regulatory functions. The most promising associating SNPs will be further analyzed in independent replication population to verify real associations. Later, functional studies of relevant candidate genes will be performed.

936T

A genome-wide association study using a custom genotyping array identifies variant in *GPR158* associated with reduced energy expenditure and increased Body Mass Index in American Indians. *P. Piaggi[1], I. Masindova[1], Y. L. Muller[1], G. B. Wiessner[1], P. Chen[1], A. Gale[1], S. Kobes[2], R. L. Hanson[2], C. Bogardus[1], L. J. Baier[1].* 1) Diabetes Molecular Genetics Section, Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Disease, Phoenix, AZ; 2) Diabetes Epidemiology and Clinical Research Section, Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Disease, National Institutes of Health, Phoenix, AZ.

A relatively lower energy expenditure (EE) predicts long-term increases in body weight and fat mass, implying that energy metabolism contributes to the pathogenesis of human obesity. Heritable factors are estimated to explain 40-70% and 10% of the inter-individual variance in body weight and EE, respectively. The aim of this study was to identify genetic variants that negatively affect EE and thereby positively influence Body Mass Index (BMI) in American Indians, an ethnic group with a high prevalence of obesity. 521,709 tag SNPs (minor allele frequency≥5%, r2≥0. 85) derived from whole-genome sequence data of 335 Pima Indians (40x coverage) were genotyped using an Affymetrix Axiom Custom Array in a population-based sample of 3494 full-heritage Pima Indians. Data were analyzed for associations with maximum lifetime BMI (n=2762) and two separate measures of EE when subjects were non-diabetic: resting metabolic rate (RMR) after overnight fasting measured by a ventilated hood system (n=496) and 24h EE measured by a whole-room calorimeter (n=406). Results were adjusted for age, sex, body composition, heritage, family membership and genomic control. Rs11014566 (A/G), which maps to an intron in *GPR158*, demonstrated one of the strongest associations for EE measures (both p<10-2), where the G allele (frequency=0. 60) was associated with higher maximum lifetime BMI (b=+1. 6% per copy, p=2x10-2) and lower EE as assessed by both RMR (b=−39 kcal/day, p=3x10-3) and 24h EE (b=−37 kcal/day, p=7x10-4). To assess replication of the BMI association, rs11014566 was further genotyped in 3950 mixed-heritage American Indians, and it was associated with BMI in this replication sample (b=+2. 0% per copy, p=2x10-3) and in the combined data (b=+1. 8%, p=5x10-5). *GPR158* encodes the G protein-coupled receptor 158 that binds to an N-type voltage-gated calcium channel (*CACNA1B*) in the brain. Cacna1b-deficient mice are protected from obesity. Rs11014566 tags 3 other SNPs, rs144895904, rs34673593, and rs16925884 (r2=0. 86-0. 99) in intron 4, any of which could potentially affect transcriptional regulation. *In vitro* luciferase assays are ongoing to functionally characterize these SNPs. Our results suggest that common variation in *GPR158* may influence EE and obesity in American Indians. Identification of novel genes/gene pathways that influence EE and BMI in humans will lead to a better understanding of the complex pathophysiology of obesity and could potentially lead to new drug targets.

937F

Genetic Variants assocaiated with Facial Shape across Continents. *L. Qiao[1], Ch. Liu[1], Y. J. Yang[2], L. Jin[2], K. Tang[1].* 1) CAS-MPG Partner Institute for Computational Biology, SIBS, Shanghai, China; 2) State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China.

It's a long standing question why people from different ethnic backgrounds look rather different on face. A recent work from our group indicated that there exists strong phenotypic divergence between Europeans (EU) and East Asians (EA), however the genetic basis of such divergence is unknown. In this study, we carried out a genome-wide association study (GWAS) to look for genetic loci responsible for facial morphological differences between Europeans and Han Chinese. Around 1000 Uyghurs from Xinjiang, China were sampled for the stage I GWAS, as Uyghur is a nearly half-way admixed of both Caucasian and East Asian ancestries. Facial features at various biometric spaces were examined, ranging from landmark geometrics to dense shape geometrics. The traits showing strong divergence between EU and EA were queried for signals of association in Uyghurs. A number of candidate loci were identified with GWAS significance. Interestingly, these loci do seem to affect various facial features, including mouth, nose, eyes and cheek bones along an EU-EA trend. Several candidate SNPs were validated in an independent Uyghur sample and a Han Chinese sample, and showed consistent effects on faces. In the upcoming analysis, we are trying to obtain a panel of SNP markers that together can explain a substantial portion of the EU-EA facial divergence.

938W

Genome-wide association study of complication of diabetes mellitus (DM) in Korean populations. *SY. Rhee[1], YA. Shin[2], HM. Kim[2], DW. Lee[2], S. Chon[1], KM. Park[2], HJ. Jung[2], JT. Woo[1].* 1) Kyung Hee University School of Medicine, Seoul, Seoul, South Korea; 2) Theragen-Etex Bio Institute, Advanced Institute of Convergence Technology, Kwanggyo Technovalley, Korea.

Diabetes mellitus (DM) is the most prevalent metabolic diseaseand the leading cause of death worldwide. DM is associated with various complications (retinopathy, nephropathy, coronary artery disease and so on). Multiple genetic factors and environmental exposures have been shown to affect pathogenesis of DM together, and therefore it is very important to find related factors cause for DM complications. Genetic researches have been found several genetic variants relevant to DM complications, but in case of genetic factors for DM retinopathy have not been well-reported in any populations. In this study, we carried out GWA analysis to identify genetic factors influencing DM retinopathy with Korean hospital cohorts. We have collected DM retinopathy diagnosis data from hospital populations through Korea National Diabetes Program (KNDP) during about last 10 years (from 2006 to 2015). The subjects with DM retinopathy were genotyped using the Axiom™ Genome-Wide ASI 1 array comprised of 598,100 SNPs. Association analysis was preformed and built logistic regression model (1. d. f.) using 506,945 SNPs fulfilled the quality control criteria of GWA stages. We are expected to recruit replication cohorts from hospital through KNDP program for further replication analysis. Meta-analysis will be performed for combining statistical values of the GWAS and a replication stage to identify loci associated with DM retinopathy.

939T

Genome-wide association studies of vitiligo implicate 100 loci in disease risk. *S. Santorico[1,2], Y. Jin[2,3], D. Yorgov[1], T. Ferrara[2], R. A. Spritz[2,3].* 1) Mathematical and Statistical Sciences, University of Colorado, Denver, USA; 2) Human Medical Genetics and Genomics Program, University of Colorado School of Medicine, Aurora, Colorado, USA; 3) Department of Pediatrics, University of Colorado School of Medicine, Aurora, Colorado, USA.

Generalized vitiligo (GV) is an autoimmune disease in which white patches of skin and hair result from destruction of melanocytes. Striking skin depigmentation in GV particularly impacts persons of color, with frequent social isolation and psychiatric co-morbidity. In addition, GV patients have ~20% risk of other autoimmune diseases, pointing to shared genetic risk factors and perhaps environmental triggers. In previous work, we discovered and replicated 26 GV loci via two GWAS and 2 via candidate gene studies. Despite this success, only 18% of risk heritability is explained by these loci. We have just completed a third GWAS and combined analysis over all GWAS. After imputation using the 1KGP Phase I reference panel, QC, and ancestry matching, ~660,000 genotyped and ~20,000,000 imputed SNPs were compared in 2853 European-derived white (EUR) cases and 37,412 EUR controls, providing power for common variants to OR 1. 2. The combined analysis yielded 72 novel loci, several with multiple independent association signals, which are undergoing replication in an independent set of 2138 EUR cases and 2262 controls. Of these loci, ~45% encode proteins involved in immune/inflammatory regulation, ~10% apoptosis/autophagy, ~10% encode melanocyte proteins likely involved in GV triggering/targeting, and ~35% encode proteins of unknown function, functions with no evident relationship to GV, or gene deserts. About half of the immunoregulatory loci are shared with other autoimmune diseases with which GV is often comorbid. Many of the immune/inflammatory and apoptosis/autophagy proteins reside in biological pathways of immune triggering, inflammatory response, T-cell activation, and cell killing that may mediate autoimmune triggering by and killing of cells that express the cognate melanocyte protein antigens. For several loci we have already identified causal variation, both coding and regulatory, and for some (*TYR*, *HLA-A*, *HLA-DQA1*/*DRB1*, *NLRP1, SLA*) we have already or are now carrying out functional analyses to determine specific mechanisms of disease causation and functional interaction. We plan to assess heritability explained by all loci achieving replication, which will then be compared to a polygenic risk model, with the long-term goal of optimized prediction of vitiligo risks and subtypes to facilitate clinical application of optimal therapies based on genetic subtyping of disease.

940F

Common Variants in PTGES Associated with Asthma Susceptibility in African Americans. *P. Sleiman[1,2], B. Almoguera[1], J. Connolly[1], F. Mentch[1], L. Vazquez[1], E. Hysinger[1], H. Hakonarson[1,2], The electronic Medical Records and Genomics (eMERGE) Network.* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, PA, USA; 2) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, PA, USA.

Asthma is one of the most common chronic diseases in the United States affecting more than 22 million adults and 6 million children. It is characterized by relapsing and remitting symptoms, reversible airway hyper-responsiveness, and chronic inflammation. We performed a meta-analysis of site and ancestry-stratified GWAS to identify variants associated with asthma susceptibility in subjects from the electronic Medical Records and Genomics (eMERGE) Network. Subjects from the eMERGE Network represent a multi-ethnic cohort linked to electronic medical records for phenotype mining from nine participating sites in the US. All cases and controls were defined by an electronic health record (EHR)-based algorithm, which mines health EHRs for pertinent diagnostic (i. e. ICD9 codes) and medication information. The algorithm was validated by manual chart review both internally at CHOP and at two external sites within eMERGE. Positive predictive values were estimated at 96% for the cases and 98% for controls. The meta-analysis included 6,696 cases and 20,114 controls from eight participating eMERGE sites. We identified two genome wide significant loci, the previously reported ORMDL3 locus on chr17q12 in the European ancestry cases and a novel locus on chr9q34 encompassing the prostaglandin E synthase (PTGES) gene in the African American ancestry analysis. PTGES encodes the critical terminal synthetic enzyme required for the processing of COX-2–derived PGH2 to PGE2. Data from animal models and studies of human patients implicate PGE2 in the pathogenesis of asthma. In models of allergic lung disease PGE2 has been shown to be bronchoprotective and anti-inflammatory. In patients with atopic asthma, inhalation of PGE2 blocks bronchoconstriction in response to challenge with inhaled specific allergen and attenuates the recruitment of eosinophils and basophils to the airway. Reduced PGE2 levels also correlate with airway hyperresponsiveness (AHR), a cardinal feature of asthma. Aspirin-exacerbated respiratory disease (AERD), a severe variant of asthma, is associated with deficient COX-2–derived PGE2 production as well as marked tissue eosinophilia and bronchoconstrictive responses to the administration of nonselective COX inhibitors. Here we report the association of common PTGES variants with asthma susceptibility. PTGES, as the dominant PGE2 synthase, may explain the observed PGE2 involvement in asthma and AERD and serve as a potential therapeutic target for both disorders.

941W

Genome-wide association study of urinary electrolytes-adjusted blood pressure. *B. Tayo[1], H. Kramer[1], G. Cao[1], A. Luke[1], X. Zhu[2], R. Cooper[1].* 1) Department of Public Health Sciences, Loyola University Chicago Stritch School of Medicine, Maywood, IL; 2) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.

Introduction: Genome-wide association studies have led to identification and validation of about 40 susceptibility loci for blood pressure (BP) and hypertension especially among individuals of European ancestry. Even though these genetic variants collectively explain only a small fraction of the heritability for blood pressure phenotypes, similar associations with blood pressure phenotypes remain to be demonstrated in individuals of African ancestry. **Goal:** The aim of our study was to identify genetic variants associated with variations in urinary electrolytes-adjusted blood pressure among individuals of African origin. **Methods:** We obtained mean daily urinary sodium and potassium of three 24-hour urine samples collected from 691 unrelated adult African Americans from Maywood, IL. The study subjects consisted of 251 females and 440 males. All subjects were genotyped using the Affymetrix Genome-Wide Human SNP Array 6. 0. Analyses included genotypes on 785268 autosomal SNPs that passed standard quality control procedures. We tested for associations with systolic and diastolic BP controlling for mean urinary sodium and potassium under additive genetic mode of inheritance using linear regression models. The models also included age, sex, body mass index and age-by-sex interaction as covariates. Population stratification was controlled for by inclusion of principal components in the models. Similarly, we tested for genetic associations with urinary electrolytes. **Results & conclusion:** The top most associated SNPs were found on chromosomes 10 (rs11248449, *p*=2. 49 x 10-6), 8 (rs6586892, *p*=3. 57 x 10-7), 14 (rs4377089, *p*=3. 69 x 10-6) and 1 (rs638859, *p*=1. 70 x 10-6) for diastolic BP, systolic BP, urinary sodium and urinary potassium, respectively. These data represent the first genome-wide association study of electrolytes-adjusted BP in an African ancestry population and also provide preliminary evidence on genetic associations with urinary electrolytes and electrolytes-adjusted BP; additional studies to confirm these findings are required.

942T

A scalable Bayesian method for integrating functional SNP annotations in genome-wide association studies. *J. Yang, X. Zhou, G. Abecasis.* Department of Biostatistics, University of Michigan, Ann Arbor, MI.

Although genome-wide association studies (GWASs) have identified many SNPs associated with complex traits and common diseases, the biological mechanisms underlying these associations are largely unknown. Here, we propose a Bayesian variable selection model to integrate variant functional annotations, and help understand and prioritize causal variants and mechanisms. Our method improves upon previous approaches by accounting for multiple categories of functional annotations, for genotype correlation due to linkage disequilibrium (LD) and, importantly, by quantifying the proportion of causal variants and relative effect sizes of variants with different functional annotation. To apply our model to very large GWAS and sequencing data sets, we present a novel scalable Bayesian computation method through a block-wise expectation maximization Markov Chain Monte Carlo (EM-MCMC) algorithm. Our algorithm dramatically improves both computational speed and posterior sampling convergence by taking advantage of the block-like LD structure of the human genome. In simulations, we show that our method increases power and identifies more true signals compared with competing methods. In real data, we show that previous greedy approaches and MCMC implementations lead to apparently sub-optimal sets of likely causal variants because they fail to fully explore the set of possible causal variants. We applied our method to a genomewide association study of age-related macular degeneration with ~33 thousand individuals and >12 million genotyped and imputed variants. Our results show that the non-synonymous markers are about 20 times more likely to be causal than the other markers, and that the effect size of associated non-synonymous variants is about 3 times larger than for other variants. Importantly, our method can help prioritize likely functional candidates for follow-up while disentangling the effects of genotype, linkage disequilibrium and functional annotation. In conclusion, our method has the potential to shed light on the biological mechanism of SNP associations and can help prioritize SNPs for downstream analysis.

943F

GWAS of 25(OH) Vitamin D concentrations in Punjabi Sikhs: Results of the Asian Indian Diabetic Heart Study. *B. R. Sapkota[1], R. Hopkins[1], A. Bjonnes[6], S. Ralhan[3], G. S. Wander[3], N. K. Mehra[4], J. R. Singh[5], P. R. Blackett[2], R. Saxena[6], D. K. Sanghera[1].* 1) Department of Pediatrics, Section of Genetics, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK; 2) Department of Pediatrics, Section of Endocrinology, College of Medicine, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA; 3) Hero DMC Heart Institute, Ludhiana, India; 4) All India Institute of Medical Sciences and Research, New Delhi, India; 5) Central University of Punjab, Bathinda, Punjab, India; 6) Broad Institute of Massachusetts Institute of Technology and Harvard, Massachusetts General Hospital, Cambridge, MA, USA.

Vitamin D deficiency is implicated in multiple disease conditions and accumulating evidence supports that the variation in serum vitamin D (25(OH)D) levels, including deficiency, is under strong genetic control. However, the underlying genetic mechanism associated with vitamin 25(OH)D concentrations is poorly understood. We earlier reported a very high prevalence of vitamin D deficiency associated with an increased risk for type 2 diabetes and obesity in a Punjabi Sikh diabetic cohort as part of the Asian Indian Diabetic Heart Study (AIDHS). Here we have performed the genome-wide association study (GWAS) of serum 25(OH) D on 3,538 individuals from this Punjabi Sikhs population. Our discovery GWAS comprised of 1,387 subjects followed by validation of 24 putative SNPs (P<10-4) using an independent replication sample (n=2,151) from the same population by direct genotyping. A novel locus at chromosome 20p11. 21 [b = -0. 13, p=4. 47x10-9] was identified to be associated with 25(OH)D concentrations. Another suggestive association signal [b = 0. 90; p=1. 36x10-6] was found within the regulatory region of the *IVL* gene on chromosome 1q21. 3. Additionally, our study replicated 3 of 5 known GWAS genes associated with 25(OH)D concentrations including *GC* (p=0. 007) and *CYP2R1* (p=0. 019) reported in Europeans and the *DAB1* (p=0. 003), reported in Hispanics. Future functional studies in biologically plausible regions associated with 25(OH)D metabolism will provide new molecular insights on genetic drivers of vitamin D status and its implications in health disparities. This study was supported by NIH grants -R01DK082766 (NIDDK) and NOT-HG-11-009 (NHGRI), and VPR Bridge Grant (OUHSC).

944W

A Genome-Wide Association Study of Multi-Comorbidities: Towards a Genomic-Decision Aid for Health and Wellness Forecasting. *K. Shameer[1,2], D. Ruderfer[3], E. Austin[4], R. Miotto[1,2], M. A. Badgeley[1], I. J. Kullo[3], J. T. Dudley[1,2,5].* 1) Department of Genetics and Genomics , Icahn Insititute of Genomics and Multiscale Biology, New YorK, NY. , USA;; 2) Harris Center for Precision Wellness, Mount Sinai Health System, New York, NY 10029, NY; 3) Division of Psychiatric Genomics, Icahn Institute of Genomics and Multiscale Biology, New York, NY 10029; 4) Division of Cardiovascular Disease, Mayo Clinic, Rochester, MN 55905; 5) Population Health Science and Policy, Mount Sinai Health System, New York, NY 10029, NY.

Genome-wide and phenome-wide association studies have expanded our understanding of disease phenotypes and associated genetic variants. No disease presentation is singular in clinical setting and disease manifestations often have pair-wise or higher degree of disease comorbidities. Leveraging genome-wide association studies could aid in the understanding of genetic variants influencing multi-comorbidity. We performed a genome-wide association using Charlson Comorbidity Index (CCI), that aggregates 22 comorbid conditionsas a quantitative trait on a cohort of 15, 335 individuals with electronic health records at the Icahn School of Medicine at Mount Sinai. We replicated the significant associations in the EHR-linked Mayo Vascular Disease Biorepository (n=7902). A phenotyping algorithm was developed based on CCI definitions using three different version count model, index severity weight and age and severity weighted version. Genetic association analyses were performed using PLINK. Four SNPs reached genome-wide significance in the discovery cohort (p<=5x10-8): rs7300363 (*STRAP, EPS8, DERA*), rs115379330 (*TK2, CMTM1, CMTM3, CMTM4, CMTM2, CKLF, CKLF-CMTM1, BEAN1*), rs73388073 (*KCNJ2-CALM2P1*) and rs58231812 (*EPSTI1*) but only rs58231812 (P= 6. 01x10-5) was replicated in the validation cohort. Collectively, the genes mapped to the SNPs are involved in a variety of biological functions (cytokine activity and deoxyribonucleotide metabolic process), pathways (pentose phosphate pathway, pyrimidine metabolism and drug metabolism), tissues (enriched among pituitary gland, parathyroid gland, adipose tissue and juvenile development) and phenotypes (QT interval, tooth development, cardiovascular, neurological, morphological and physiological phenotypes) related to organismal development and normal lifespan. Replicated SNPs in the intronic region of *EPSTI1,* highly expressed in bone marrow and differentially expressed in the setting of various infectious diseases. We plan to replicate our findings in additional samples from various EHR-linked networks and in the process of phenotyping comorbidity phenotypes such as Cumulative Illness Rating Scale (CIRS), Index of Coexisting Disease (INDEX) and Kaplan-Feinstein index (KFI) for comparative analyses. Investigation of genetic variants influencing disease comorbidity will aid in understanding genetic predisposition for disease susceptibility and help in managing complex diseases, wellness and health of patient populations.

945T

Meta-analysis of exome chip variants identifies common and rare variant associations for white blood cell counts in more than 132,000 participants. *U. Schick[1,2], S. M. Tajuddin[3], P. L. Auer[4], G. Lettre[5], A. P. Reiner[6,7] on behalf of the Blood Cell Consortium.* 1) Icahn School of Medicine at Mount Sinai, New York, NY; 2) The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, NY; 3) Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD; 4) Joseph J. Zilber School of Public Health, University of Wisconsin-Milwaukee, Milwaukee, WI; 5) Montreal Heart Institute and Université de Montréal, Montreal, Quebec H1T 1C8 Canada; 6) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 7) Department of Epidemiology, University of Washington, Seattle, WA.

Total and differential white blood cell counts (WBCC) are clinically important measures that are used to monitor health and disease states. WBCC are heritable traits and genomic association studies have identified more than thirty trait-associated loci. To further characterize genomic variation across the allelic frequency spectrum influencing WBCC traits, we genotyped ~250,000 variants in 18 studies comprising up to 132,764 individuals of European (Eu), African (Af), Hispanic and Asian descent with total and differential (neutrophil, eosinophil, monocyte, basophil, lymphocyte) WBCC. We carried out race-specific (Eu or Af) and race-combined (All) meta-analyses of exome chip variants applying single variant and gene-level tests (VT, SKAT, GRANVIL) in rareMETALS. Thresholds for significance were assessed using Bonferroni correction (single variant P<2x10-7 & gene-level P<3. 1x10-6). We confirmed many common variant associations (e. g. *SH2B3-ATXN2, DARC, ABO)* and other loci harboring rare variants (e. g. *CXCR2, LCT, JAK2*) for WBCC traits. Further, we identified 3 novel rare variant associations and more than a dozen biologically plausible common variant loci. Gene-level tests identified a novel gene *TBX3,* a transcription factor known to be aberrantly methylated in chronic lymphocytic leukemia, for lymphocyte count (GRANVIL P(All)=1. 4x10-6), and two novel genes for neutrophil count (*ZNF439*, GRANVIL P(Af)=7. 0x10-7 & *ORC4C6*, SKAT P(Eu)=2. 6x10-8). Novel common variant associations point to genes with known mutations in myeloid malignancies including *ASXL1* (rs2295764, P(Eu)= 3. 3x10-9) for basophil count and *CD33* (rs3865444 & rs35112940, P(Eu)<1. 6x10-8) for total WBCC. Also, we identified noncoding SNPs in close proximity to regulators of hematopoiesis, specifically *CEBPA (*rs736289, P(Eu)=7. 0x10-9) for basophil count, *CEBPE*-proximal *SLC7A8* (rs11625112, P(Eu)=3. 8x10-8) for monocyte count and *ZMIZ2* (rs1050331, (All)=1. 3x10-8) for total WBCC. Other novel SNPs overlap with known platelet loci including *JMJD1C* noncoding SNPs for total WBCC (rs1935 & rs12355784, P(All|Eu)<3. 6x10-8) and intergenic SNPs near *SIRPA* for total WBCC and lymphocyte count (rs6136489 & rs6045676, P(All|Af)<3. 3x10-8). Many of these SNPs were significant *cis*-eQTL in whole blood for proximate genes (e. g. *ASXL1*, *CD33*, *CEBPA*, *SIRPA)* in the blood eQTL browser. We identify several novel rare and common variant associations for WBCC and efforts are ongoing to carry out replication of these findings.

946F

Genome-wide interaction with insulin secretion risk score reveals novel loci contributing to type 2 diabetes etiology in African Americans. *J. M. Keaton[1,2,3], J. N. Hellwege[2,3], M. C. Y. Ng[2,3], N. D. Palmer[2,3,4], J. S. Pankow[5], M. Fornage[6], J. G. Wilson[7], A. Correa[7], L. J. Rasmussen-Torvik[8], J. I. Rotter[9], S. S. Rich[10], L. E. Wagenknecht[11], B. I. Freedman[12], D. W. Bowden[2,3,4].* 1) Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC; 3) Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC; 4) Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC; 5) Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, MN; 6) University of Texas Health Science Center at Houston, Houston, TX; 7) University of Mississippi Medical Center, Jackson, MS; 8) Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 9) Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute, Harbor-UCLA Medical Center, Torrance, CA; 10) Center for Public Health Genomics, University of Virginia, Charlottesville, VA; 11) Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC; 12) Department of Internal Medicine - Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, NC.

Type 2 diabetes (T2D) is the result of metabolic defects in insulin secretion and insulin sensitivity, yet most of the ~80 genes associated with T2D risk involve insulin secretion/beta cell biology. We hypothesized that T2D risk includes insulin sensitivity loci that are defined by interaction with insulin secretion loci. To test this hypothesis, SNPs associated with acute insulin response to glucose (AIR), a measure of first-phase insulin secretion, in African Americans from the IRAS Family Study (IRASFS; n=492 subjects) were used to build genetic risk scores (GRS) for T2D and tested for genome-wide interaction in the ARIC, CARDIA, JHS, MESA, and WFSM cohorts (n=2725 cases, 4167 controls). A GRS from 1) novel AIR loci (AIR GRS; AIR P<5x10-4; n=5 SNPs) and 2) a GRS from T2D GWAS SNPs trending towards association with AIR (T2D-IS GRS; n=5) were developed. Unweighted GRS (summation of risk alleles) and weighted GRS (sum of risk alleles at each locus multiplied by their effect sizes) were determined. Each GRS was tested for association with AIR and disposition index (DI), a combinatorial measure of first-phase insulin secretion and insulin sensitivity, in IRASFS. The weighted AIR GRS was associated with AIR in IRASFS (P=4. 96x10-2) and this association improved with BMI adjustment (P=3. 34x10-2). The weighted T2D-IS GRS was associated with DI (P=4. 43x10-2), and this association was robust against BMI adjustment (P=4. 51x10-2). GRS were tested for genome-wide interaction using logistic regression modeling T2D as the outcome followed by meta-analysis of interaction effects. Nominally significant (P<5x10-6) interactions with the AIR GRS were observed at several loci including a mitochondrial ribosomal gene (*MRPL36*, rs4975846, P=2. 03x10-7), a post-transcriptional regulator (*SAMD4A*, rs11627203, P=6. 57x10-7), and a neuromuscular gene (*UTRN*, rs17074194, P=1. 87x10-6). Nominally significant interactions with the T2D-IS GRS included two SNPs at the *DGKB* locus (rs6976381, P=1. 21x10-6; rs6962498, P=3. 71x10-6), a diacylglycerol kinase associated with T2D and fasting glucose levels in published GWAS. The most associated SNPs in both the AIR GRS and T2D-IS GRS interaction analyses were robust against BMI adjustment. These data support the hypothesis that the spectrum of genetic factors contributing to T2D risk includes interactions with insulin secretion loci, and that epistasis modeling is a powerful approach for the detection of candidate insulin resistance genes.

947W

Amino acid variation in HLA class II proteins is a major determinant of humoral response to common viruses. *C. Hammer[1,2,3], M. Begemann[3], P. J. McLaren[1,2], I. Bartha[1,2], A. Michel[4], B. Klose[5], C. Schmitt[5], T. Waterboer[4], M. Pawlita[4], T. F. Schulz[5], H. Ehrenreich[3,6], J. Fellay[1,2].* 1) School of Life Sciences, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; 2) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Göttingen, Germany; 4) Division of Molecular Diagnostics of Oncogenic Infections, Infections and Cancer Program, German Cancer Research Center, Heidelberg, Germany; 5) Institute of Virology, Hannover Medical School, Hannover, Germany; 6) DFG Research Center for Nanoscale Microscopy & Molecular Physiology of the Brain (CNMPB), Göttingen, Germany.

The magnitude of the human antibody response to viral antigens is highly variable. To explore the impact of human genetic variation on humoral immunity, we performed genome-wide association studies of immunoglobulin G (IgG) response to 14 human pathogenic viruses in 2,363 immunocompetent adults. Significant associations were observed in the major histocompatibility complex (MHC) region on chromosome 6 for influenza A virus, Epstein-Barr virus (EBV), JC polyomavirus (JCPyV) and Merkel Cell polyomavirus (MCPyV) (*P*min = 7 x 10-26). Using local imputation and fine mapping, we identified specific amino acid residues of HLA class II proteins as the most probable causal variants. The strongest associations were observed for amino acid positions in the HLA-DRβ1 protein. We therefore analyzed the potential impact of common HLA-DRβ1 haplotypes on IgG response to influenza A, EBV, JCPyV and MCPyV. This analysis showed no consistency in the associations: the size and directionality of the effects were virus-specific. Most prominently, the amino acid haplotypes present in the classical alleles *HLA-DRB1*15:01* or **16:01* associated with influenza A seropositivity and with higher anti-EBV IgG levels, but with JCPyV and MCPyV seronegativity. In contrast, the haplotype present in *HLA-DRB1*01:01* associated with MCPyV seropositivity and with higher anti-EBV IgG levels, but with influenza A seronegativity. Variable amino acid residues explained between 2. 5% (JCPyV serostatus) and 5. 4% (EBV EBNA IgG levels) of the phenotypic variance. We also compared our results to reported HLA associations with autoimmune disorders. Striking similarities were observed in the genetic underpinning of IgG response to influenza A and narcolepsy, and of antibodies against Epstein-Barr virus and multiple sclerosis. The results of this study emphasize that HLA class II variation plays a central and pathogen-specific role in the modulation of humoral immune response to viral infection.

948T

Integrative analysis of a GWAS for amino acids and acylcarnitines in whole blood and gene-expression data identifies six novel loci and reveals insight into regulatory mechanisms. *M. Scholz[1,2], R. Burkhardt[1,3], H. Kirsten[1,2,4], F. Beutner[1,5], L. Holdt[1,6], A. Gross[1,2], A. Teren[1,5], A. Toenjes[7], S. Becker[1,3], K. Krohn[1,8], P. Kovacs[9], M. Stumvoll[7,9], D. Teupser[1,6], J. Thiery[1,3], U. Ceglarek[1,3].* 1) LIFE Leipzig Research Center for Civilization Diseases, University of Leipzig, Leipzig Germany; 2) Institute for Medical Informatics, Statistics and Epidemiology, University of Leipzig, Leipzig, Germany; 3) Institute of Laboratory Medicine, Clinical Chemistry and Molecular Diagnostics, University Hospital Leipzig, Leipzig, Germany; 4) Department for Cell Therapy, Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany; 5) Heart Center Leipzig, Leipzig, Germany; 6) Institute for Laboratory Medicine, Ludwig-Maximilians University Munich, Munich, Germany; 7) Medical Department, Clinic for Endocrinology and Nephrology, University of Leipzig, Leipzig, Germany; 8) Interdisciplinary Centre for Clinical Research, University of Leipzig, Leipzig, Germany; 9) Integrated Research and Treatment Center Adiposity Diseases, University of Leipzig, Leipzig Germany.

Background and statement of purpose: Emerging data suggests that altered blood levels of amino acids and acylcarnitines are associated with common metabolic diseases in adults. Therefore, our aim was to identify common genetic determinants for blood metabolites to better understand pathways contributing to human physiology and common diseases. **Methods:**We applied a targeted mass-spectrometry-based method to analyze whole blood spots for 96 amino acids, acylcarnitines and metabolite ratios in a cohort of 2,107 adults. Using chip-based technology and HapMap-based imputation, we performed genome-wide association (GWA) to identify genetic modifiers of metabolite concentrations. Discovered hits were analyzed for replication in an independent cohort. Furthermore, we used Illumina HT-12 v4 Expression BeadChips to study gene expression in peripheral mononuclear cells in order to analyze causal relations between SNPs, gene-expression and metabolite levels. **Results:**We discovered and replicated six novel loci associated with blood levels of total acylcarnitine, arginine , propionylcarnitine, 2-hydroxyisovalerylcarnitine, stearoylcarnitine, and aspartic acid traits. Several SNPs associated with blood metabolites in our study overlap with previously identified loci for human diseases suggesting a shared genetic basis or pathomechanisms involving metabolic alterations. Causal analyses resulted in following putative causative genes: *SLC22A16* for total acylcarnitines, *ARG1* for arginine, *HLCS* for 2-hydroxyisovalerylcarnitine, *JAM3* for stearoylcarnitine via a trans-effect at chromosome 1, and *PPP1R16A* for aspartic acid traits. Additionally, we report replication and provide additional functional evidence for ten loci that have previously been published for metabolites in several tissues. **Conclusion:**The integrative analysis of SNP, gene-expression and metabolite data revealed novel insights into the genetic regulations of human metabolism. At several examples, we provide evidence for metabolite regulation via gene-expression and observed overlaps with GWAS loci for common diseases. In consequence, , our findings provide strong candidates for future functional studies directed to understand human metabolism and pathogenesis of related diseases.

949F

Genetic association in narcolepsy suggests autoimmune origin. *HM. Ollila[1], R. Hillary[1], J. Faraco[1], L. Ling[1], J. Hallmayer[2], F. Han[3], E. Mignot[1].* 1) Stanford University, Palo Alto, CA. , USA; 2) Department of Psychiatry & Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA; 3) Department of Surgery, Peking University People's Hospital, Beijing, China.

IntroductionType 1 Narcolepsy is characterized by sleepiness, REM sleep abnormalities and loss of muscle tone triggered by positive emotions (cataplexy). The cause of type 1 narcolepsy is a loss of neurons producing the hypocretin/orexin peptide of likely autoimmune origin. Narcolepsy is strongly associated with Human leukocyte antigen (HLA) DQB1*06:02, with 98 percent of narcoleptics carry this allele. Our aim was to further study genetic predisposition to narcolepsy and to determine if other HLA alleles also affect narcolepsy predisposition. **Methods**We performed GWAS in Asian, African American and Caucasian samples with 6,000 cases and 30,000 controls. In addition, direct (subsample) and imputed HLA genotyping at high resolution (8-digit) was performed using HIBAG. Case control marching and conditional analysis were performed. **Findings and conclusions**In addition to well-known DQB1 effects, HLA-DPA1*01:03~DPB1*04:02 was highly protective, while other alleles at DPB1 and HLA-Class I increased susceptibility. In addition, we found significant SNP associations with age of onset and with comparison before or after 2009 pandemic H1N1 influenza. Additional genome-wide significant associations were seen with variants in TCRA, TCRB, P2RY11, ZNF365, IL10RB-IFNAR1 and other genes known to be involved in autoimmunity. Our results highlight the importance of HLA in narcolepsy, and together with functional studies, support the autoimmune origins of narcolepsy.

950W

Preliminary Genetic Association Analysis of Obstructive Sleep Apnea (OSA) and Obesity in Patients from Colombia. *L. A. Morford[1], J. K. Hartsfield, Jr. [1], P. Hidalgo[2], L. Wilches-Buitrago[3], C. A. Morillo[4], L. Otero[3].* 1) Dept Oral Health Science, Center for the Biologic Basis of Oral/Systemic Diseases, Hereditary Genetics/Genomics Core, Univ of Kentucky, Lexington, KY USA; 2) Pontificia Universidad Javeriana - Sleep Clinic, Hospital Universitario San Ignacio. Bogotá, Colombia; 3) Pontificia Universidad Javeriana. Bogotá, Colombia; 4) McMaster University. Hamilton, Ontario, Canada.

Obstructive Sleep apnea (OSA) is a highly prevalent breathing disorder that is frequently associated with obesity. In this pilot-level Genome-wide Association Study (GWAS) of 542,585 Single Nucleotide Polymorphisms (SNPs), we sought to identify genetic markers associated with the combined phenotype of **BOTH** OSA and obesity. Ethics approval for the study was obtained from the ethical committee of Faculty of Medicine of Pontificia Universidad Javeriana and Hospital Universitario San Ignacio. DNA obtained from 58 Colombian subjects (29 males and 29 females; 40 to 75 years old) diagnosed with **BOTH** OSA and obesity was compared to the DNA obtained from 232 individuals (132 males and 100 females; 40 to 75 years old) who presented with **ONLY** one of the two phenotypes or neither phenotype. Patients were classified as having OSA by overnight polysomnography with an AHI ≥ 5 events/hour, and all individuals with a Body Mass Index (BMI) ≥ 30 kg/m2 were classified as obese. We excluded patients diagnosed with any mental and/or neurological diseases, as well as individuals with Central Sleep Apnea. The preliminary data showed a significant association with the combined phenotype of OSA plus obesity and SNP rs2292662 within the Ataxin-7 (ATXN7) gene on Chromosome 3 (p= 2. 70E-08; OR=4. 3). ATXN7 is a transcription factor that plays a role in histone acetylation and deubiquitination during the process of chromatin remodelling. Borderline associations were also independently detected between the combined phenotype and genetic markers near or within genes for p21 Protein-Activated Kinase-7 (PAK7) on Chromosome 20 (p= 1. 92E-07; OR= 34. 3), Ubiquitin Specific-Protease 25 (USP25) on Chromosome 21 (p=2. 28E-07; OR=19. 4) and Neuron Navigator-2 (NAV2) on Chromosome 11 (p=2. 02E-06; OR=3. 3). Although genes like ATXN7 and USP25 have not previously been associated with the duel phenotype of OSA and obesity, these preliminary results suggest that chromatin remodelling may play a role in duel disease generation and/or maintenance. Our ongoing and future investigations should help to better understand the complex gene interaction networks involved in obesity and other diseases associated to OSA. This research was funded by Colciencias. Grant 537, contract 369.

951T

Phenome Wide Association Study of Human MHC Region in Marshfield Clinic PMRP. *J. Liu[1], Z. Ye[2], J. Mayer[2], C. Green[3], L. Rolak[4], C. Cold[5], S. J. Hebbring[1].* 1) Center for Human Genetics, Marshfield Clinic Research Foundation, Marshfield, WI; 2) Biomedical Informatics Research Center, Marshfield Clinic Research Foundation, Marshfield, Wisconsin, USA; 3) Department of Dermatology, Marshfield Clinic, Marshfield, Wisconsin, USA; 4) Department of Neurology, Marshfield Clinic, Marshfield, Wisconsin, USA; 5) Department of Pathologist, Marshfield Clinic, Marshfield, Wisconsin, USA.

Over 160 disease phenotypes have been mapped to the major histocompatibility complex (MHC) region on chromosome 6 by GWAS, suggesting that the MHC region as a whole may be involved in the etiologies of many phenotypes, including unstudied diseases. The phenome-wide association study (PheWAS), a powerful and complementary approach to GWAS, has demonstrated its ability to discover and rediscover genetic associations. In the current study, we systematically explored the MHC region by associating 3067 variants with 6221 phenotypes by PheWAS in a cohort of 7481 subjects from Marshfield Clinic Personalize Medicine Research Projects (PMRP). Results showed that expected associations previously identified by GWAS could be identified by PheWAS (eg. psoriasis, ankylosing spondylitis, and type I diabetes). Importantly, novel associations with 13 diseases not previously assessed by GWAS (eg. Lichen Planus) were also identified and replicated in an independent population. Many of these associated diseases appear to be immune or immune related disorders. These results demonstrate that the PheWAS approach is a powerful and novel method to discover and rediscover SNP-disease associations, and further emphasize the importance of the MHC region in human health.

952F

Pervasive pleiotropy between psychiatric disorders and immune disorders revealed by integrative analysis of multiple GWAS. *Q. Wang[1,2,3], C. Yang[3,4,5], J. Gelernter[2,3,6,7], H. Zhao[1,4,7,8].* 1) Program in Computational Biology and Bioinformatics, Yale University, New Haven, Connecticut, USA; 2) Department of Psychiatry, Yale School of Medicine, New Haven, Connecticut, USA; 3) VA CT Healthcare Center, West Haven, Connecticut, USA; 4) Department of Biostatistics, Yale School of Public Health, New Haven, Con- necticut, USA; 5) Department of Mathematics, Hong Kong Baptist University, Hong Kong SAR; 6) Department of Neurobiology, Yale School of Medicine, New Haven, Connecti- cut, USA; 7) Department of Genetics, Yale School of Medicine, West Haven, Connecticut, USA; 8) VA Cooperative Studies Program Coordinating Center, West Haven, Con- necticut, USA.

Although some existing epidemiological observations and molecular experiments suggested that brain disorders in the realm of psychiatry may be influenced by immune dysregulation, the degree of genetic overlap between psychiatric disorders and immune disorders has not been well established. We investigated this issue by integrative analysis of genome-wide association studies of 18 complex human traits/diseases (five psychiatric disorders, seven immune disorders, and others) and multiple genome-wide annotation resources (Central nervous system genes, immune-related expression-quantitative trait loci (eQTL) and DNase I hypertensive sites from 98 cell-lines). We detected pleiotropy in 24 of the 35 psychiatric-immune disorder pairs. The strongest pleiotropy was observed for schizophrenia-rheumatoid arthritis with MHC region included in the analysis (p = 3. 9e-285), and schizophrenia-Crohns disease with MHC region excluded (p = 1. 1e-36). Significant enrichment (>1. 4 fold) of immune-related eQTL was observed in four psychiatric disorders. Genomic regions responsible for pleiotropy between psychiatric disorders and immune disorders were detected. The MHC region on chromosome 6 appears to be the most important with other regions, such as cytoband 1p13. 2, also playing significant roles in pleiotropy. We also found that most alleles shared between schizophrenia and Crohns disease have the same effect direction, with similar trend found for other disorder pairs, such as bipolar-Crohn's disease. Our results offer a novel birds-eye view of the genetic relationship and demonstrate strong evidence for pervasive pleiotropy between psychiatric disorders and immune disorders. Our findings might open new routes for prevention and treatment strategies for these disorders based on a new appreciation of the importance of immunological mechanisms in mediating risk of many psychiatric diseases.

953W

Common susceptibility variants are shared between schizophrenia and psoriasis in the Han Chinese population. *XY. Yin[1,2,3,4], NE. Wineinger[2], KC. Wilhelmsen[4], JJ. Liu[5], NJ. Schork[3], XJ. Zhang[1].* 1) Institute of Dermatology, Anhui Medical University, Hefei, Anhui, China; 2) The Scripps Translational Science Institute, La Jolla, USA; 3) Human Biology, J. Craig Venter Institute, La Jolla, USA; 4) Department of Genetics, and Renaissance Computing Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC 27517, USA; 5) Human Genetics, Genome Institute of Singapore, A*ASTAR, Singapore 138672, Singapore.

Background: Previous studies have shown that individuals with schizophrenia have a higher likelihood of developing psoriasis than a typical person. This suggests that there might be a shared genetic etiology between the two conditions. **Aims**: To characterize the potential shared genetic susceptibility between schizophrenia and psoriasis using genome-wide marker genotype data. **Methods**: We applied a marker-based co-heritability estimation procedure, polygenic score analysis, gene set enrichment test and a LASSO regression model on 1,139 psoriatic and 744 schizophrenic subjects, and 1,678 controls both genotyped on ~500,000 single nucleotide variants (SNVs). **Results**: In the combined cohort, we estimated that a large fraction of psoriasis and schizophrenia risk could be attributed to common variants (*h*2SNP =29%, s. e. = 5. 0%, P = 2. 00 ×10-8), with a co-heritability estimate between the traits of 21%. We found that five variants within HLA region reached genome wide significance when both conditions were analyzed together (P < 5. 00×10-8), which collectively conferred a significant risk effect (odds ratio of highest risk quartile=6. 03, P < 2. 00 ×10-16). Variants contributing most to the shared heritable component between psoriasis and schizophrenia were enriched in antigen processing and cell endoplasmic reticulum. **Conclusions**: We found evidences for a shared genetic etiology between schizophrenia and psoriasis. The mechanism for this shared genetic basis likely involves immune and calcium signaling pathways.

954T

Meta-analysis of exome array data identifies two novel regions associated with lung function. *V. E. Jackson[1,2], L. V. Wain[1,2], I. Sayers[3], I. P. Hall[3], M. D. Tobin[1,2], SpiroMeta Consortium.* 1) Departments of Health Sciences and Genetics, Adrian Building, University of Leicester, Leicester, UK; 2) National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, UK; 3) Division of Respiratory Medicine, University Hospital of Nottingham, Nottingham, UK.

Lung function measures are an important predictor of mortality and morbidity and are used in the diagnosis of a number of diseases, including chronic obstructive pulmonary disease (COPD). A number of large-scale genome-wide association studies (GWAS) have successfully identified single nucleotide polymorphisms (SNPs) influencing lung function in over 50 regions; however these so far identified regions only account for a small proportion of the estimated heritability. We utilised genotype data from the exome array, which predominately measures low frequency, coding variation, in attempt to uncover some of the so-called missing heritability. We carried out discovery meta-analyses of exome array data and three lung function measures: forced expiratory volume in one second (FEV1), forced vital capacity (FVC) and the ratio ofFEV1 to FVC (FEV1/FVC). These analyses included 23,751 individuals of European ancestry, from 11 studies, genotyped using the Illumina Human Exome BeadChip. Each trait was adjusted for age, sex, height and ancestry principal components and familial relationships, as appropriate, and stratified by ever smoking status. We have utilised single variant association analysis methods, traditionally employed in GWAS, along with gene-based methods, which test for the joint effect of several variants in a gene. We performed all analyses in ever smokers and never smokers separately, in addition to in all samples combined. We followed up the most significant single variant and gene associations (P<10-4) in up to 93,390 independent samples. Through our single variant association analyses, we identified signals within two novel regions: both associations were with intergenic SNPs located close to *LY86* (P=1. 1x10-13) and *FGF10* (P=1. 9x10-8). *LY86* may play a role in the immune response to bacterial lipopolysaccharide (LPS) and cytokine production. *FGF10* is involved in several biological processes, including branching morphogenesis of the lung. Bioinformatics follow-up of these SNPs are now underway, to identify potential mechanisms underlying these associations.

955F

Genome-wide association analysis of a custom Pima Indian genotyping array identifies a female specific effect of common variation near *TH*/*MIR4686* on risk for type 2 diabetes and reduced insulin secretion. *L. Baier, P. Chen, P. Piaggi, Y. Muller, A. Nair, S. Kobes, W. Knowler, C. Bogardus, R. Hanson.* PECRB, NIDDK/NIH, Phoenix, AZ.

To optimize capture of genetic information in American Indians, we designed an entirelycustom Affymetrix Axiom array based on whole genome sequence data from 335 Pima Indians. High quality genotypes were obtained on 548,206 variants which tagged (r2 within 300 kb windows) 92% of the 5M common variants (minor allele frequency ≥0. 05) detected in the Pima genomes. This array was used to genotype 3,637 full heritage Pimas (stage 1) who had participated in a longitudinal study of type 2 diabetes (T2D); stage 1 data analysis is complete. Recently, 4,060American Indians from the same longitudinal study, who are largely of mixed heritage,were genotyped with the array (stage 2), but analysis of stage 2 array data is ongoing. However, 408 variants associated with T2D in stage 1 were prioritized for expedited genotyping (Illumina BeadXpress) in the mixed heritage sample and analysis of these 408 variants iscomplete. Among the prioritized variants, rs11564707 (C/G; frequency of G= 0. 45 and 0. 40 in full and mixed heritage Pima Indians) provided the strongest association with T2D, where it independently associated in each sample (full heritage: *P*=8x10-7; OR[95%CI] per copy of G allele=1. 33[1. 19-1. 49] andmixed heritage: *P*=5x10-3; OR= 1. 22[1. 10-1. 40], adjusted for age, sex, birth year, family membership and ancestry). Combining both samples provided a genome-wide significant association with T2D(*P* = 2. 2x10-8, OR=1. 28[1. 18-1. 40]). However, there was a significant gender x genotype interaction (*P*=0. 001) for T2D such that this association was entirely driven by the 4,175 females (*P*=2. 5x10-9; OR=1. 41[1. 26-1. 58]); the association in 3,314 males wasnot signficant(*P*=0. 14; OR= 1. 11[0. 97-1. 27]). This female specific association was also observed for measures of early insulin secretion assessed in full heritage Pima Indians who were normal glucose tolerant. Both the 30 minute insulin levels during a 75g oral glucose tolerance test and the insulin response to a 25g intravenous glucose bolus were reduced in females (both *P*<0. 01) but not males. Rs11564707 maps near the microRNA 4686 and the 5'region of the *TH* gene, and tags 11 nearby variants with an r2>0. 99 (including a novel T>TTTTGTTTG insertion at chr11:2,204,322). *TH* encodes tyrosine hydroxylase and studies have suggested TH activity in pancreatic islets and in the paracrine control of glucose induced insulin secretion; however, future functional studies are required to establish the causative variant and affected gene.

956W

Analysis of sex-by-SNP interaction among geographically distinct cohorts reveals novel suggestive loci for IgA nephropathy. *N. Mladkova[1], Y. Li[1], F. Scolari[2], D. Cusi[3], J. Novak[4], B. Julian[5], J. Feehally[6,7], B. Stengel[8,11], H. Zhang[9], RP. Lifton[10], AG. Gharavi[1], K. Kiryluk[1].* 1) Department of Nephrology, Columbia University Medical Center, Columbia University, New York, NY. USA; 2) Division of Nephrology, Azienda Ospedaliera Spedali Civili of Brescia, Montichiari Hospital, University of Brescia, Brescia, Italy. [2] Department of Medical and Surgical Specialties, Radiological Sciences, University of Brescia, Brescia, Italy; 3) Renal Division, DMCO (Dipartimento di Medicina, Chirurgia e Odontoiatria), San Paolo Hospital, School of Medicine, University of Milan, Milan, Italy; 4) Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama, USA; 5) Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, USA; 6) John Walls Renal Unit, University Hospitals of Leicester, Leicester, UK; 7) Department of Infection, Immunity and Inflammation, University of Leicester, Leicester, UK; 8) INSERM, Centre for Research in Epidemiology and Population Health, Villejuif, France; 9) Renal Division, Peking University First Hospital, Peking University Institute of Nephrology, Beijing, China; 10) Department of Genetics, Yale University School of Medicine, New Haven, Connecticut, USA. [2] Howard Hughes Medical Institute, Yale University School of Medicine, New Haven, Connecticut, USA; 11) University Paris-Sud, Villejuif, France.

Background: IgA nephropathy (IgAN) represents the most common form of primary glomerulonephritis in the world. The disease is characterized by deposits of IgA1-containing immune complexes in kidney filtration units leading to kidney damage. IgAN is highly prevalent in East Asia, of intermediate prevalence in Europe and extremely rare in Africa. Although the disease is equally frequent in males and females in Asia, it is more common in European males, with a male-to-female ratio exceeding 3:1. This suggests a complex interaction between gender and ethnicity on disease risk. Methods: We performed genome-wide screen for sex-by-SNP interaction in previously describedbi-ethnic cohort of European ancestry (2017 cases and 8199 controls) and Asians (1194 cases and 902 controls). All individuals were typed with high-density SNP chips and imputed using ethnicity-matched 1000 Genomes reference panels (Minimac2). We tested for sex-by-SNP interaction using logistic regression under additive genotype coding. The results from individual cohorts were combined using fixed-effects meta-analysis (METAL). Results: The top interaction signal resides in the inter-genic region on chr. 13 (rs9595757, P≤3. 4e-07). The second suggestive locus resides on chr. 14 (rs11108444, P= 1. 0e-06, intronic variant in *ELK3*, a transcription factor involved in B-cell development and *IGH* gene regulation). The third locus resides within the Smith-Magenis syndrome region on chr. 17 (rs6502579, P= 2. 2e-06, intronic variant in *NT5M*). Another promising locus intersects *HS3ST3A1* on chr. 17 (rs8079179, P=5. 5e-06). The genetic variation in *HS3ST3A1* has previously been associated with Plasmodium falciparum parasitemia. Conclusions: This analysis of sex-by-SNP interaction in a large bi-ethnic cohort of IgAN identified several new suggestive loci. Many of those loci are located within genes involved in the immune system regulation and warrant further investigation in additional Asian and European cohorts.

957T

Do myopia genes vary with age? *J. W. L. Tideman[1,2], J. Guggenheim[3], Q. Fan[4], S. M. Saw[4,5,6,7], C. Williams[8], V. W. V. Jaddoe[2], J. R. Polling[1,9], C. C. W. Klaver[1,2], CREAM Consortium.* 1) Epidemiology, Erasmus Medical Center, Rotterdam, Zuid-Holland, Netherlands; 2) Department of Ophthalmology; Erasmus Medical Centre Rotterdam, The Netherlands; 3) School of Optometry & Vision Sciences, Cardiff University, Cardiff, UK; 4) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore; 5) Saw Swee Hock School of Public Health, National University Health Systems, National University of Singapore, Singapore, Singapore; 6) Department of Ophthalmology, National University Health Systems, National University of Singapore, Singapore; 7) DUKE-National University of Singapore Graduate Medical School, Singapore, Singapore; 8) School of Social and Community Medicine, University of Bristol, Bristol, UK; 9) Department of Orthoptics & Optometry, University of Applied Sciences, Faculty of Health, Utrecht, the Netherlands.

Purpose:GWA-studies performed in adults identified many genetic loci associated with refractive error and myopia. Knowledge of their age of onset in children may help understand the development of myopia, and identify a timeframe for specific therapies. We compared the effect of the individual risk variants on ocular biometry in three age categories (<10, 10 – 25 and >25 years). **Methods:** The study population consisted of three age-groups identified from the international CREAM consortium: 5,490 children aged <10 years; 5,000 children aged 10-25 years; and 13,274 adults >25 years. In these age-groups, we compared the effect sizes (betas) of the 39 top hits of the currently known genetic loci on the axial length/corneal radius (AL/CR) ratio, a good proxy for refractive error. The association between individual SNPs and AL/CR was studied using linear regression analysis, adjusting for age and sex, and for principal components when applicable. In addition, we also investigated a total genetic risk score based on the sum of all risk alleles. All analyses were bonferroni-adjusted. **Results:** In the age-group <10 years, three genetic variants (near *GJD2, KCNQ5, ZIC2,* respectively) were found to be significantly associated with AL/CR. In the age-group 10-25 years, five genetic variants (*BMP2, TOX, KCNQ5, A2BP1* and *CACNA1D*) were significantly associated. Genetic variants near *BMP2* and *BICC1* showed a significant increase between <10 years and 10 – 25 years; genetic variants near *LAMA2* and *MYO1D* showed a significant increase from the two younger age-groups to adults. The genetic risk score was associated with AL/CR in all three age-groups, and increased from 0. 0018 (P = <5*10-9) in <10 years, 0. 0032 (P = 5*10^-15) in 10-25 year olds, to 0. 0051 (P 1*10^-72) per risk allele in adults. **Conclusion:** A number of genes associated with refractive error in adults already have a significant effect on ocular biometry in young children. The genes *KCNQ5* and *ZIC2* appear to have all their effect on AL/CR early in life. Most genes (*LAMA2, GJD2*) have an early effect which progresses with age, and some genes (*MYO1D,PABPCP2*) have no observable effect at all <10 years. Our results provide insights on the age period in which myopia genes exert their effect.

958F

Common genetic variants associated with thyroid function may be risk alleles for Hashimoto's disease and Graves' disease. *P. Campbell[1], T. H. Brix[2], S. G. Wilson[1,3,4], L. C. Ward[1], J. Hui[5], J. P. Beilby[5,6], L. Hegedus[2], J. P. Walsh[1,3].* 1) Endocrinology & Diabetes, Sir Charles Gairdner Hospital, Perth, Western Australia, Australia; 2) Department of Endocrinology and Metabolism, Odense University Hospital, Odense, Denmark; 3) School of Medicine and Pharmacology, University of Western Australia, Crawley, Australia; 4) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 5) Pathwest Laboratory Medicine WA, Nedlands, WA, Australia; 6) School of Pathology and Laboratory Medicine, University of Western Australia, Nedlands, WA, Australia.

Recent studies have identified common genetic variants associated with TSH, free T4 and thyroid peroxidase antibodies, but it is unclear whether these differ between patients with Hashimoto's disease and Graves' disease. We genotyped 11 common variants in a discovery cohort of 203 Australian patients with autoimmune thyroid disease (AITD). Two variants with significant or suggestive associations were analysed in a replication cohort of 384 Danish patients. In addition, we compared our results against 3960 healthy controls defined as TPOAb negative with TSH levels within 0·4-4·0 mU/l and no history of thyroid disease. Genotyping was performed using TaqMan allelic discrimination 5¤nuclease assays (Applied Biosystems) with fluorescence measured by a Victor2 Multilabel Plate-Reader (Perkin-Elmer) as per the manufacturer's protocol. Binary association tests were performed using general linear modelling in r v3. 1. 0 and snptest v2. 3. 0, adjusting for sex. For rs753760 (*PDE10A*), the minor allele frequency in Graves' disease and Hashimoto's disease was 0·38 vs. 0·23, respectively, (P=6·42×10-4) in the discovery cohort, 0·29 vs. 0·24 (P=0·147) in the replication cohort and 0·32 vs. 0·24 in combined analysis (P=0·0021; all analyses adjusted for sex). In healthy controls, the frequency was 0·29, significantly different from Hashimoto's disease but not Graves' disease. *PDE10A* has been implicated in cAMP degradation in response to TSH stimulation of thyrocytes and has been positively associated with TSH. It is plausible, therefore that the variant influences disease ascertainment rather than pathogenesis. For rs4889009 (*MAF* gene region), the frequency of the minor G-allele in Graves' disease and Hashimoto's disease was 0·48 vs. 0·36 (P=0·0156) in the discovery cohort, 0·48 vs. 0·34 (P=1·83 × 10-4) in the replication cohort and 0·48 vs. 0·35 in the combined analysis (P=7·53×10-6); in controls, the frequency was 0·38, significantly different from Graves' disease but not Hashimoto's disease. There are three variants within 4kb of each other associated with thyroid volume (rs17767491), goitre (rs3813579) and serum TSH (rs3813582). Two of these variants (rs17767491 and rs3813582) are in moderate LD with rs4889009 (r^2=0·686 for each) suggesting that this region may be a susceptibility locus for Graves' disease. Our data shows that common variants in *PDE10A* and *MAF* gene regions may influence whether patients with AITD develop Graves' disease or Hashimoto's disease.

959W

DQA1 gene and type 1 diabetes. *Y. Lee[1,2,3,4], F. Lo[6,7], C. Huang[1,5], W. Ting[1,5], Y. Lien[1], C. Lin[1], W. Lin[1], C. Chan[1].* 1) Dept Pediatrics & Med Res, Mackay Memorial Hosp, New Taipei City, Taiwan; 2) Department of Pediatrics, School of Medicine, College of Medicine, Taipei Medical University; 3) Institute of Biomedical Sciences, MacKay Medical College; 4) Department of Medicine, MacKay Medical College; 5) Department of Nursing, MacKay Medicine, Nursing and Management College; 6) College of Medicine, Chang Gung University; 7) Department of Pediatrics, Chang Gung Memorial Hospital, Taiwan.

Type 1 diabetes (T1D) is an autoimmune disease characterized by T cell-mediated destruction of pancreatic b cells and the presence of autoantibodies against several b cell autoantigens. Both genetic susceptibility and environmental factors contribute to the pathogenesis. An analysis of combined data of Caucasians, Africans, and Japanese shows that DR4 haplotype –associated susceptibility might be determined by the DQA1 locus. We reported association between DQA1 gene and type 1 diabetes in Han Chinese in Taiwan. **Patients**The subjects were 359 T1D patients. T1D was diagnosed on the basis of clinical manifestations and laboratory evidence. **Controls** The 385 control subjects included hospital personnel and individuals who underwent routine health examinations or minor surgery. None had a history of autoimmune disease. All patients and controls were Han Chinese in Taiwan. **Genotyping of the DQA1 gene**We used sequencing-based typing to genotype the DQA1 gene. **Statistical analysis**Statistical difference in allele distributions between patients and controls were assessed by the chi-square test. Odds ratios (OR) and 95% confidence intervals were also calculated. The Bonferroni correction, $Pc = 1 - (1-P)^n$, was used for multiple comparisons where *Pc* is the corrected *P* value, *P* the uncorrected value, and n the number of comparisons. In this study, n is 8. A *Pc* value of less than 0. 05 was considered statistically significant. **Results**The DQA1*03:01 allele was the most frequent in patients (51. 3%) and controls (31. 4%). It was also significant more frequent in patients than in controls, OR (95%CI) = 2. 29 (1. 85-2. 83), *Pc* = 6. 31E-14. The DQA1*05:01 allele was the second most frequent in patients (38. 3%) and in controls (20. 5%) and conferred susceptibility of T1D, OR = 2. 40 (1. 91-3. 02), *Pc* = 3. 60E-13. Allele DQA1*06:01 was significantly less frequent in patients than in controls (0. 3% vs. 9. 6%) and rendered protection against T1D, OR = 0. 03 (0. 01-0. 06), *Pc* = 2. 66E-15. So were *01:01 [OR=0. 19 (0. 11-0. 32), *Pc* = 1. 23E-08], 01:02 [OR = 0. 21 (0. 14-0. 30), *Pc* <1. 00E-15], and 01:03 [OR = 0. 16 (0. 09-0. 28), *Pc* = 6. 82E-09]. **Conclusion**DQA1 is associated with type 1 diabetes in Han Chinese in Taiwan. DQA1*03:01 and *05:01 confer susceptibility of T1D while *01:01, *01:02, *01:03, and *06:01 render protection against T1D.

960T

Estrogen receptor alpha gene (ESR1) variant is associated with specific clinical features of systemic lupus erythematosus patients in a Brazilian population . *S. E. Lofgren, D. G. Suterio, M. N. Drehmer, I. R. Souza.* Biologia Celular, Embriologia e Genetica, Universidade Federal de Santa Catarina, Florianopolis, SC, Brazil.

Background: Systemic Lupus Erythematosus (SLE), as well as several other autoimmune diseases, is considered to be a women's disease, since around 90% of patients are female. The mechanisms behind this gender bias are not well understood but several studies suggest an important role of male and female hormones. Estrogens are known to have a modulatory effect on several immune responses, many of which are correlated to autoimmune diseases. Here we analyzed a genetic variant in the estrogen receptor-αgene (*ESR1*), implicated in the hormonal regulation of immune cells, for susceptibility and severity of SLE. Material and methods: Genomic DNA was extracted from peripheral white blood cells from 420 individuals (260 healthy controls and 166 SLE patients) by phenol-chlorophorm protocol. Samples were genotyped for the putative functional polymorphism rs2234693 in the*ESR1*gene by pre-designed TaqMan genotyping assay. Allele and genotype frequencies were analyzed in cases and controls and also considering all ACR classification criteria for SLE and other relevant clinical and familial features. Statistical analyses were performed with Unphased software. Results: The tested SNP was not associated with SLE susceptibility*per se*but within the patient group could be related to some specific clinical manifestations. The minor allele rs2234693-C was correlated with the absence of arthritis as well as of anti-ANA and anti-RNP autoantibodies. It could also be associated withnon-occurrence of other autoimmune diseases in the family. On the other hand, the same variant was correlated with the presence of nephritis and discoid rash. Conclusions: In the context of SLE, the minor allele of the*ESR1*gene rs2234693-C was associated with renal and cutaneous involvement. On the contrary, the associated C allele may be considered a protective variant forarthritis and the production of unspecific autoantibodies as well asfamilial occurrence of autoimmune diseases. Overall, our study indicates that the minor C allele is correlated withclinical features thatmay be considereda milder manifestation of the disease. Our results might contribute to the understanding of the influence of genetic polymorphisms in the estrogen pathway and SLE susceptibility.

961F

Association of WNT pathway gene polymorphisms with nonsyndromic oligodontia. *N. Dinckan[1,2,3], Z. O. Uyguner[3], H. Kayserili[4], J. T. Hecht[2,5,6], A. Letra[1,2,6].* 1) Department of Endodontics, University of Texas Health Science Center School of Dentistry at Houston, Houston, TX; 2) Center for Craniofacial Research, University of Texas Health Science Center School of Dentistry at Houston, Houston, TX; 3) Department of Medical Genetics, School of Medicine, Istanbul University, Istanbul, Turkey; 4) Department of Medical Genetics, School of Medicine, Koc University, Istanbul, Turkey; 5) Department of Pediatrics, University of Texas Health Science Center at Houston Medical School, Houston, TX; 6) Pediatric Research Center, University of Texas Health Science Center at Houston Medical School, Houston TX.

Tooth agenesis is the most common developmental abnormality of human dentition and is characterized by the congenital absence of one or more permanent teeth. Oligodontia is the term used to describe severe tooth agenesis, where six or more permanent teeth are missing. Oligodontia may occur as an isolated clinical entity or as a part of a syndrome. Investigations on model organisms have revealed hundreds of genes either directly or indirectly involved in the regulation of tooth development that could contribute to the etiology of oligodontia. The WNT gene pathway regulates multiple developmental processes during craniofacial and tooth development, and variations in WNT pathway genes have been reported in individuals with tooth agenesis. We investigated 34 SNPs in/nearby 12 WNT pathway genes (*WNT3, WNT3A, WNT5A, WNT8A, WNT9B, WNT10A, WNT11, AXIN1, AXIN2, APC, LRP5, LRP6*) for association with oligodontia in 23 Turkish Caucasian multiplex families presenting an average of 6-18 missing teeth. Families were ascertained through probands, and additional relatives were recruited. Genotypes were generated using Taqman Chemistry in a ViiA7 sequence detection instrument. Family-based single SNP association tests were performed using FBAT software. Pairwise-haplotype analysis was performed using the 'hbat' function in FBAT. Bonferroni correction was used to adjust for multiple testing and p-values ≤ 0. 001 were considered statistically significant for the single SNP analyses. We found borderline association for *WNT3A* rs752107 (P=0. 006) and rs1745420 (P=0. 04), *APC* rs861674 (P= 0. 006) and *AXIN2* rs7591 (P=0. 04) with oligodontia. *APC* SNP haplotypes (rs861674; rs2431238) showed strong association with oligodontia (P=0. 001). *WNT3A* rs752107 and *AXIN2* rs7591 are located in the 3' UTR and are predicted to bind to miR-128, and miR-205, respectively, and could have a regulatory role. Our results continue to support a role for WNT pathway genes in the pathogenesis of tooth agenesis, particularly in oligodontia. Further studies should elucidate the mechanism of how these genes contribute to tooth agenesis phenotypes.

962W

Meta-analysis of genome-wide imputed GWAS confirms known significant loci and identifies new suggestive loci in patients with Dupuytren´s disease. *K. Becker[1,2], S. Siegert[1], M. R. Toliat[1], G. H. Dolmans[3], P. M. N. Werker[3], M. Nothnagel[1], P. Nürnberg[1,2], H. C. Hennies[1,2,4].* 1) Cologne Center for Genomics, Cologne, North Rhine-Westphalia, Germany; 2) Cluster of Excellence on Cellular Stress Responses in Aging-associated Diseases, University of Cologne, Cologne, Germany; 3) University of Groningen and Dept. of Plastic Surgery, University Medical Center Groningen, Groningen, the Netherlands; 4) Div. of Human Genetics, Innsbruck Medical University, Innsbruck, Austria.

Dupuytren´s disease, a fibromatosis of the connective tissue in the palm leading to flexion contracture of single fingers, is a common complex disease with a strong genetic component. In a first large GWAS study, nine loci were found to be associated with this disease. Much of the inherited risk in Dupuytren´s disease still needs to be discovered, and the already identified loci explain ~1% of the heritability in this disease. We have now included 1,580 cases and 4,491 controls in a meta-analysis of three genome-wide imputed GWAS datasets. The same quality control criteria were applied to each GWAS dataset and each dataset was imputed separately with IMPUTE2 and the 1000 Genomes reference set. GWAMA was used for the genome-wide association meta-analysis. We used a fixed effects model adjusting for population stratification. We used both the meta-analysis dataset and a whole transcriptome expression dataset (12 cases and 12 controls) to perform extensive network and pathway analysis. Here we have replicated seven previously identified loci, six of these on genome-wide significance level (p-values < 5E-08). The strongest association signal was observed on chromosome 7p14, for SNP rs17171229, with p=1. 11E-28; OR: 2. 02. This imputed SNP gave a slightly stronger signal than the previous top SNP rs16879765 (p=1. 52E-27; OR: 2. 02). We did not observe suggestive association for two previous significant loci on chromosomes 1p36 and 7q31. In addition we identified 14 new suggestive loci (p-value < 1E-05). Strikingly, four of the significantly confirmed loci contain genes that code for Wnt signaling proteins. Some of our newly identified suggestive loci also contain genes associated with Wnt signaling and thus present excellent candidates for replication. In addition we compared the genome-wide expression of genes in patient- and control-derived tissue samples and found the Wnt/β-catenin pathway to be the major deregulated pathway in patient samples. Network analysis provided further evidence that the Wnt signaling pathway in conjunction with other pathways may play an important role in Dupuytren´s disease. Thus we were able to replicate for the first time previously identified susceptibility loci for Dupuytren´s disease in a meta-analysis of genome-wide imputed GWAS datasets. Our findings further corroborate the strong genetic basis for Dupuytren´s disease. Genomic sequencing of the GWAS loci is one next step to identify underlying causative genetic variants.

963T

Association of *PLEKHA7*, *COL11A1*, *ST18 -PCMTD1* and *ABCC5* Polymorphisms with primary angle-closure glaucoma. *J. Chen[1], Y. Jiang[1], Y. Zheng[1], Y. Huang[1], C. Huang[1], H. Chen[1], C. Pang[1,2], M. Zhang[1].* 1) Joint Shantou International Eye Center, Shantou University & the Chinese University of Hong Kong, Shantou, Guangdong, China; 2) Department of Ophthalmology & Visual Sciences, Chinese University of Hong Kong, Hong Kong, China.

Purpose: Primary angle-closure glaucoma (PACG) is one of the leading causes of irreversible blindness in Asians. Recently genome-wide association studies (GWAS) have identified four susceptibility loci associated with PACG: *PLEKHA7, COL11A1*, *ST18-PCMTD1* and *ABCC5*. The current study aimed to investigate the association of these four loci with two subtypes of PACG and ocular biometric parameters, including central corneal thickness, anterior chamber depth, lens thickness, vitreous chamber depth and axial length. **Methods:** In total 762 unrelated controls and 762 unrelated PACG patients, including 400 acute PACG (APACG) and 362 chronic PACG (CPACG) were recruited from Joint Shantou International Eye Center. Four single nucleotide polymorphisms (SNP) including *PLEKHA7* rs11024102, *COL11A1* rs3753841*, ST18-PCMTD1* rs1015213 and *ABCC5* rs1401999 reported by previous GWAS were genotyped in all participants with Taqman SNP genotype assay. Disease association and quantitative trait association were analyzed by logistic and linear regression respectively, controlled for sex and age. All *P* values were adjusted by Bonferroni correction for multiple comparisons. **Results:** *PLEKHA7* rs11024102 was significantly associated with PACG in the dominant model (*P* = 0. 05, odds ratio [OR] = 1. 35). *COL11A1* rs3753841 was significantly associated with PACG and with APACG in the recessive (*P* = 0. 028, OR = 1. 64 for PACG and *P* = 0. 012, OR = 1. 91 for APACG respectively) and additive models (*P* = 0. 008, OR = 1. 30 for PACG; *P* = 0. 047, OR = 1. 30 for APACG respectively). However, for *PCMTD1-ST18* rs1015213 and *ABCC5* rs1401999, there was no any significant association with patient groups in any model (all *P* > 0. 05). Moreover, comparison between APACG and CPACG showed no significant difference between the two subtypes in the four SNPs (all *P* > 0. 05). In quantitative trait association analysis, *COL11A1* rs3753841 was significantly associated with anterior chamber depth (*P* = 0. 004, b ± SE = -0. 082 ± 0. 019 mm). No significant association with any ocular biometric parameter was found in the other three SNPs (all *P* > 0. 05). **Conclusions:** In our study, *PLEKHA7* rs11024102 and *COL11A1* rs3753841 were associated with PACG, confirming the previous GWAS reports. Our results further showed that *COL11A1* was associated with APACG. However, no statistically difference was found between APACG and CPACG in the four SNPs. In addition, the minor allele G of *COL11AL* rs3753841 was correlated with narrower anterior chamber.

964F

Polymorphisms in Macrophage Migration Inhibitory Factor (MIF) And Disposition To Biliary Atresia Among Children. *k. Sadek[1, 2], M. Abdel-Rahman[2, 3, 5], S. Ezzat[4, 5].* 1) Immunogenetics Department, Division of Human Genetics and Genome Research, National Research Centre Egypt; 2) Division of Human Genetics and Department of Ophthalmology, The Ohio State University, Columbus, Ohio, US; 3) Department of Pathology, National Liver Institute, Menoufiya University, Egypt; 4) Department of Public Health, National Liver Institute, Menoufiya University, Egypt; 5) National Liver Institute Sustainable Sciences Institute Collaborative Research Center, Menoufiya University, Egypt.

BACKGROUNDTwo polymorphisms in the promoter region of *MIF* gene, rs755622 (-173 G/C) and rs5844572 (-794 CATT5–8) have been linked to susceptibility or severity of several inflammatory diseases. The two polymorphisms influence the basal and/or induced transcriptional activity of *MIF.* One of these polymorphisms, rs755622, has been implicated in the susceptibility to biliary atresia (BA), while the other polymorphism has not been studied in BA patients. The aim of this study was to investigate the association between these two polymorphisms and BA susceptibility and outcome in Egyptian patients. **METHODS**We conducted a retrospective study of 33 Egyptian infants with confirmed diagnosis of BA. The allele frequency in the population was assessed in 141 healthy adults from the same geographical location. In BA cases, DNA was extracted from paraffin embedded surgically resected tissues. DNA was extracted from the peripheral blood of the controls. Samples were tested for the rs755622 using Taqman real-time PCR and for the rs5844572 using and fluorescence-based genotyping. **RESULTS**Table 1 and 2 summarizes the results of the study. No statistical differences were observed in the allele frequencies between patients and controls. The frequency of the rs755622 C/G and C/C alleles (associated with activation of *MIF)* were observed more frequently in BA patients with marked fibrosis (9/14, 64%) compared with BA patients with mild to moderate fibrosis (5/14, 36%). The difference was statistically not significant. For the rs5844572 polymorphism the frequency of homozygous and heterozygous short repeats (5/5 or 5,X) was higher (11/18, 61. 1%) in BA patients with mild to moderate fibrosis compared with BA patients with marked fibrosis (2/14, 14. 3%), the difference was statistically significant (p= . 007). **CONCLUSION**The result of this study shows no association between *MIF* rs755622 and rs5844572 polymorphisms and susceptibility to BA. However, it suggests that these polymorphisms, in particular rs5844572, may be linked to the degree of fibrosis. This suggests that in patients with high-risk allelotypes of the rs755622 and rs5844572 earlier surgical intervention may be required for better outcome for native liver. **ACKNOWLEDGEMENTS** This project was funded in part by the Sustainable Sciences Institute.

965W

Interactive analyses of *INS* (rs689), *INSR* (rs1799816) and *PPP1R3A* (rs1799999) polymorphisms with Type 2 Diabetes risk in Brahmin group of North-West India. *A. Bhanwer, J. Sokhi.* Human Genetics, Guru Nanak Dev University, Amritsar, Punjab, India.

Type 2 diabetes (T2D) is a complex disorder and its pathogenesis is assumed to involve genetic variations in the candidate genes. The interactions between the genes involved in insulin signaling and secretory pathways are believed to play an important role in determining an individual predisposition to T2D. The 'Brahmins' who constitute a prominent community in India, are the highest of the four Hindu castes. Their traditional occupation is to concern themselves with the spiritual guidance of the people, conduct rituals at marriages, births, deaths and other ceremonies. Till date, strict endogamy practices makes them unique as compared to the rest of the world. Moreover, the studies have reported that people belonging to different caste groups are responsible for their variable susceptibility to disease predisposition. So, in the present study, polymorphisms in genes influencing insulin signaling pathways are studied for their association with T2D in Brahmins from North-West India. The present case-control study enrolled 366 participants (166 T2D cases and age and sex matched 200 healthy controls) with written informed consent. Genotyping was done using PCR-RFLP method. The statistical analysis was carried out using SPSS v16. 0 and Gene-gene interactions were studied through MDR software v3. 0. 2. In case of *INS* (rs689 A>T) and *INSR* (rs1799816 G>A) polymorphism, a significant association was observed in the genotype and allele frequency distribution among T2D cases and controls (p<0. 05). Further, genotype model analysis indicated that the AT genotype of *INS* A>T polymorphism provided >2. 0 fold increased risk towards T2D (p=0. 001), the GA genotype of *INSR* G>A polymorphism showed nearly 2. 4 fold risk of T2D (p=0. 005) after correction for age, sex, BMI, WC, WHR whereas no association of *PPP1R3A* C>A polymorphism with T2D was observed in Brahmins. The interactive analyses revealed nearly 2 fold increased risk of T2D (p<0. 0001). Furthermore, AA-GA-CA and AT-GG-CA genotypes combinations of three polymorphisms (p=0. 0432; p=0. 0026 respectively) indicated 3 fold higher chances of developing T2D in Brahmins of North-West India. In conclusion, our results provided the evidences for the interactions between the genes belonging to same signaling cascade. It also revealed that the cumulative effect of more than one polymorphism can help understanding the better picture of a complex trait like T2D in different population groups.

966T

Genetic Risk Score Based on Type 2 Diabetes Variants Predicts Longitudinal Decline in β-cell Function in Mexican Americans at Risk for Type 2 Diabetes. *M. Black[1], J. Wu[1], YH. Shu[1], E. Trigo[2], K. Taylor[3], J. Rotter[3], R. M. Watanabe[4,5,6], T. A. Buchanan[2,4,6], A. H. Xiang[1].* 1) Department of Research & Evaluation, Kaiser Permanente Southern California, Pasadena, CA; 2) Department of Medicine Keck School of Medicine, the University of Southern California, Los Angeles, CA; 3) Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute at Harbor-University of California Los Angeles Medical Center, Torrance, CA; 4) Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA; 5) USC Diabetes and Obesity Research Institute, Los Angeles, CA; 6) Department of Physiology & Biophysics, Keck School of Medicine of USC, Los Angeles, CA.

Across the genome, associations between 62 variants and type 2 diabetes mellitus (T2DM) have been robustly replicated in various populations. However, association of these variants with longitudinal decline in β-cell compensation for insulin resistance has not been comprehensively assessed. We examined association between these variants, each residing in a distinct gene or intergenic region, and the rate of change in β-cell function in 373 BetaGene participants (mean±SD age: 35. 0±8. 2 years; 74. 8% female) with baseline and follow-up phenotyping (mean±SD length of follow-up: 4. 6±1. 5 years). β-cell function (disposition index; DI) was estimated from frequently sampled intravenous glucose tolerance tests with Minimal Model analysis. Genotyping was performed on the Illumina OmniExpress platform. An additive genetic model was assumed for all SNPs, and all models were adjusted for age, sex and baseline BMI. We tested each variant individually and jointly using an unweighted genetic risk score (GRS) constructed by summing the number of risk alleles across the 62 loci for each subject, where risk allele was defined by the observed direction of effect in univariate association analyses of SNP vs. rate of change in DI using BetaGene data. The GRS was approximately normally distributed (mean±SD: 63. 1±5. 3). Only 3 SNPs were individually associated with the rate of change in DI at a nominal significance level: *KCNK16* rs3734621 (p=0. 009), *HMG20A* rs7177055 (p=0. 019), *PRC1* rs12899811 (p=0. 033). However, the GRS was highly associated with the rate of change in DI (p=1. 40x10-8). Participants in the highest GRS tertile (65<GRS<77) had an average decline in DI of 885 units per year, whereas those in the lowest GRS tertile (48<GRS<61) had an average decline in DI of 62 units per year. These data provide evidence that a GRS based on SNPs associated with T2DM can significantly predict longitudinal decline in β-cell function over time in Mexican Americans at risk for T2DM.

967F

Interaction Analysis of *CAPN10* and *PGC-1*αGenes with Type 2 Diabetes in Three Unrelated Endogamous Groups of North-West India (Punjab): A Case-Control and Meta-Analysis Study. *R. Sharma[1], R. Kapoor[2], A. Bhanwer[1].* 1) Human Genetics, Guru Nanak Dev University, AMRITSAR, Punjab, India; 2) Heart Station and Diabetes Clinic, Amritsar, India.

Type 2 Diabetes (T2D) is caused by αcell dysfunction and/or insulin resistance, which is promoted by multifactorial genetic or environmental factors. To assess the influence of ethnicity, the association of *CAPN10* and *PGC-1ć* genes was investigated in total of 1125 samples (554 T2D cases and 571 controls) from Bania, Brahmin and Jat Sikh endogamous groups of North-West India (Punjab). Single-locus analysis showed significant difference in association of *CAPN10* (SNP-43, SNP-19, SNP-63) and *PGC-1ć*(Thr612Met, Thr528Thr, Gly482Ser, Thr394Thr, IVS2+52C>A) variants with T2D among studied groups. However, no association of *PGC-1ć* (Asp475Asp) polymorphism was seen in any group. Haplotype analysis depicted that in *CAPN10*, 212-haplotype in Bania group, 221-haplotype in Brahmin, 211 and 221-haplotypes in Jat Sikh group conferred increased risk towards T2D. *PGC-1ć* haplotypes GAGGCC, GAGGTA in Bania, GGGGCA, GGGGTC in Brahmin; GAAGTC, GGGGTC in Jat Sikh group provide 1. 5-5 fold increased T2D risk. Strong synergistic interactions were observed between *CAPN10* SNP-63, *PGC-1ć*Gly482Ser polymorphisms in Bania; *CAPN10* SNP-43, *PGC-1ć*Thr528Thr and Thr394Thr SNPs in Brahmin; *PGC-1ć*Thr528Thr, Gly482Ser, Thr394Thr variants in Jat Sikh group. Further, meta-analyses revealed that D-allele of *CAPN10* SNP-19 [p=0. 01, OR=1. 08 (1. 01-1. 2) at 95% CI] confer 1. 08-fold increased risk towards T2D predisposition. Similarly, minor alleles of *PGC-1ć*Gly482Ser [p=0. 006, OR=1. 23 (1. 06-1. 45) at 95% CI] and Thr394Thr [p=0. 00, OR=1. 41 (1. 20-1. 67) at 95% CI] showed significant overall association with T2D risk after conducting meta-analyses. Differential pattern of association of *CAPN10* and *PGC-1ć* genes affecting risk of T2D in studied ethnic groups emphasize the role of genetic heterogeneity. Knowledge of the ethnicity and nature of variation in risk factors needs serious attention for disease risk calculation and proper medical intervention. Therefore, further large-scale and better designed population-based studies are needed to ascertain the impact of ethnicity and studied variants on diabetes risk.

968W

The mechanism of cleft palate after palatal fusion. *S. Suzuki, H. Imura, N. Natsume.* Div. Research and Treatment for Oral and Maxillofac. Congenital Anomalies. , Aichi-Gakuin University, Nagoya City, Aichi, Japan.

Palatal formation starts with an elevation in bilateral palatine processes onto the tongue in the form of vertical shelves, followed by a downward migration (elevation stage). The elevated palatine processes then elongate and contact each other at the midline (contact stage). Adhesion and self-decomposition of the epithelium occur at the contact location, and mesenchymal tissues then fuse (fusion stage). Cleft palate must therefore be induced by a disruption during the developmental process of the palatal shelves. Cleft palate can be occured due to a failure in the elevation of palatal shelves, failure in elevated palatal shelves to initiate contact, failure in proper fusion, or breakdown subsequent to fusion. However, we reported that some embryos appeared to develop cleft palates after palatal fusion by TCDD(2,3,7,8-Tetrachlorodibenzodioxin). On the other hand, the same present status has been reported in mesenchyme homeobox 2(MEOX2) knockout mouse. We observed MEOX2 gene from the blood from patients with cleft palate and performed PCR-Direct Sequence for 99 cleft palate patients. All patients were informed and consented. This study was approved by institutional review boards. There were two genetic variants , rs2237493 and rs113582077,were observed but we could not find significant differences between cases and controls. Further functional investigation will be required.

969T

An investigation of toll-like receptors in tuberculosis susceptibility reveals sex-specific associations for *TLR8* polymorphisms. *M. Salie, M. Daya, L. A. Lucas, R. M. Warren, G. D. van der Spuy, P. D. van Helden, E. G. Hoal, M. Moller.* Biomedical Science, Stellenbosch University, Tygerberg, Western Cape, South Africa.

Toll-like receptors (TLRs) are involved in the recognition of conserved microbial structures, leading to activation of an inflammatory response and formation of an adaptive immune response. Twenty-three polymorphisms in five *TLR* genes were genotyped in 623 tuberculosis cases and 473 healthy controls in a population-based case-control association study in a South African population. We detected sex-specific associations for *TLR8* polymorphisms, with rs3761624 (OR=1. 54, *P*<0. 001), rs3764879 (OR=1. 41, *P*=0. 011) and rs3764880 (OR=1. 42, *P*=0. 011) associated in females and rs3764879 (OR=0. 72, *P*=0. 013) and rs3764880 (OR=0. 75, *P*=0. 036) associated in males. Epistatic interactions between the *TLR* genes were investigated and the *TLR1*_rs4833095 polymorphism was shown to interact with *TLR2*_rs3804100 and (GT)n microsattelite (*P*=0. 002) and alter susceptibility to TB. We also studied the role of *TLRs* in disease caused by different *Mycobacterium tuberculosis* genotypes in 257 tuberculosis cases, and identified associations between specific *TLR* polymorphisms and disease caused by specific strains. This study provides further evidence that the TLRs play an important role in the outcome of tuberculosis disease, and suggests a partial explanation for the male bias in tuberculosis ratios.

970F

Genetic associations with non-fasting metabolite concentrations. *D. O. Mook-Kanamori[1, 2, 3], R. Li-Gao[1], R. de Mutsert[1], K. Willems van Dijk[4, 5], F. R. Rosendaal[1].* 1) Clinical Epidemiology, Leiden University Medical Center, Leiden, South Holland, Netherlands; 2) Public Health and Primary Care, Leiden University Medical Center, Leiden, South Holland, Netherlands; 3) BESC, Epidemiology Section, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia; 4) Human Genetics, Leiden University Medical Center, Leiden, South Holland, Netherlands; 5) Endocrinology, Leiden University Medical Center, Leiden, South Holland, Netherlands.

In 2011, two genome-wide association studies described fifteen loci (*FADS1, ACADS, ACADM, ACADL, ELOVL2, CPS1, PLEKHH1, SYNE2, SPTLC3, ETFDH, SLC16A9, ACSL1, SCD, SLC22A4* and *PHGDH*) that showed strong associations with a wide range of metabolite concentrations as measured by the Biocrates Absolute*IDQ*TM p150 kit. These studies suggested that genetic variants could explain a large proportion of phenotypic variance by taking metabolites as endophenotypes. Their findings were limited to metabolite concentrations in the fasting state. However, the human body resides in a non-fasting state for the greater part of the day. Therefore, our objective was to examine these associations in non-fasting metabolite concentrations. This study was embedded in the Netherlands Epidemiology of Obesity (NEO) study, a population-based prospective cohort of 6,671 persons aged 45-65 years. In 477 participants (56% males, mean BMI 30. 4 kg/m2) the metabolite concentrations were measured by the Biocrates Absolute*IDQ*TM p150 kit both fasting and 150 minutes after a meal challenge (600 kCal, 16 En% protein, 50 En% CHO, 34 En% fat). Genotyping was performed using the Illumina HumanCoreExome array. Additive models were used to assess the associations between the genetic variants in the fifteen loci with the metabolites in a fasting state, a postprandial state and with the metabolite response to the meal. After Bonferroni correction, we replicated thirteen out of the fifteen loci that were associated with the fasting metabolites (except *SYNE2* and *SCD*). For nine of the fifteen associated loci, the metabolite concentrations changed significantly after the meal (except for *FADS1*, *ELOVL2, PLEKHH1, SYNE2, SPTLC3* and *ACSL1*), ranging from 6% to 23%. For seven of these nine loci there were also associations with the postprandial metabolite concentrations (Beta (P-value)): *ACADS* (-0. 13 (9. 6*10-12)*, ACADM* (0. 06 (6. 8*10-5)*, ACADL* (0. 17 (7. 7*10-18)*, CPS1* (0. 15 (4. 6*10-8)*, SLC16A9* (-0. 10 (5. 2*10-9)*, SLC22A4* (-0. 07 (2. 0*10-4) and *PHGDH* (0. 05 (2. 7*10-3). The fifteen loci were not associated with the metabolite responses to the meal. In conclusion, our findings suggest that the associations between these loci and metabolite concentrations are not influenced by the fasting state of the participants. A genome-wide association study on non-fasting metabolites is required to identify genetic variants that are specifically associated with non-fasting metabolite concentrations.

971W

Rare genomic variants in *PARD3* are potential risk factors for human cranial neural tube defec. *X Chen[1,2,3], Yu An[4], Yonghui Gao[1,5], Liu Guo[6], Lei Rui[7], Hua Xie[1], Mei Sun[3], Hong Shao[2], Siv. I Hung[2], Xiaoming Sheng[2], Jizhen Zou[8], Yihua Bao[1], Hongyan Guan[9], Bo Niu[10], Zandong Li[7], Richard Finnell[11], James F Gusella[3,12], Ting Zhang[1], Bai-Lin Wu[2,4,13].* 1) Capital Institute of Pediatrics, Beijing, Beijing, China; 2) Department of Laboratory Medicine, Children's Hospital Boston, Boston, MA, 02115, USA; 3) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, 02114, US; 4) Children's Hospital and Institutes of Biomedical Science, Shanghai Medical College of Fudan University, Shanghai, 200032, China; 5) Institute of Acu-moxibustion, China Academy of Chinese Medical Sciences, Beijing, 100700, China; 6) Department of Neurology, Affiliated Children's Hospital of Capital Institute of Pediatrics, Beijing, 100020, China; 7) State Key Laboratories for Agrobiotechnology, China Agricultural University, Beijing, 100193, China; 8) Department of Pathology, Affiliated Children's Hospital of Capital Institute of Pediatrics, Beijing, 100020, China; 9) Department of Integrated Early Childhood Development, Capital Institute of Pediatrics, Beijing, 100020, China; 10) Department of Biotechnology, Capital Institute of Pediatrics, Beijing, 100020, China; 11) Dell Pediatric Research Institute, Department of Nutritional Sciences, The University of Texas at Austin, Austin, Texas 78723 USA; 12) Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA; 13) Department of Pathology, Harvard Medical School, Boston, MA 02115, USA.

The genetic etiology of human neural tube defects (NTDs) is largely unknown, although rare mutations of the planar cell polarity (PCP) genes have been shown to contribute to human NTD susceptibility. The increasing evidence for interactions between PCP and the apical–basal polarity (ABP) pathway, and the confirmed role of ABP proteins in zebrafish neurulation, prompted us to hypothesize that rare genomic variants of the core ABP gene, *PARD3,* are associated with susceptibility to human NTDs. In this study, we identified five rare genomic variants of *PARD3* from 100 Chinese cranial NTD cases, including one microdeletion and four single nucleotide variants. Overall, rare *PARD3* genomic variants are significantly associated with increased risk for cranial NTDs (5/100 *vs.* 0/100, $p<0.05$). By contrast, none of rare, NTD-specific SNV was found in 1044 Han Chinese controls. Over-expression analysis in HEK293T and MDCK cells confirmed a loss-of function model for two *PARD3* variants (D783G and P913Q), resulting in defective tight junction formation via disrupted aPKC binding. *PARD3* knockdown by siRNA in human neural progenitor cells gave rise to abnormal cell polarity. *cPard3* knockdown in chick embryos using a shRNA lentivirus delivered by *in ovo* microinjection resulted in significant more embryos with small tail phenotype, reduced epithelial polarization and ectopic hinge-like invagination occurs on the neuroepithelial tissue of chick neural tube due to the insufficiency of cPard3. Our studies suggest that rare, loss-of-function variants of *PARD3* are potential risk factors for human cranial NTDs, possibly by disrupting apical tight junction formation and epithelial polarization process.

972T

Replication of GWAS findings for cleft lip/palate in a Han Chinese population. *Z. Jia[1,2], Q. Yu[1,2], J. Ma[1,2], S. He[1,2], B. Shi[1,2].* 1) State Key Laboratory of Oral Disease, West China hospital of Stomatology, Sichuan University, Chengdu, Sichuan, China; 2) Department of Cleft Lip and Palate Surgery, West China Hospital of Stomatology, Sichuan University, chengdu, China;.

Background: Isolated cleft lip with or without cleft palate (CL/P) is one of the most common birth defects, with a prevalence of 1/700 in Asian. Genome wide association (GWA) study as the most powerful tool already identified numerous susceptible loci, however, these loci require multiple replications across populations to validate for etiologic SNP. **Methods:** In this study, we replicated 37 SNPs which has the most significant p values in published GWA studies, genotyping by using SNPscan among 309 NSCL/P trios (143 NSCLO and 166 NSCLP) from Western Han Chinese. We performed the transmission disequilibrium test (TDT) on individual SNPs and gene-gene (GxG) interaction analyses on the family data. Parent-of-Origin effects were assessed by separately considering transmissions from heterozygous fathers versus heterozygous mothers to affected offspring. **Results:** Allelic TDT results showed that T allele at rs2235371 (IRF6) and C allele at rs12063989 (24kb 3' of DIEXF) was uniquely associated with NSCLO (p=0. 00015 and 0. 00016, ORtransmission=0. 48 and 0. 49, 95%CI: 0. 33-0. 71 and 0. 34-0. 72, respectively). Individual SNP TDT analysis did not show rs742071 was associated with any of the three cleft groups (NSCLO, NSCLP and NSCL/P); While, T allele of rs742071 (PAX7) is paternally under-transmitted (p= 0. 0043) and parent-of-origin effects is also significant (Z=-2. 88 and p=0. 004). Genotypic TDT for epistatic interactions showed that rs2235371 interacts with rs11119388 (SYT14) (p= 1. 10E-12) among NSCL/P group. **Conclusions:**This study further confirmed that IRF6 gene rs2235371 was associated NSCLO from Asian populations. Rs2235371 interacts with rs11119388 (p= 1. 10E-12) among NSCL/P group, suggesting they may act in the same pathway. These findings provide solid support for IRF6 gene playing a critical role in human CL/P.

973F

Genotyping analysis of 3 RET polymorphisms demonstrates low somatic mutation rate in Chinese Hirschsprung disease patient. *Q. Jiang[1], Z. Zhang[2], Q. Li[2], W. Cheng[3], GL. Qiao[2], P. Xiao[4], L. Gan[2], L. Su[5], CY. Miao[6], L. Li[2].* 1) Department of Medical Genetics, Capital Institute of Pediatrics, Beijing, China; 2) Department of Pediatric Surgery, Capital Institute of Pediatrics, Beijing, China; 3) Department of Surgery, Beijing United Family Hospital, China; Department of Paediatrics and Surgery, Faculty of Medicine, Nursing and Health Sciences, Monash University, Victoria, Australia; 4) Department of Pathology, Capital Institute of Pediatrics Affiliated Children's Hospital, Beijing, China; 5) Anhui Medical University, Hefei, China; 6) Peking University Capital Institute of Pediatrics Teaching Hospital, Beijing, China.

Background: Genetic mosaicism has been reported for both coding and non-coding sequences in the RET gene in Hirschsprung disease (HSCR) patients. This study aimed to investigate somatic mutation rate in Chinese population by comparing both homozygous genotype percentage and risk allele frequency of 3 RET single nucleotide polymorphisms (SNPs) among blood and colon samples. Methods: DNA was extracted from 59 HSCR blood samples, 59 control blood samples and 76 fresh frozen colon tissue samples (grouped into ganglionic, transitional and aganglionic level). Genotype status of rs2435357 and rs2506030 was examined by competitive allele specific hydrolysis probes (Taqman) PCR technology, and rs2506004 was examined by Sanger sequencing. Homozygous genotype percentage and risk allele frequency were calculated for each type of sample and compared by chi-square test. P<0. 05 was regarded as being statistically significant. Results: Colon tissue DNA samples showed similar frequency of SNPs as that of the blood DNA samples in HSCR patients, both of which are significantly higher than the control blood group (rs2435357 TT genotype: 71. 2%, 74. 7% versus 22. 0% in HSCR blood, HSCR colon and control blood DNA respectively, P=0. 000; rs2506004 AA genotype: 72. 4%, 83. 1% versus 25. 5%, P=0. 000; rs2506030 GG genotype: 79. 7%, 77. 2% versus 54. 2%, P=0. 000 and 0. 004). With respect to DNA extracted from ganglionic, transitional and aganglionic levels, no statistically significant difference was demonstrated in those 3 regions (rs2435357: p=0. 897; rs2506004: p=0. 740; rs2506030: P=0. 901). Conclusion: Our data does not support the notion that high frequency of somatic changes as an underlying etiology of Chinese HSCR population.

974W

Targeted next generation sequencing identifies genetic regions influencing the development of chronic obstructive pulmonary disorder in a population from northern Sweden. *J. Klar[1], H. Matsson[2], C. Söderhäll[2], M. Lamontagne[3], S. Gudmundsson[1], H. Backman[4], A. Lindberg[4], E. Rönmark[4], D. Sin[5], D. S. Postma[6], Y. Bossé[3], J. Kere[2], B. Lundbäck[7], N. Dahl[1].* 1) Department for Immunology, genetics and pathology, Uppsala University, Uppsala, Sweden; 2) Department of Biosciences and Nutrition and Center for Innovative Medicine (CIMED), Karolinska Institutet, Huddinge, Sweden; 3) Institut universitaire de cardiologie et de pneumologie de Québec, Department of Molecular Medicine, Laval University, Québec, Canada; 4) The OLIN studies, Sunderby Hospital of Norrbotten, Luleå, Sweden; 5) The University of British Columbia Center for Heart Lung Innovation, St-Paul's Hospital, Vancouver, Canada; 6) Center Groningen, GRIAC research institute, University of Groningen, Groningen, The Netherlands; 7) Krefting Research Centre, Institute of Medicine, University of Gothenburg, Sweden.

Background: Chronic obstructive pulmonary disease (COPD) is a common disease in Sweden, affecting approximately one in six in age >40 years, with severe impact on health and quality of life. COPD results from environmental factors, especially cigarette smoking, with contributions from yet unknown genetic background factors and by gene-environment interactions. The general objective of this project is to get insight into the mechanisms and interactions of specific genes and genetic pathways that contribute to the development of COPD. **Aim:** To investigate candidate genes, with emphasis on genes important for lung development and homeostasis, in defined populations of smokers and non-smokers in search for genetic variations associated with development of COPD. **Methods:** We conducted a candidate gene analysis on 22 genes implicated in lung development and 71 genes and regions previously associated with COPD. Targeted enrichment (HaloPlex; Agilent) and high throughput sequencing (Illumina) was performed on patients and healthy controls retrieved from the Swedish Obstructive Lung Disease in Norrbotten (OLIN) Studies sample set. Pairwise linkage disequilibrium (LD) analysis was performed using SNAP and the variant data set from the CEU population panel of the 1000 genome project. Lung eQTL was calculated using genotyping and gene expression data from patients who underwent lung surgery. **Results:** We identified a total of 2,151 SNPs of which 74, distributed in 45 gene regions, have significantly different allele frequencies in COPD cases than controls. We identified six regions with associated variants in pairwise LD ($r2 > 0.7$, $D' = 1.0$). The variants in all six regions were confirmed to likely form haplotypes in the carriers. Seven associated variants were detected and significantly associated with gene expression. **Conclusion:** In summary, this study identified genetic variants at several loci significantly associated with COPD and replicated several previous GWAS results. In addition, we identified several regions in LD associated to COPD in the Swedish population. We further indicate the advantage of using less heterogeneous populations in the studies of complex disorders, as we, although using a small sample set, were able to detect genetic variants significantly associated to COPD.

975T

Genetic variant near *PLXDC2* gene influences the risk of primary open-angle glaucoma by increasing intraocular pressure. *F. Mabuchi[1], N. Mabuchi[1], Y. Sakurada[1], S. Yoneyama[1], K. Kashiwagi[1], Z. Yamagata[2], H. Iijima[1].* 1) Dept Ophthalmology, Univ Yamanashi, Chuo, Yamanashi, Japan; 2) Dept Health Sciences, Univ Yamanashi, Chuo, Yamanashi, Japan.

Purpose: To assess the association between the genetic variants previously reported to be associated with primary open-angle glaucoma (POAG) in the Japanese population and the phenotypic features, including intraocular pressure (IOP). **Methods:** Six hundred sixty one Japanese patients, including 417 patients with POAG (210 patients with normal tension glaucoma: NTG and 207 patients with high tension glaucoma: HTG) and 244 control subjects without glaucoma, were analyzed for 3 genetic variants reported to be associated with POAG; rs547984 (near gene: *ZP4*), rs7081455 (*PLXDC2*), and rs7961953 (*TMTC2*). Allele frequency differences were compared between POAG (NTG or HTG) patients and control subjects. A logistic regression model was used to study the effects of genotyped alleles when comparing POAG patients with control subjects. Demographic and clinical features, including maximum IOP and vertical cup-to-disc ratio (VCDR), were compared between alleles or genotypes. A multiple linear regression analysis was carried out to confirm the association between the risk alleles and maximum IOP or VCDR. **Results:** There were significant differences in the rs7081455 (*PLXDC2*) allele frequencies between the HTG (P = 0. 011) or POAG (P = 0. 005) patients and control subjects. Adjusted for age and gender, an almost 1. 5 increased risk of HTG (P = 0. 012) and POAG (P = 0. 0042) was found with the G allele of rs7081455 (*PLXDC2*). Maximum IOP (23. 5 ± 10. 3mmHg, mean ± standard deviation) in patients with GG genotype of rs7081455 (*PLXDC2*) was significantly higher (P = 0. 0073) than that (19. 9 ± 7. 4mmHg) in patients with TT genotype. VCDR (0. 72 ± 0. 27) in patients with risk G allele of rs7081455 (*PLXDC2*) was significantly larger (P = 0. 0088) than that (0. 72 ± 0. 27) in patients without G allele. A linear regression analysis confirmed significant association between the G allele of rs7081455 (*PLXDC2*) and maximum IOP (P = 0. 012) or VCDR (P = 0. 014). **Conclusion:** Genetic variant near *PLXDC2* gene is considered to influence the risk of POAG by increasing IOP.

976F

Gene-hormone therapy interaction and fracture risk in postmenopausal women. *Y. Wang[1], J. Wactawski-Wende[1], L. Sucheston-Campbell[2], L. Preus[1,2], J. Nie[1], R. D. Jackson[3], S. Handelman[4], R. Nassir[5], C. Crandall[6], H. M. Ochs-Balcom[1].* 1) Department of Epidemiology and Environmental Health, University at Buffalo, Buffalo, NY; 2) Cancer Prevention and Control, Division of Cancer Prevention and Population Sciences, Roswell Park Cancer Institute, Buffalo, NY; 3) Department of Internal Medicine, Division of Endocrinology, Diabetes and Metabolism, The Ohio State University, Columbus, OH; 4) Center for Pharmacogenomics, Department of Molecular Virology, Immunology and Medical Genetics, The Ohio State University, Columbus, OH; 5) Department of Biochemistry and Molecular Medicine, University of California Davis, Davis, CA; 6) Division of General Internal Medicine and Health Sciences Research, David Geffen School of Medicine at University of California, Los Angeles, Los Angeles, CA.

Given the completed meta-genome-wide association studies (GWAS) and known protective effects of hormone therapy (HT) on bone, we hypothesized that HT modifies the association of a fracture genetic risk score (GRS) comprising 16 meta-GWAS significant fracture-associated SNPs with fracture (total body, central body, lower limb, and upper limb). In 9,922 European Americans in Women's Health Initiative (WHI) HT randomized trials, including 1608 incident fractures from 1993 to 2005, we studied this interaction using Cox proportional hazards models and case-only analyses. In Cox models, the fracture region with the most significant interaction was central body fracture (pelvis, spine, hip, clavicle and scapula); however this multiplicative interaction term did not meet statistical significance (p-interaction=0. 19) regardless of covariate inclusion (age, trial, and sub-study (GARNET, WHIMS+)). In case-only analyses, we saw borderline evidence for multiplicative GRS-HT interaction for central body fracture; p=0. 13. In z-score-stratified models, women in the highest GRS category who were not randomized to HT were at the highest fracture risk: adjusted hazard ratio (HR)=2. 38 (95% CI: 1. 15-4. 91) compared to women in the lowest genetic risk category who were randomized to HT: 0. 97 (95% CI: 0. 43-2. 18). Additive interaction for central body fracture estimated using RERI (relative excess risk due to interaction) was marginally significant in the highest GRS category for central fracture; RERI=1. 10 (95% CI: -0. 06-2. 27), p=0. 06. Our results support exploration of genetic risk score as another factor in determining potential risk and benefit ratio of hormone therapy use for fracture prevention.

977W

Exploring interactions between genetic variation under linkage peaks and cigarette smoking on body weight in families. *K. L. Edwards[1], J. Y. Wan[1], C. Ziadeh[1], N. Shridhar[2], T. M. Norden-Krichmar[1], S. A. Santorico[3].* 1) Dept. of Epidemiology, University of California, Irvine, CA; 2) Dept. of Epidemiology, University of Washington, Seattle, WA; 3) Dept. of Mathematical and Statistical Sciences, University of Colorado, Denver.

The Metabolic Syndrome (MetS) is a complex condition characterized by a cluster of risk factors, including obesity, lipid abnormalities, hypertension and glucose intolerance. Previous work using families and linkage analysis identified chromosomal regions with evidence for linkage to both individual and clusters of MetS traits. Examining evidence for gene x environment (GxE) interactions under these linkage peaks provides a unique opportunity to further characterize the genetic and environmental underpinnings of complex disease and to explore sources of missing heritability. Cigarette smoking is an ubiquitous exposure that is known to influence MetS phenotypes and regulate gene expression via epigenetic modification. Ordered subset analysis (OSA) was used to examine evidence for interactions between smoking and genetic variation on body weight in 69 European American families, comprised of 495 subjects. OSA aims to reduce genetic heterogeneity by finding the subset of families with similar smoking backgrounds that maximize evidence for linkage. OSA was conduced using a variance components approach as implemented in SOLAR. Statistical significance was assessed by permutation testing. Families were subsetted using the proportion of smoking family members and in a separate analysis the average number of packs per day for the family. A subset of 12 families with 49 total subjects (large proportion of whom reported ever smoking) provided evidencefor linkage with body weight (LOD=1. 74, P=0. 05 at 132 cM in the chromome 2 linkage region). When the average number of packs per day was used to subset families, evidence for linkage was found in the chromsome 2 linkage region between 136-137 cM (LOD=2. 2, P=0. 01, 40 families, 293 total subjects). These results support not only the presence of a gene(s) that influences body weight in the previously identified linkage region on chromosome 2, but one that is also modified by smoking. Additional results will be presented for additional previously identified linkage regions, including those in Mexican, Japanese and African American families. Overall, it appears that evidence for linkage with body weight is modified by smoking. Targeting previously identified linkage regions may be a fruitful way to identify GxE.

978T

Genome-wide analysis of gene interactions with an environmental risk score for colorectal cancer risk. *M. Du[1], J. Gong[1], M. J. Gunter[2], A. Toland[3], S. I. Berndt[4], H. Brenner[5], A. T. Chan[6], J. Chang-Claude[5], M. Hoffmeister[5], J. Jeon[1], L. Le Marchand[7], W. Li[8], Y. Lin[1], J. D. Potter[1], M. L. Slattery[9], A. Vargas[10], P. A. Newcomb[1], U. Peters[1], E. White[1], L. Hsu[1].* 1) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) School of Public Health, Department of Epidemiology and Biostatistics, Imperial College London, London, W2 1PG, United Kingdom; 3) Department of Molecular Virology, Immunology, & Medical Genetics, Department of Internal Medicine Division of Human Genetics, The Ohio State University, Columbus, OH; 4) Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD; 5) Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany; 6) Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA; 7) Epidemiology Program, University of Hawaii Cancer Center, Honolulu, HI; 8) Division of Preventive and Behavioral Medicine, Departments of Medicine, University of Massachusetts Medical School, Worcester, MA; 9) Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT; 10) Cancer Prevention Fellowship Program, National Cancer Institute, Bethesda, MD.

Numerous environmental factors contribute to colorectal cancer (CRC) development. Studies examining gene interactions with individual risk factors have provided biological insight into these associations. However, many predictors have additive effects and a score combining known CRC risk factors may better characterize an individual's overall environmental risk profile. Use of such a composite score may increase statistical power to detect gene interactions, which can, in turn, inform preventive strategies. To study genome-wide gene interactions with combined environmental risk of CRC, we examined 7,945 CRC cases and 8,893 controls of European ancestry in the Genetics and Epidemiology of Colorectal Cancer Consortium. Using a reference panel of 955 participants who underwent sequencing, we imputed whole-genome sequence data into remaining participants with genome-wide array data. To assess overall environmental risk of CRC, we calculated an environmental risk score (ERS) comprising 17 established or probable CRC risk factors based on published findings. To test for multiplicative interactions, we used multivariable logistic regression and defined statistical significance using the conventional genome-wide α=5E-08. We further explored interactions using the potentially more powerful 2-stage Cocktail approach. The ERS was associated with increased risk of CRC (OR highest vs. lowest quartile=2. 41, 95% CI: 2. 20-2. 64). We observed a quantitative interaction between the ERS and a variant on chromosome 12q21. 33 using logistic regression (MAF=0. 26, OR-interaction=0. 87, *P*=6. 5E-09). For the highest vs. lowest ERS quartile, the association with CRC risk was stronger in participants carrying one (OR=3. 00, 95% CI: 2. 58-3. 50) or two (OR=3. 68, 95% CI: 2. 52-5. 39) copies of the minor allele compared to those with two copies of the major allele (OR=1. 97, 95% CI: 1. 74-2. 22). This variant is in the intergenic region between *BTG1* and *CLLU1*, genes which regulate cell growth and differentiation and have been implicated in cancer (chronic lymphocytic leukemia) and cardiovascular disease. There were no statistically significant interactions using the Cocktail approach. We identified a novel SNP interaction with an ERS for CRC risk—suggesting that aggregating known risk factors can help uncover gene interactions. These data suggest that the combined influence of environmental factors on CRC may be modified by common genetic variation, and warrant replication and further investigation.

979F

Interaction between variants in *CLU* and *MS4A4E* modulates Alzheimer's disease risk. *M. Ebbert[1], K. Boehme[1], M. Wadsworth[1], S. Mukherjeeb[2], P. Crane[2], P. Ridge[1], J. Kauwe[1], Alzheimer's Disease Neuroimaging Initiative, Alzheimer's Disease Genetics Consortium.* 1) Department of Biology, Brigham Young University, Provo, UT; 2) Department of Medicine, University of Washington, Seattle, Washington 98105, USA.

Epistasis affects many human phenotypes and likely plays a critical role in Alzheimer's disease (AD), but most AD studies focus on single loci. Here, we present strong evidence supporting epistasis between two top AD loci that modulate AD risk with a combined odds ratio (OR) of 2. 51, which is higher than any single AD locus except *APOEϵ4*, *APP*, and *TREM2.* This analysis was performed using the largest dataset employed in an epistasis study, to date. Ebbert et al. recently reported two potential gene-gene interactions between rs11136000 (*CLU*) and rs670139 (*MS4A4E*) (SF=3. 81, p=0. 016), and rs3865444 (*CD33*) and rs670139 (*MS4A4E*) (SF=5. 31, p=0. 003) using the Cache County data, but the interactions have not yet been replicated. Using nine independent datasets (3838 cases, 4532 controls) from the Alzheimer's Disease Genetics Consortium (ADGC), we tested each interaction, controlling for age, gender, and *APOEϵ4* dose. We then performed two meta-analyses per interaction (ADGC only and with Cache) using METAL, and performed a rigorous 10,000 permutations analysis to empirically test the resulting p-value's probability. We repeated the meta-analyses in *APOEϵ4* carrier and non-carrier strata, estimated the combined OR and combined population attributable fraction (cPAF) for both. We also collected CSF and plasma from 500+ individuals and measured nearly 200 human disease-specific biomarkers, including Aβ40, Aβ42, Tau, and pTau, to test their association with the interactions. Our results support the *CLU-MS4A4E* interaction (ADGC: SF=2. 37, p=0. 007; with Cache: SF=2. 71, p=0. 0004) and found a potential dosage effect using the ADGC data between rs11136000:C/C and rs670139:G/T (with Cache: SF=1. 73, p=0. 02). Empirical permutation p-values support the main interaction (ADGC: p=0. 03; with Cache: p=0. 002). The *CD33-MS4A4E* interaction did not replicate. We found an association for *CLU-MS4A4E* in Cache County with*APOEϵ4* negative individuals (SF=4. 75, p=0. 005) that replicates when including Cache in the meta-analysis (SF=2. 08, p=0. 004). The estimated combined OR and cPAF for *CLU-MS4A4E* are 2. 51 and 8. 0, respectively. Association tests between the interactions and biomarker levels are in process. The *CLU-MS4A4E* interaction replicates in a large case-control series, with possible dosage and APOE*ϵ4* effects. The *CLU-MS4A4E* OR is higher than any single AD locus except *APOEϵ4*, *APP*, and *TREM2*. We estimate an 8% decrease in Alzheimer's disease incidence without the *CLU-MS4A4E* risk alleles.

980W

A Genome-Wide Gene-Environment Interaction Study on Dyslipidemia between Genes and Obesity Traits. *M. Kang, J. Sung.* Complex Disease and Genetic Epidemiology Branch, Department of Epidemiology, School of Public Health, Seoul National University, Seoul, Republic of Korea.

IntroductionOne purpose of the human genomic studies is applying the best knowledge to personalized health promotion. A genome-wide gene-environment interaction(GWGEI) is an approach particularly designed for this end. Dyslipidemia is a known risk factor of CVD. Using common types of lipid problems, reduced HDL, elevated LDL, total cholesterol, and TG, we investigated the genes that modify the effect of obesity on dyslipidemia. Despite successful identification of dozens of SNPs related to lipid traits, relatively fewer studies were performed about genes that modify the effects of important risk factors. **Methods-**To examine GWGEI between genes and obesity traits, 12,321 individuals were involved from two population cohorts in Korea: 3,479 from the Healthy Twin Study, 8,842 from the Korean Association Resource(KARE). We conducted an exhaustive GWI search for 4. 1 million SNP markers and their interactions with obesity traits: BMI over 25kg/m2(overweight), WHR above 0. 90 for males, 0. 85 for females(obesity), and waist circumference above 102cm for males, 88cm for females(abdominal obesity). We applied EDGxE, the method using marginal association and case-only test statistics for GxE, and Cocktail methods, where significance of interaction p-value is gradually increased by the initial marginal p-value of the trait GWAS. All two-way interactions, case-only tests, and GxE correlation tests were performed to estimate significance of GxE interactions. **Results**For HDL, we identified 2 loci interacted with WHR near *ASAH2*, and 5 loci interacted with abdominal obesity near *CEP112*. For log-transformed TG, we identified 6 loci interacted with WHR near *APOA5* and *RCAN1*, and 49 loci interacted with abdominal obesity near *BUD13*, *NPR2*, and *LINGO2*. As WHR exceeds the criteria for obesity, about 3. 0% with two minor alleles of *RCAN1* showed steeper rises in TG levels than others. On the other hand, about 3. 5% with two minor alleles of *LINGO2* did not show changes in TG levels by any changes in waist circumference. **Conclusion**Our findings suggest that some individuals are particularly susceptible to abnormal lipid profile even if they are in normal range of obesity indices. The identified SNPs might be used as genetic markers to give a personalized guideline for management of lipid profiles. **Acknowledgements**This work was supported by the Post-Genome Technology Development Program(10050164, Developing Korean Reference Genome) funded by the Ministry of Trade, Industry&Energy(MOTIE, Korea).

981T

Interaction between genotypes at *OPRM1* and *CHRNA5* and an adolescent substance prevention intervention on smoking during high school. *D. J. Vandenbergh[1,2,3], GL. Schlomer[1,4], HH. Cleveland[4], AE. Schink[3], ME. Feinberg[5], JM. Neiderhiser[6], MT. Greenberg[1,5], RL. Spoth[7], C. Redmond[7].* 1) Biobehavioral Health, Penn State University, University Park, PA; 2) The Institute for the Neurosciences, Penn State University, University Park, PA; 3) Molecular and Cellular Integrative Biosciences Program, Penn State University, University Park, PA; 4) Department of Human Development and Family Studies, Penn State University, University Park, PA; 5) Bennett Pierce Prevention Research Center, Penn State University, University Park, PA; 6) Department of Psychology, Penn State University, University Park, PA; 7) Partnerships in Prevention Science Institute, Iowa State University, Ames IA.

Adolescent smoking is reduced by prevention intervention programs, but not all individuals respond. We tested whether genetic risk of smoking, based on genotype at SNP rs1799971 (A118G) in the mu opioid receptor (*OPRM1*) gene, would be reduced in response to a substance use prevention/intervention program. We further tested for epistasis between rs1799971 and rs16969968 (N398D) in the *CHRNA5* gene on smoking, to incorporate the effect of this nicotinic receptor subunit gene, which we have shown previously to interact with the intervention to reduce smoking. Participants (N = 1,688) were randomly assigned to either control condition or a family based intervention in grade 6 and a school-based drug preventive intervention in grade 7. Smoking in the past month was assessed in grades 9–12 using a four-point scale (0 = never smoked, 1 = smoked but not in last month, 2 = one or a few times, 3 = about once a week or more). There was no main effect of genotype at *OPRM1* (coded as presence/absence of the G allele) or intervention, but the interaction was significant (b = –0. 27, p < 0. 01). In addition, we created a polygenic risk score by summing rs1799971 and rs16969968 (coded 1 for A/A genotype, 0 for others). Although there was no main effect of the polygenic risk score, the interaction with condition was significant (b = –0. 27, p < 0. 001). These results suggest that preventive interventions can reduce the genetic risk for smoking, which may involve interaction of multiple addiction-relevant genes.

982F

Genome-wide gene by disease interaction analysis identifies common SNPs at 17q21. 2 that are associated with increased body mass index only among asthmatics. *L. Wang, W. Murk, A. DeWan.* Chronic Disease Epidemiology, Yale University, New Haven, CT.

Background: Asthma and obesity are risk factors for one another, and their comorbidity is a growing medical problem. Although genetic factors are believed to play an essential role in this relationship, to our knowledge, no single nucleotide polymorphism (SNP) has been identified to be associated with both conditions. To explore comorbid susceptibility loci, we hypothesized that the genetic effects on one of the phenotypes is modified by the other phenotype, which provides a novel model supporting a gene by disease interaction analysis in a GWAS framework. **Methods:** We performed a genome-wide gene by asthma interaction analysis for the outcome of body mass index (BMI) in the Multi-Ethnic Study of Atherosclerosis (MESA) study (N=2474 Caucasians, 257 asthmatics) and replicated findings in the Framingham Heart Study (FHS) offspring cohort (N=1408 Caucasians, 382 asthmatics). The tagging SNP in the replicated region rs2107212, was further examined in stratified analyses. **Results:** In MESA, we identified three genetic regions: 7p21. 3, 10q25. 3 and 17q21. 2. Interactions involving SNPs clustered in 17q21. 2 were replicated in FHS. In both MESA and FHS asthmatics, subjects carrying the A allele for rs2107212 had higher BMI than non-carriers, which was not the case for non-asthmatics. Meta-analysis of MESA and FHS showed that the overall odds of being obese increased by 1. 89 fold for each additional A allele in the asthmatic population (p= 4. 3 × 10-5). We further examined BMI change subsequent to asthma diagnosis over a period of 26 years in FHS and demonstrated greater BMI increase among asthmatics compared to non-asthmatics. Asthmatics carrying the A allele at rs2107212 had significantly greater net BMI increase over the 26-year period compared to non-asthmatics. **Conclusions:** We applied a genome-wide gene by disease interaction analysis and found that common variants in 17q21. 2 are associated with asthma-related obesity. Chromosome 17q21 has been repeatedly reported to be associated with asthma. Although the region covering the *ORMDL3*, *GSDMA* and *GSDMB* genes has been the most replicated one, the identified 17q21. 2 keratin cluster has also been shown to be associated with asthma in a previous GWAS study. Our findings could benefit targeted obesity prevention among asthmatics and help elucidate pathways involved in the comorbidity of asthma and obesity.

983W

Metabolites of the one-carbon metabolism and the beneficial effects of the gene-Mediterranean diet interaction. *A. Hadjisavvas[1,2], M. G Kakkoura[1,2], K. Sokratous[1], C. A Demetriou[3], M. A Loizidou[1], K. Kyriacou[1,2].* 1) Electron Microscope/Molecular Pathology, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; 2) The Cyprus School of Molecular Medicine, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus; 3) Neurology Clinic D, The Cyprus Institute of Neurology and Genetics, Nicosia, Cyprus.

Mediterranean diet (MD) has been shown to decrease breast cancer (BC) risk in the Cypriot women of the MASTOS case-control study (Demetriou *et al.* 2012). However, the prominent molecular mechanisms involved in the interplay between MD and BC risk are still poorly understood. One-carbon metabolism (OCM), where the biological processes of DNA methylation and DNA synthesis are interconnected, is a prominent pathway for studying gene-diet interactions and dietary effects. Particularly, we have previously shown that single nucleotide polymorphisms (SNPs) in two OCM genes, the methylenetetrahydrofolate reductase (*MTHFR* 677C>T, rs1801133; 1298A>C, rs1801131) and methionine synthase (*MTR* 2756A>G, rs1805087) could act as effect modifiers on the association between BC and MD (Kakkoura *et al.* 2015). In our study, we selected to quantify a range of one-carbon metabolites in the serum of women, in order to investigate further the importance of the association between SNPs, MD and BC and to gain insights into the underlying molecular mechanisms. Sera from the controls of the MASTOS study were divided into two groups; low and high adherence to the MD, which was previously, determined as a dietary pattern rich in vegetables, fruit, legumes, fish and olive oil (Demetriou *et al.* 2012). A rapid ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) method was used to quantify four one-carbon metabolites of interest (Acquity *I*-Class system coupled to a Xevo-TQD MS system, Waters, UK). These included Vitamins B2, B6, B12 and glutathione. Metabolites were monitored in multiple-reaction monitoring mode and their levels were quantified in 92 samples using external calibration curves. Adjusted coefficient values were determined by linear regression analysis. Vitamin B6 (pyridoxal) was found to be statistically significantly higher (by 0. 4 ng/mL) in serum of women with a high adherence to MD compared to those with a low adherence to MD. No statistically significant associations were observed for the remaining metabolites. Vitamin B6 is a critical coenzyme in the OCM, facilitating the transfer of methyl groups for DNA synthesis and methylation, which are important processes preventing carcinogenesis (Lurie *et al.* 2012). High adherence to the MD was associated with higher serum levels of Vitamin B6, suggesting that Vitamin B6 could be a key player in the OCM via which the MD exerts its anticarcinogenic effects.

984T

African ancestry is associated with intraocular pressure in Latinos. *X. Gao[1], D. Nannini[1], M. Torres[2], Y. Chen[3], K. Taylor[3], J. Rotter[3], R. Varma[2].* 1) Department of Ophthalmology and Visual Sciences, University of Illinois at Chicago, Chicago, IL; 2) USC Eye Institute, Department of Ophthalmology, University of Southern California, Los Angeles, CA; 3) Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA, Torrance, CA.

Intraocular pressure (IOP) is a main risk factor, as well as the only modifiable risk factor, for glaucoma. Racial differences have been observed in IOP measurements with individuals of African descent experiencing the highest IOP measurements when compared to Non-Hispanic Whites and Latinos. The purpose of this study is to examine the relationship between genetic ancestry and IOP in Latinos. We evaluated the genetic ancestry of 3541 Latinos recruited from the Los Angeles Latino Eye Study and assessed the association between IOP and genetic ancestry. Study participants were genotyped using the Illumina OmniExpress BeadChip. We used the program STRUCTURE to estimate individual genetic ancestry. Univariate and multivariate regression analyses were performed to investigate the relationship between genetic ancestry and IOP. African ancestry was significantly associated with IOP in Latinos in our univariate analysis (P = 0. 002). After adjusting for known risk factors, such as age, sex, body mass index, systolic blood pressure, central corneal thickness and type 2 diabetes, African ancestry remained positively associated with IOP (P = 0. 0005). The main association was modified by a significant interaction between African ancestry and systolic blood pressure (P = 0. 037). When stratified by hypertension status, individuals with hypertension experienced a larger increase in IOP with increasing African ancestry. To our knowledge, we demonstrate for the first time that African ancestry and its interaction with hypertension are positively associated with IOP in Latinos.

985F

Reproductive Performance and PON1 L55M Gene Polymorphism in Women Occupationally Exposed to Organophosphate Pesticides. *M. Hema Prasad[1], P. Priyanka[1], B. Chaitanya[1], L. Chandana[1], M. Sailaja[1], P. P Reddy[2], A. Jyothy[1].* 1) Dept of Environmental Toxicology, Institute Of Genetics And Hospital For Genetic Diseases, Hyderabad, Telangana, India; 2) Bhagawan Mahavir Hospital and Research Centre ,Masab Tank,Hyderabad ,Telangana,India.

Background: Organophosphate pesticides (OPs) are a group of toxic insecticides used in agriculture for protection against pests. They persist as environmental contaminants and pose significant risk to human health by causing respiratory disorders, nervous system disorders and certain cancers. Human paraoxonase 1 (PON1) is a lipoprotein-associated enzyme involved in the detoxification of OPs by hydrolyzing the bioactive oxons. PON1 L55M gene polymorphism has been associated with reduced paraoxonase enzyme stability in humans. Earlier studies indicated that exposure to OPs affect reproductive system , which can lead to infertility and an adverse birth out come. In India women agricultural workers are at high risk as they are constantly exposed to OPs during their work. Hence, the present investigation is aimed to study the reproductive performance and association of PON1 L55M gene polymorphism in women occupationally exposed to OPs. **Materials and methods:** In the present study 150 women agricultural workers and 150 age and sex matched healthy controls with same socio economic status were selected. Epidemiological data was collected from exposed and controls using a standard questionnaire. Serum paraoxonase levels were analyzed by ELISA kit method. Gene polymorphisms were detected by using PCR and RFLP. **Results:** Reproductive history of exposed group showed an increased incidence of abortions(13. 10%), still births(3. 77%), premature births(4. 31%), neonatal deaths(3. 41%) and congenital malformations(2. 15%) when compared to controls (4. 88%, 1. 05%, 2. 62%, 1. 41% and 0. 71%) respectively. Serum paraoxonase levels were significantly decreased in the exposed (7. 35 ng/mL) when compared to controls (11. 63 ng/mL). PON1 L55M gene polymorphism results revealed that the frequency of mutant 'A' allele was significantly higher in exposed (34%) when compared to controls (25%) (p=0. 03 & OR=1. 492). **Conclusions:** The epidemiological data revealed an impaired reproductive performance in exposed group when compared to controls which might be attributed to the exposure of women agricultural workers to Organophosphate pesticides. The results also showed significant association of PON-1 L55M gene polymorphism and decreased level of PON-1 in women occupationally exposed to pesticides. **Keywords:** Infertility, Polymorphism, Respiratory System, Nervous System, Cancer.

986W

Obesity-mediated eQTLs identify sex-hormone effects in obese Finnish men. *A. Ko[1,2], A. Gusev[3], B. Pasaniuc[1,4,5], R. M. Cantor[1], E. Nikkola[1], M. Alvarez[1], M. Civelek[6], A. J. Lusis[1,6], F. S. Collins[7], J. Kuusisto[8], A. L. Price[3,9], M. Boehnke[10], K. L. Mohlke[11], M. Laakso[8], P. Pajukanta[1,2,5].* 1) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA; 2) Molecular Biology Institute at UCLA, Los Angeles, CA; 3) Harvard T. H. Chan School of Public Health, Boston, MA; 4) Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA; 5) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, CA; 6) Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA; 7) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 8) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland; 9) Program in Medical an Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA; 10) Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, MI; 11) Department of Genetics, University of North Carolina, Chapel Hill, NC.

Obesity is a serious risk factor for cardiometabolic disease. However, variants identified in the recent meta-GWAS on body mass index (BMI) explain only 2. 7% of BMI variance. Taken together with the continuous increase in the worldwide prevalence of obesity, gene-environment interactions (GxEs) likely play a major role in obesity. Accordingly, we hypothesized that obesity activates regulatory DNA variants in adipose tissue and influences gene expression profiles, leading to allele-dependent responses to obesity. Consequently, these regulatory variants may display context-specific expression signatures that differ in the obese vs. non-obese conditions and can be mapped as context-specific expression quantitative trait loci (cseQTL). To identify obesity cseQTLs, we performed eQTL mapping (FDR<1%) using 7,932,277 SNPs with adipose RNA-sequence data on 17,210 expressed genes in 582 men from the Finnish METSIM cohort. We considered the eQTLs only observed in the obese group, but not in the non-obese or overall groups as obese eQTLs (OB), and vice versa for non-obese (NOB). We tested the OB and NOB eQTL genes (egenes) for replication in an independent METSIM cohort (n=771) using adipose microarray data (FDR<5%). Overall, we discovered 1,631 OB egenes, regulated by 17,851 cis (+/-1Mb) variants, and 840 NOB egenes, regulated by 6,134 cis variants. Of these, 55% of the OB and 41% of the NOB egenes were consistent across RNA-seq and microarrays. In both groups, 95% of the independent eQTLs display a nominally significant interaction (P<0. 05) between variants and obesity status using a likelihood ratio test. Furthermore, the OB egenes exhibit higher cis heritability in the OB than NOB groups (p<0. 001 by permutation), suggesting different local regulatory mechanisms. The expression of the top 64 replicated genes with the highest differential cis heritability explains 4. 6% of BMI variance and 3. 6-9. 3% of other metabolic traits, including lipid and amino acid levels. When searching for potential drivers of context-specificity, we observed that estrogen signaling genes were enriched among OB vs. NOB egenes (P=0. 001), including a GWAS BMI gene, *HSD17B12*, that converts estrone into the most active estrogen, estradiol, in adipocytes. In summary, we demonstrate cseQTLs as a new form of GxE in obesity. Specifically, OB egenes exhibit variant-dependent transcriptional regulation mediated by obesity and involve genes in the estrogen pathway as regulators of male obesity.

987T

Evidence for a causal association between obesity and multiple sclerosis through direct and indirect pathways: Results from the Kaiser Permanente MS Research Program. *L. F. Barcellos[1,2], M. Gianfrancesco[1], X. Shao[1], B. Rhead[1], L. Shen[2], H. Quach[1], A. Bernstein[3], C. Shaefer[2,4].* 1) Div Epidemiology-SPH, Univ California, Berkeley, Berkeley, CA; 2) Kaiser Permanente Division of Research, Oakland, CA; 3) Palm Drive Hospital, Sebastopol, CA; 4) Research Program on Genes, Environment and Health, Kaiser Permanente, Oakland, CA.

Multiple sclerosis (MS) [MIM 126200] is a demyelinating autoimmune disease that involves both genetic and environmental risk factors. While childhood and adolescent obesity has been associated with a two-fold increased risk of MS, a causal relationship has not been determined. Direct and indirect effects of 97 obesity variants based on a recent genome-wide association study were examined using regression-based mediation analysis to estimate the causal association between obesity and MS susceptibility. Each obesity variant was analyzed to measure the direct and indirect effect of having no risk alleles (*a*=0) versus having two risk alleles (*a*=1) on MS susceptibility, controlled for gender, year of birth, ancestry as determined from EIGENSTRAT components, smoking, education, and number of *HLA-DRB1*15:01* alleles, the strongest genetic predictor of MS. The mediator was defined as self-reported body-mass index (BMI) at age 18 or in one's 20's. Analyses were bootstrapped using 100 replications. Participants included non-Hispanic White members of Kaiser Permanente (1,200 MS cases, 10,000 controls). Common and rare variant data for each candidate gene were obtained through genome-wide association profiling and imputation. Results showed that 29 of the 97 obesity variants demonstrated indirect effects on MS through BMI (odds ratios [ORs] ranging from 1. 01 – 1. 05; p<0. 001). Additionally, 15 variants were directly associated with MS; specifically, 7 were associated with an increased risk of disease (rs1460676, rs11727676, rs13107325, rs17405819, rs12286929, rs2650492, rs6091540; ORs ranging from 1. 18-1. 65; p<0. 001), and 8 were protective of MS (rs7599312, rs11057405, rs9540493, rs10132280, rs4787491, rs9925964, rs3810291, rs2836754; ORs ranging from 0. 67-0. 82; p<0. 001), implying that alternate mechanisms are involved in MS susceptibility that are not mediated by BMI. This research for the first time establishes a causal association between obesity and MS and identifies new variants demonstrating a direct effect on disease susceptibility. Findings have implications for interventions targeting a reduction in MS and extend the understanding of how obesity might increase the risk of developing other autoimmune and neurologic diseases.

988F

Identification of novel candidate genes and pathways for Parkinson Disease through gene-based association tests of rare sequence variants. *K. Nuytemans, L. Maldonado, B. J. Rich, K. John-Williams, E. R. Martin, G. Beecham, W. K. Scott, J. M. Vance.* John P Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL.

Previously, we reported association of several biological pathways (KEGG) with Parkinson Disease (PD) risk using GWAS data (based on common variants with low risk effects) (Edwards et al. PlosOne 2010). Alternatively, the presence of rare variants (RVs) with intermediate risk are believed to explain some of the missing heritability in neurodegenerative disorders. Exons are enriched for these rare intermediate-risk RVs, which can be detected through high throughput sequencing. Identification of genes or pathways with accumulation of RV differences in cases versus controls will further our understanding of PD mechanisms. In order to assess association of genes with PD risk, we performed SKAT-O analyses *in* two datasets with RVs filtered on frequency and/or gene location: WES on 396 cases vs 222 controls and follow-up targeted sequencing of 375 genes in 330 cases vs 166 controls. Results were combined across datasets using SeqMeta. Nominally significant genes (p<0. 05) in both datasets include *CASP6* (apoptosis) and *AP5M1* (lysosomal transport). Top genes nominally significant in the SeqMeta analyses include *TUBA4A* (recently identified in amyotrophic lateral sclerosis) and *ABCB6* (mitochondrial transporter). None of these gene associations remain significant after multiple testing adjustment. Subsequent enrichment analyses in KEGG pathways of the gene-based results were performed using WEB-based GEne SeT AnaLysis Toolkit and GeneSet Enrichment Analyses (GSEA). We observed evidence (adjusted p<0. 05) for association with PD risk for the ABC transporters, olfactory transduction, pancreatic secretion and vascular smooth muscle contraction pathways. Over 50% of the genes contributing to the latter two pathways overlap with the larger calcium signaling pathway (associated in original GWAS pathway study and nominally significant here (p=0. 07)), suggesting that these signals might reflect the effect of the calcium signaling pathway. These data indicate we can detect novel candidate genes for PD using the RV gene-based approach; either genes with functions previously linked to PD or other neurodegenerative disorders (e. g. *CASP6*, *AP5M1, TUBA4A*) or genes with novel functions (e. g. *ABCB6*). The pathway analyses confirm the involvement of olfactory function and calcium signaling in PD development, and provide further evidence for ABC transporters contributing to PD. Further studies are needed to assess the overall involvement of ABC transporters in PD pathogenesis.

989W

Mitochondria haplogroup analysis in primary open angle glaucoma suggests an association with haplogroup T2. *J. L. Wiggs[1], D. Navarro[1], L. Shen[1], J. N. Cooke Bailey[2], J. L. Haines[2], L. R. Pasquale[1], X. Gai[1], NEIGHBORHOOD consortium.* 1) Dept Ophthalmology, Harvard Med Sch, MEEI, Boston, MA; 2) Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, OH.

Primary open angle glaucoma (POAG) is an age-related neurodegenerative disease that is a leading cause of blindness worldwide. Several lines of evidence suggest that mitochondrial dysfunction contributes to glaucoma, and may be an important factor influencing susceptibility to retinal ganglion cell apoptosis and optic nerve degeneration. Previous studies investigating mitochondrial DNA variants in glaucoma have reported conflicting results and have been generally underpowered. The purpose of this study is to investigate the association of mitochondrial DNA (mtDNA) variants and haplogroups with POAG in a large case/control cohort. 2871 cases and 2866 controls were genotyped using both the Illumina 660W and the Illumina HumanExome arrays. Combining the mtDNA SNPs from both platforms, 154 variants were available for single variant analysis after quality control (>0. 98 call rate, MAF > 0. 005). One SNP (m. 12705C>T) demonstrated nominally significant association (initial p-value=0. 00086 and Bonferroni corrected P-value= 0. 13) in the single variant analysis using a logistic regression model that included age and sex as co-variants. Using custom tools based on PhyloTree, 15 mtDNA European haplogroups were identified using 294 unique mtDNA SNP positions represented on the two platforms combined. After comparing haplogroup frequencies in cases and controls the T2 haplogroup was significantly associated with POAG (p = 0. 002, Fisher's exact test Bonferroni corrected). These results support a role for mitochondria in glaucoma and suggest that further studies investigating mtDNA and epistatic interactions with nuclear genes would be of interest.

990T

Cytochrome P450 and Matrix Metalloproteinase Genetic Modifiers of Disease Severity in Cerebral Cavernous Malformation Type 1. *H. Choquet[1], E. Trapani[2, 10], L. Goitre[2, 10], L. Trabalzini[3, 10], A. Akers[4], M. Fontanella[5, 10], B. L. Hart[6], L. A. Morrison[7], L. Pawlikowska[1, 8], H. Kim[1, 8, 9], S. F. Retta[2, 10].* 1) Center for Cerebrovascular Research, Department of Anesthesia & Perioperative Care, University of California, San Francisco, San Francisco, CA, USA; 2) Department of Clinical and Biological Sciences, University of Torino, Orbassano (TO), Italy; 3) Department of Biotechnology, Chemistry and Pharmacy, University of Siena, Siena, Italy; 4) Angioma Alliance, Durham, North Carolina, USA; 5) Department of Neurosurgery, Spedali Civili and University of Brescia, Brescia, Italy; 6) Department of Radiology, University of New Mexico, Albuquerque, New Mexico, USA; 7) Departments of Neurology and Pediatrics, University of New Mexico, Albuquerque, New Mexico, USA; 8) Institute for Human Genetics, University of California, San Francisco, California, USA; 9) Department of Epidemiology and Biostatistics, University of California, San Francisco, California, USA; 10) CCM Italia research network (www. ccmitalia. unito. it).

Background: Familial Cerebral Cavernous Malformation Type 1 (CCM1) is an autosomal dominant disease caused by *CCM1* mutations, and characterized by multiple brain lesions. CCM lesions manifest across a range of different phenotypes, including wide differences in lesion number, size and susceptibility to intracerebral hemorrhage (ICH). Oxidative stress plays an important role in cerebrovascular disease pathogenesis, raising the possibility that inter-individual variability in genes related to oxidative stress may contribute to the phenotypic differences observed in CCM1 disease. Here, we investigated whether candidate oxidative stress-related cytochrome P450 (CYP) and matrix metalloproteinase (MMP) single nucleotide polymorphisms (SNPs) grouped by superfamily, family or gene, or analyzed individually influence the severity of CCM1 disease. *Methods:* Clinical assessment and cerebral susceptibility-weighted magnetic resonance imaging were performed to determine total and large (≥5 mm in diameter) lesion counts as well as ICH in 188 CCM1 patients harboring the founder *CCM1* 'common Hispanic mutation' (CCM1-CHM). Samples were genotyped on the Affymetrix Axiom Genome-Wide LAT1 Human Array. We analyzed 1,122 SNPs grouped by CYP and MMP superfamily, family or gene for association with total or large lesion count and ICH adjusted for age and gender, using the SNP "set-based test" (PLINK v1. 07) which corrects *P*-values for the multiple SNPs tested within a set. SNPs bearing the associations were then analyzed individually. *Results:* The CYP superfamily was borderline associated with total lesion count (*P*=0. 057) and large lesion count (*P*=0. 088) in contrast to the MMP superfamily. The CYP4 and CYP8 families were associated with either large lesion count or total lesion count (*P*=0. 014), and two other families (CYP46 and the MMP Stromelysins) were associated with ICH (*P*=0. 011 and 0. 007, respectively). *CYP4F12* rs11085971, *CYP8A1* rs5628, *CYP46A1* rs10151332, and *MMP3* rs117153070 single SNPs mainly bearing the above-mentioned associations were also individually associated with CCM1 disease severity. *Conclusions:* Overall, our candidate oxidative stress-related SNP set approach outlined CYP and MMP families and identified suggestive SNPs that may impact the severity of CCM1 disease, including the development of numerous and large CCM lesions and ICH. These novel genetic risk factors of prognostic value could serve as early objective predictors of disease outcome.

991F

Identification of NSCL/P candidate genes following knockdown of *crispld2*. *B. Chiquet[1,2], L. Maili[1], R. Plant[1], J. Dyke[4], R. Boyer[4], Q. Yuan[1], E. Swindell[1], A. Letra[1,3], S. Blanton[5], J. Hecht[1,4].* 1) Dept Pediatrics, Univ Texas Med Sch, Houston, Houston, TX; 2) Department of Pediatric Dentistry, UTHouston School of Dentistry, Houston, TX; 3) Department of Endodontics, UTHouston School of Dentistry, Houston, TX; 4) University of Texas School of Dentistry, Houston, TX; 5) John P. Hussman Institute for Human Genomics, University of Miami Health System, Miami, FL.

Orofacial development is a multifaceted process involving many tightly regulated gene signaling pathways. Perturbation of any of these genes can lead to malformations, including cleft lip and palate. We and others have shown that *CRISPLD2* is associated with nonsyndromic cleft lip with or without cleft palate (NSCL/P). Recently we have shown that Crispld2 knockdown in zebrafish alters neural crest cell migration patterns, resulting in abnormal jaw and palate development. To identify genes that are affected by knockdown of Crispld2, RNA profiling was completed comparing RNA expression levels of wild-type and Crispld2 knockdown zebrafish embryo heads at 16 hours. 349 unique genes were found to be differentially expressed. Pathway analysis of these genes identified a network of eight genes previously implicated in orofacial development: *BAG3, CASP8, FGFR1, FOS, HOXb1b, JAG2, KIF1B,* and *MMP2.* Quantitative mRNA expression analysis confirmed the differential expression of *CASP8, FGFR1, FOS, KIF1B,* and *MMP2* transcripts in total body RNA extracted from the anterior half of developing 16h post fertilization zebrafish embryos. We genotyped our nonHispanic white and Hispanic NSCLP datasets with 26 SNPs in these eight genes; data were then stratified by ethnicity and presence or absence of a family history of clefting and analyzed using FBAT and APL. We identified significant associations between *FOS*/rs1046117 and NSCL/P in our nonHispanic white all and negative family history dataset (p=0. 0005) and *MMP2*/rs243836 in our Hispanic positive family history dataset (p=0. 002). Significant haplotype associations were also identified in *FGFR1* and *JAG2* in our nonHispanic white dataset and in *MMP2* for our Hispanic dataset (p<0. 002). Gene-gene interaction analysis was also performed revealing twenty-two significant interactions among *KIF1B-MMP2*, *CASP8-JAG2*, *CASP-FOS*, *HOXA1-JAG2*, *BAG3-FOS*, *BAG3-JAG2*, *FOS-JAG2*, and *JAG2-MMP2* (p≤0. 002). Overall, the strongest associations with NSCL/P were seen with *FOS* and *MMP2* genes. The results of this study demonstrate a unique method of identifying NSCL/P candidate genes. Additionally, they may offer valuable insights towards potential mechanisms through which *CRISPLD2* may function during orofacial development.

992W

SNPs, Linkage Disequilibrium and Transcriptional Factor Binding Sites Associated with Acute Mountain Sickness among Han Chinese at the Qinghai-Tibetan Plateau. *N. Buroker[1], X-H. Ning[2], Z-N. Zhou[4], K. Li[3], W-J. Cen[3], X-F. Wu[4], W-Z. Zhu[5], CR. Scott[1], S-H. Chen[1].* 1) Pediatrics, 356320, University of Washington, Seattle, WA; 2) Division of Cardiology, Seattle Children's Hospital. Institute. Foundation, Seattle, WA USA; 3) Lhasa People Hospital, Tibet; 4) Laboratory of Hypoxia Physiology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, China; 5) Center for Cardiovascular Biology and Regenerative Medicine, University of Washington, Seattle, WA USA.

Acute mountain sickness (AMS) occurs in up to 50% of individuals ascending to high altitudes greater than 2600 meters. An AMS Han Chinese and a normal Han group were compared. 17 simple nucleotide polymorphisms (SNPs) within 9 genes that have been associated with AMS were analyzed with respect to linkage disequilibrium (LD) between intra- and intergenic SNP alleles and alterations in transcriptional factor binding sites (TFBS). Included in the study was the angiotensin-converting enzyme (ACE) (rs4340), the angiotensinogen (*AGT*) (rs699) and the angotensin II type 1 receptor (*AGTR1*) (rs5186) SNPs from the renin-angiotension system (RAS) as well as the *GNB3* (rs2071057) SNP from G-protein signaling and a LDL apolipoprotein B (*APOB*) (rs693) SNP. The endothetal Per-Arnt-Sim (PAS) domain protein 1 (*EPAS1*) SNP and two egl nine homolog 1 (*EGLN1*) SNPS (rs480902 and rs516651) from the hypoxia-inducible factor (HIF) oxygen signaling pathway were included. SNPs analyzed in the vascular endothelial growth factor (VEGF*)* signaling pathway are the v-akt murine thymoma viral oncogene homolog 3 (*AKT3*) (rs4590656 and rs2291409), the endothelial cell nitric oxide synthase 3 (*eNOS3*) (rs1007311 and rs1799983) and the (*VEGFA*) (rs79469752, rs13207351, rs28357093, rs1570360 and rs3025039). These SNP alleles alter the TFBS for TF binding. Pair-wise LD was computed between SNPs. An increase in LD occurred in 32 pair-wise comparisons while a decrease was found in 22 pair-wise comparisons between the AMS and controls. Increases and decreases in LD pairs were found within and between signaling pathways and systems indicating the interaction of SNP alleles or potential TFBS from different areas of the genome. The most drastic change in TFBS occurs with ACE (I/D) SNP (rs4340) where the ACE-I allele generates 84 potential TFBS while the ACE-D allele generates only four binding sites. The alteration in TFBS generated by the 17 SNPs is discussed with respect to AMS.

993T

Neural Tube Defect candidate genes are associated with a fumonisin B1 biomarker among at-risk Guatemalan women. *M. Garrett[1], K. Soldano[1], R. Riley[4], O. Torres[5,6], J. Matute[6], K. Voss[4], J. Gelineau-Van Waes[3], S. Gregory[2], A. Ashley-Koch[1].* 1) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC; 2) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC; 3) Department of Pharmacology, Creighton University, School of Medicine, Omaha, NE; 4) Toxicology and Mycotoxin Research Unit, R. B. Russell Research Center, USDA - ARS, Athens, GA; 5) Laboratorio Diagnostico Molecular S. A. , Guatemala City, Guatemala; 6) Centro de Investigaciones en Nutricion y Salud, Guatemala City, Guatemala.

Neural tube defects (NTDs) are congenital defects occurring when the neural tube fails to close during embryonic development. The incidence of NTDs is particularly high in Guatemala, China, and the Transkei of South Africa, where maize is a dietary staple. Fumonisin B1 (FB1), a mycotoxin produced by the mold *Fusarium verticillioides* which contaminates corn, may contribute to this increased risk (Marasas et al. , 2004). FB1 inhibits ceramide synthase causing depletion of complex sphingolipids and increased production of the sphingolipid metabolites sphingosine 1-P (So1P) and sphinganine 1-P (Sa1P), which can be used as biomarkers of FB1 exposure (Merrill et al. , 2001). To investigate genetic mechanisms disrupting sphingolipid biosynthesis, we ascertained women from Guatemalan departamentos with high and low FB1 exposure. FB1 was measured in urine; So1P and Sa1P levels were measured in blood. DNA was extracted from blood spots and 259 single nucleotide polymorphisms (SNPs) in 25 genes that are either known NTD or sphingolipid metabolism genes were genotyped. Quality control (QC) was performed with PLINK (Purcell et al. , 2007) and principal component (PC) analysis was run using EIGENSOFT (Patterson et al. , 2006). After QC, 627 samples and 253 SNPs remained. Linear regression in PLINK was used to test for association between each SNP and sphingolipid metabolites (So1P and Sa1P), controlling for age, departamento, two PCs, and FB1. False-discovery rate (FDR) q-values were calculated in SAS (SAS Systems, Cary, NC). rs2296783 in TCF7L2 was significantly associated with Sa1P, such that for each additional T allele, Sa1P concentration decreased (FDR q=0. 0311). Five SNPs in TCF7L2, S1PR3, TEAD1, DAAM1, and VANGL1 were nominally associated with Sa1P levels (FDR q-values < 0. 2). TCF7L2 is involved in blood glucose homeostasis, common SNPs in which have been associated with type 2 diabetes (Grant et al. , 2006). However, several TCF7L2 splice variants have been identified that disrupt the wnt signaling pathway (Prokunina-Olsson et al. , 2009), which has been heavily implicated in NTDs. Also, maternal SNPs in TCF7L2 have been associated with increased risk of having an NTD-affected pregnancy among obese women (Lupo et al. , 2012). These data provide increased support for a connection between sphingolipid metabolism (altered by FB1 exposure) and NTDs. Future work is needed to replicate these findings and understand the functional mechanisms leading to these associations.

994F

Regulatory annotations implicate thousands of independent disease-associated variants. *A. K. Sarkar[1,2], L. D. Ward[1,2], M. Kellis[1,2].* 1) CSAIL, Massachusetts Institute of Technology, Cambridge, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA.

For most complex traits, known genetic associations only explain a small fraction of the narrow sense heritability (the “missing heritability problem”), prompting intense debate on the genetic basis of complex traits. Here, we use epigenomic annotations across 127 tissues and cell types to investigate the role of weakly associated variants in complex disease. Focusing on Type 1 Diabetes (T1D), we show that weakly associated variants are significantly enriched in immune cell enhancers across thousands of independent loci and additionally enriched in accessible chromatin, in T1D–relevant pathways, and near relevant regulatory motifs. These results are robust to controlling for linkage disequilibrium both between weakly associated variants and between non-coding and coding variants. Moreover, weakly associated variants explain the bulk of heritability attributable to all common variants, suggesting a dramatically larger number of loci than currently suspected may underlie T1D and other complex traits.

995W

Cholesterol transport and beta-amyloid regulation polygenic risk scores as predictors of beta-amyloid levels in a sample at increased risk for Alzheimer’s disease. *B. F. Darst[1], R. L. Koscik[2], B. T. Christian[3], A. M. Racine[4], J. M. Oh[4], B. P. Hermann[2,5], C. M. Carlsson[2,4,6], K. Hogan[7], S. C. Johnson[2,4,7], C. D. Engelman[1,2].* 1) Department of Population Health Sciences, University of Wisconsin, Madison, Madison, WI; 2) Wisconsin Alzheimer’s Institute, University of Wisconsin School of Medicine and Public Health, Madison, WI; 3) Department of Medical Physics, University of Wisconsin School of Medicine and Public Health, Madison, WI; 4) Alzheimer’s Diseases Research Center, University of Wisconsin School of Medicine and Public Health, Madison, WI; 5) Department of Neurology, University of Wisconsin School of Medicine and Public Health, Madison, WI; 6) Geriatric Research Education and Clinical Center, Wm. S. Middleton Memorial VA Hospital, Madison, WI; 7) Department of Anesthesiology, University of Wisconsin School of Medicine and Public Health, Madison, WI.

Background: Although large genome-wide association studies have collectively identified thousands of genetic variants associated with complex diseases, individual variants often have very small effects. Polygenic risk scores (PRS) combine the effects of these variants, often increasing power. Use of disease-related biomarkers can further increase power. For example, accumulation of beta-amyloid (Ab) between nerve cells in the brain is a hallmark of Alzheimer’s disease (AD) pathology and may be used as a biomarker of AD. Methods: Participants were from the Wisconsin Registry for Alzheimer’s Prevention, a longitudinal family-based study of initially asymptomatic adults enriched for a parental history of AD. Using the genes associated with AD in the International Genomics of Alzheimer’s Project’s (IGAP) meta-analysis, we identified clusters of genes that grouped into pathways involved in Ab deposition: Ab regulation (*SORL1*, *PICALM*, *ABCA7*, *CLU*, and *APOE*) and cholesterol transport (*ABCA7*, *CLU*, and *APOE*). Two pathway-specific weighted PRS were formed using the single most significant variant in each gene in the IGAP meta-analysis. Mixed models were used to assess whether each PRS was associated with cerebral Ab deposition in eight regions-of-interest measured using Pittsburgh compound B positron emission tomography (PET) in 168 individuals, and with cerebral spinal fluid (CSF) Ab42/Ab40 levels in 111 individuals. Results: A higher cholesterol PRS was associated with greater PET Ab (P=. 003) and lower CSF Ab (P<. 0001), both indicating a higher risk for developing AD. *APOE* alone was a significant predictor of PET Ab (P=. 01) and CSF Ab (P<. 0001) and, as expected, removing *APOE* from the cholesterol PRS attenuated the associations (P=. 19 and P=. 02, respectively). Similarly, a higher Ab PRS was associated with greater PET Ab (P=. 002) and lower CSF Ab (P<. 0001), and removing *APOE* from the Ab PRS attenuated the associations (P=. 21 and P=. 11, respectively). Conclusion: Although the statistical significance between the pathway-specific PRS and Ab levels are largely driven by *APOE*, there is consistent evidence for a combined effect of variants in other genes with elevated Ab deposition, despite the relatively small sample size. Cholesterol transport and Ab regulation PRS may be informative in explaining variations in Ab levels. Moreover, the use of pathway-specific PRS to examine specific biomarkers of disease is a promising new use of genome-wide association findings.

996T

Genome-wide gene-set analysis identifies different patterns of genetic sharing across complex phenotypes. *H. Gui[1], J. Kwan[2], P. Sham[1,2,3], S. Cherny[1,2,3], M. Li[1,2].* 1) Centre for Genomic Sciences, The University of Hong Kong, Hong Kong SAR, China; 2) Department of Psychiatry, The University of Hong Kong, Hong Kong SAR, China; 3) The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong SAR, China.

As an important complement to individual SNP analysis and gene-based analysis for a typical genome-wide association study (GWAS), GWAS gene-set analysis has the potential to discover hidden disease susceptibility genes by combining statistical evidence with biological knowledge. Recently a gene-set analytical tool 'MAGMA' has been developed to handle polygenic traits using more reasonable and powerful competitive test. Here we adopt both classical gene-set enrichment analysis (hypergeometric test implemented in KGG; http://statgenpro.psychiatry.hku.hk/limx/kgg/) and MAGMA on GWAS summary statistics of six diseases or traits (Crohn's disease (CD), rheumatoid arthritis (RA), schizophrenia (SCZ), bipolar disorder (BPD), low-density cholesterol (LDL) and high-density cholesterol (HDL) level) to look for pathways/gene-sets important for their normal functioning or abnormal pathogenesis. Though no gene-set was significantly associated with two psychiatric diseases, we found a few gene-sets enriched with susceptibility genes for other four phenotypes. Interestingly those LDL/HDL shared gene-sets involving in lipid metabolism or transport are mainly due to pleiotropic apolipoprotein genes for both phenotypes; however, those CD/RA shared gene-sets are ascribed to different phenotype-specific genes which are all important to immune response. Our study reveals that genetic sharing at advanced gene-set level can sometimes provide better perspective to explain disease comorbidity.

997F

Integrated molecular phenotyping in chondrocytes identifies genes and pathways disrupted in osteoarthritis. *J. Steinberg[1], G. Ritchie[1,2], T. I. Roumeliotis[1], A. L. A. Binch[3], R. Coyle[1], M. Pardo[1], C. L. Le Maitre[3], J. Choudhary[1], J. M. Wilkinson[4], E. Zeggini[1].* 1) Wellcome Trust Sanger Institute, Cambridge, UK; 2) European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge, UK; 3) Biomolecular Sciences Research Centre, Sheffield Hallam University, Sheffield, UK; 4) Department of Human Metabolism, The Mellanby Centre for Bone Research, University of Sheffield, Beech Hill Road, Sheffield, UK.

Osteoarthritis (OA) is a degenerative joint disease that affects over 40% of individuals over the age of 70, and is a leading cause of pain and loss of physical function. OA is a disease of cartilage degeneration, but the underlying molecular changes are poorly understood. There is no treatment for OA; instead, disease management targets pain and joint stiffness, and culminates in joint replacement surgery. To gain insights into the molecular etiology of OA, we applied a genome-wide multi-omics approach based on DNA methylation, gene transcription and relative protein abundance. We compared chondrocyte samples extracted from affected and relatively healthy articular cartilage from the joints of 12 OA patients. Importantly, we aimed to identify genes and biological pathways that show association on several molecular levels. In the first step, we identified 271 differentially methylated regions containing 296 unique genes, 349 differentially expressed genes, and 209 proteins with evidence for differential abundance. Integration analysis highlighted genes that are consistently affected, including *AQP1*, *CLEC3B* and *COL1A1*, coding for aquaporin-1, tetranectin and a collagen protein, respectively. The carbonic anhydrase inhibitor Acetazolamide is a known inhibitor of aquaporin-1, suggesting a possible therapeutic route. The thrombolytic agent Tenecteplase, an engineered form of fibrin-specific tPA, binds to tetranectin; the consequences of this binding in OA warrant investigation. We carried out gene set enrichment tests for Gene Ontology (GO), and separately, KEGG and Reactome gene sets, in all three data types. For each gene set, we combined the three association *p*-values using the geometric mean, and obtained an empirical *p*-value for the combined statistic. We identified several consistently implicated biological pathways converging on extracellular matrix organization, collagen formation, cell signaling or adhesion, and skeletal system development. We found that several of these gene sets also show evidence for association from OA GWAS data, although not more so than some other gene sets. These results provide a first view of the integrated molecular landscape underpinning OA development in primary chondrocytes, point to potential therapeutic avenues, and highlight the translational potential of integrated functional genomics.

998W

Identification of disease-specific eQTLs in asthma and COPD cases but not healthy controls. *D. C. Croteau-Chonka[1], M. van den Berge[2], P. J. Castaldi[1], C. P. Hersh[1], V. J. Carey[1], E. K. Silverman[1], A. Spira[3], P. Zanen[4], C. Wijmenga[5], H. M. Boezen[6], H. J. M. Groen[2], W. Mali[7], G. H. Koppelman[8], D. S. Postma[2], S. T. Weiss[1], B. A. Raby[1].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 2) University of Groningen, University Medical Center Groningen, Department of Pulmonology, Groningen, The Netherlands; 3) Division of Computational Biomedicine, Departments of Medicine and Pathology, Boston University, Boston, MA; 4) Department of Pulmonology, University Medical Center Utrecht, Utrecht, The Netherlands; 5) Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 6) University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, The Netherlands; 7) Department of Radiology, University Medical Center Utrecht, Utrecht, The Netherlands; 8) University of Groningen, University Medical Center Groningen, Department of Pediatric Pulmonology and Pediatric Allergology, Beatrix Children's Hospital, Groningen, The Netherlands.

RATIONALE: Epidemiological data supports an etiological link between asthma and chronic obstructive pulmonary disease (COPD), suggesting that COPD susceptibility is increased among asthmatics. We hypothesize that fixed airway obstruction (and ultimately COPD) susceptibility in asthma is conferred by regulatory genetic variants that induce critical perturbations of specific genes in immunoregulatory cells. The aim of this study was to determine the overlap of expression quantitative trait loci (eQTL) identified in asthma and COPD cases and to prioritize candidates for further disease-susceptibility testing and functional characterization. **METHODS:** We performed a fixed-effects meta-analysis of eQTL associations in whole blood samples from three cohorts: Childhood Asthma Management Program [CAMP] (202 asthma cases), COPD Pathology: Addressing Critical gaps, Early Treatment & diagnosis and Innovative Concepts [COPACETIC] (158 COPD cases and 267 controls), and Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints [ECLIPSE] (121 COPD cases). To detect disease-specific eQTL shared between asthma and COPD cases but not observed in healthy controls, we formally tested eQTL identified by meta-analysis for evidence of heterogeneity (Cochran's *Q* test). **RESULTS:** Using this combined cohort of 748 subjects, we identified 4,513 significant eQTLs (444 genes) at a very conservative Bonferronithreshold of $P < 7.7E-08$ among 646,919 SNP-gene pairs tested. For a subset of 135 eQTLs (28 genes), significant associations were noted in the asthma and COPD cases ($P < 0.05$) but not in the healthy controls. Among these were regulatory variants targeting *ASRGL1*, a gene previously associated with the COPD biomarker CC16, and ones targeting *CHI3L2* and *ORMDL1*, which are in the same families as two reported asthma genes. Pathways enriched among shared disease-specific eQTL genes (FDR < 5%) included targets of *MIRNA-34B* (expressed in COPD sputum and associated with emphysema) and of *MIRNA-18A* (associated with asthma). **CONCLUSION:** Preliminary association meta-analysis results suggest the existence of disease eQTLs both common and specific to asthma and COPD with potential links to known genetic risk factors for respiratory diseases and to relevant inflammatory biology pathways. The identification of eQTLs via disease-ascertained study cohorts helps inform our understanding of important genetically-modified disease pathways.

999T

A cell type specific gene expression signature of Alzheimer's disease pathogenesis from meta-analysis of publically available microarray data. *J. Chen, A. Zhang.* Verge Analytics Inc. , Mountain View, CA.

Alzheimer's disease (AD) is the leading cause of dementia worldwide and a growing public health concern. However, the key molecular pathways and gene networks involved in AD are not clearly understood. In order to gain a clearer understanding of disease pathogenesis, we performed a meta-analysis of 31 published gene expression studies in AD patients, mouse models, and cell models, across several cell types and microarray platforms. The raw data from each study was preprocessed (background subtraction and normalization) and quality controlled. Differentially expressed genes between the AD and control conditions were identified for each condition. Using Weighted Gene Coexpression Network Analysis (WGCNA), we found several gene network modules correlated with disease and that were preserved between multiple experiments. We then leveraged data from the International Genomics of Alzheimer's Project (IGAP) genome-wide association meta-analysis to predict the causality of the identified gene network modules. Using these approaches, we derived a gene network signature of Alzheimer's disease through meta-analysis of raw microarray data.

1000F

Promoter polymorphisms in Wnt signaling pathway are associated with tuberculosis risk in Chinese subjects. *X. Hu, J. Y. Zhang, Z. Z. Zhao, X. B. Song, X. J. Lu, B. W. Ying, L. L. Wang.* West China Hospital, Chengdu , China.

Introduction: Compelling studies have indicated that Wnt signaling pathway plays an important role in the development of tuberculosis, however, the exact mechanism remains unknown. For the first time, we took a pathway based candidate gene approach in Chinese population to investigate the associations of promoter polymorphisms within Wnt signaling pathway with tuberculosis. **Methods:** Twenty-three single nucleotide polymorphisms (SNPs) from promoter regions of five key genes in the Wnt signaling pathway (Wif1, Dkk1, Sfrp1, Wnt3a, Wnt5a, Ctnnbl) were selected via searching the dbSNP database. The genotype analyses were conducted by the Massarray method for 973 unrelated Chinese subjects. Inflammatory markers C-reactive protein (CRP), erythrocyte sedimentation rate (ESR) and hematocrit (Hct) were also detected using IMMAGE® 800 and Test® 1. Gene mRNA expressions were measured by real-time PCR on a subset of 416 subjects whose buffy coat fractions were available. Logistic regression analyses were conducted to test the associations of polymorphisms with tuberculosis risk and inflammatory markers. **Results:** The study group comprised of 478 tuberculosis patients and 495 control subjects, and no significant differences in age and sex were observed (p=0. 105, 0. 343, respectively). Genotype distributions of all SNPs agreed with Hardy-Weinberg equilibrium (p all > 0. 05). Rs3864004, rs9859392 and rs9870255 of Ctnnb1 gene and rs4736958 in Sfrp1 gene were associated with decreased tuberculosis risk, from allele and genotype distributions and dominant model (p all <0. 05). The T allele and recessive mode of rs7832767 in Sfrp1 may increase the risk of tuberculosis [OR = 1. 26(1. 03-1. 53), 1. 99(1. 25-3. 16), p=0. 022, 0. 004, respectively]. MRNA expressions of Ctnnb1 and Wnt3a in tuberculosis cases were significantly higher than those in healthy controls (15. 21 ± 3. 10 *vs.* 5. 28 ± 1. 09, 0. 09 ± 0. 02 *vs.* 0. 59 ± 0. 16, p both<0. 001, respectively), while the genotype distributions of these target SNPs were not associated with mRNA expression (p all>0. 05). Rs4736958 and rs7832767 related to CRP expression level (p=0. 016, 0. 018), and heterozygous TC patients showed a higher level of CRP. **Conclusions:** Ctnnb1 and Sfrp1 promoter polymorphisms are associated with tuberculosis susceptibility. Further functional studies are warranted to verify our results.

1001W

Mining Longitudinal BMI, as a Cachexia Proxy, for Common Genetic Determinants among COPD and Cancer Cases in the Framingham Heart Study. *M-L. N. McDonald[1], M. Mattheisen[1,2], S. Won[3], M. H. Cho[1,4], M. Hardin[1,4], G. O'Connor[5,6], C. Hersh[1,4], E. K. Silverman[1,4].* 1) Channing Division of Network Medicine, Brigham & Women's Hosp/Harvard Sch Med, Boston, MA; 2) Department of Biomedicine and Centre for integrative Sequencing (iSEQ), Aarhus University, Aarhus, Denmark; 3) Department of Public Health Science, Seoul National University, South Korea; 4) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 5) The National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, MA; 6) Section of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, Boston University School of Medicine, Boston, MA.

Cachexia, a rapid change in weight including the loss of muscle, is an ominous prognostic indicator for many chronic diseases including chronic obstructive pulmonary disease (COPD) and cancer. However, the majority of population-based studies of cachexia employ cross-sectional body mass index (BMI) to classify cases. To test the hypothesis that shared genetic variants influence cachexia in COPD and cancer cases, we analyzed longitudinal BMI data as a proxy for cachexia using data from the Framingham Heart Study (FHS). To achieve this aim, we ran 6 genome-wide association studies (GWAS) within COPD (N=344), all cancer (N=1085), breast cancer (N=233), prostate cancer (N=231), gastrointestinal (GI) cancer (N=231) and lung cancer (N=112) cases. We used a two step procedure by first calculating a BMI slope phenotype using a random slope and intercept linear mixed model accounting for age, sex and change in smoking status as fixed effects. Only participants with BMI measurements at 3 consecutive time points in the Original or Offspring cohorts were included in the analysis. As the second step, generalized estimating equations (GEE) were used to test for association between 418,365 SNPs and BMI slopes. Two SNPs reached a level of genome-wide significance: rs2024759, an intergenic variant, among lung cancer cases (b=-0. 083, *P*=2. 9x10-8) and rs4751240, in *DOCK1,* among GI cancer cases (b=0. 10, *P*=4. 0x10-9). We performed gene-set enrichment analysis of 7 genes (*ASIC2, CDH13, CNTN4, CSMD1, DLG2, ERBB4, FHIT*) common to the top 1% of the 6 GWAS and *DOCK1* to discover gene-sets relevant to cachexia including those involved with steroid hormone pathways. This list of genes was input to GeneMANIA as a query list to reveal a highly connected disease network module via physical interactions, genetic interactions, co-expression and co-localization. In summary, we were unable to demonstrate that shared genetic variants influence changes in BMI among COPD and cancer cases in the FHS. However, we were able to provide evidence for a common disease network module comprising genes relevant to cachexia in COPD and cancer cases.

1002T

Novel pathways identified from blood-based transcriptomic profiling of kidney function and chronic kidney disease in the Framingham Heart Study. *A. Y. Chu[1], Q. Yang[2], S. Hwang[1], D. Levy[1], C. S. Fox[1].* 1) NIH/NHLBI Framingham, Framingham, MA; 2) Department of Biostatistics, Boston University School of Public Health, Boston MA.

Chronic kidney disease (CKD) is a global public health concern and is associated with increased risk for end-stage renal disease, cardiovascular disease and all-cause mortality. Over 50 genetic loci have been identified in genome-wide associations studies of estimated glomerular filtration rate (eGFR), a measurement of kidney filtration used to diagnose CKD. However, these variants explain a small proportion of total heritability of eGFR indicating that additional factors, such as gene expression and DNA methylation may be involved in kidney function and subsequently CKD. We aimed to identify novel genes and pathways associated with kidney function by testing the correlation of gene expression levels measured on the Affymetrix Human Exon ST Array from whole blood of 5,624 Framingham Heart Study participants with cross-sectionally measured eGFR (calculated using serum creatinine level and the Modification of Diet in Renal Disease [MDRD] equation) and CKD (eGFR<60mL/min/1. 73 m2; Ncases=340 and Ncontrols=5,284) and to perform pathway analysis using the Gene Ontology (GO) database using the WebGestalt interface (http://bioinfo. vanderbilt. edu/webgestalt/). Of the 17,873 transcripts tested, 219 were correlated with both eGFR (false discovery rate [FDR]<0. 01) and differentially expressed in CKD (FDR<0. 01) including *PIP5K1B*, a known eGFR locus from GWAS. The majority of pathways identified by analysis of the GO database (all padj[multiple testing adjusted] <0. 01) were involved in renal system development, and macromolecule modification and metabolism. However, additional pathways related to DNA damage and cell cycle regulation (padj<0. 01) were also significantly enriched for genes correlated with kidney function and CKD. When the list of 219 genes was stratified by direction of differential expression with CKD status, the 157 genes with higher expression levels among those with CKD were enriched in pathways involving protein metabolism and modification (padj<0. 01) whereas the 62 genes with lower expression levels among those with CKD were enriched in pathways involving metabolism of organic and aromatic compounds (padj<0. 01). Our results suggest that in addition to pathways for kidney development (the organ's structural framework) and processing and filtering xenobiotics (a primary function of the renal system), mechanisms and genes involved in DNA stability and cellular turnover may potentially influence CKD risk.

1003F

Whole exome sequencing of very early-onset inflammatory bowel disease shows both family-based and case/control associations with *CCDC40* and *MAPK12*. *C. J. Cardinale[1], Z. Wei[2], L. Tian[1], L. Lima[1], N. Dawany[3], M. Devoto[4], R. N. Baldassano[3], H. Hakonarson[1], J. R. Kelsen[3].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Computer Science, New Jersey Institute of Technology, Newark, NJ; 3) Division of Gastroenterology, Children's Hospital of Philadelphia, Philadelphia, PA; 4) Division of Genetics and Molecular Biology, Children's Hospital of Philadelphia, Philadelphia, PA.

We performed whole exome sequencing on 309 samples including 118 probands and 191 parents and family members in a trio setting for very early-onset inflammatory bowel disease (VEOIBD), that is, IBD with an onset before age 5 years. GATK was employed for multi-sample variant calling for these trios, followed by rigorous QC, which yielded a set of 292,084 variants with high quality. We then performed the transmission disequilibrium test (TDT) for these variants to identify overtransmission of non-synonymous SNVs. We found statistically significant nsSNPs in *PZP*, *SLC7A6OS*, *PCSK5*, and *MAPK12* ($P < 0.001$, $\alpha 2$). In a complimentary analysis we performed case/control association by utilizing a healthy control cohort of 253 exomes for prioritizing nsSNPs (genomic inflation = 1. 0). The top hits were two intron-retaining variants in *CCDC40* (coiled coil domain containing 40) a protein involved in the function of motile cilia ($P < 5.19 \cdot 10\text{-}8$). *MAPK12* was also found in the case/control analysis at $P < 4.97 \cdot 10\text{-}5$. *MAPK12* encodes ERK6, a p38 MAP kinase which is part of the intracellular signaling pathway activated by proinflammatory cytokines or physical stress leading to direct activation of transcription factors. We conclude that rare non-synonymous SNVs within the *CCDC40*

and *MAPK12* genes are associated with VEOIBD.

1004W

Mutational burden in candidate genes for Chiari Type I Malformation is associated with the co-occurrence of connective tissue disorders. *K. Soldano[1], M. Garrett[1], H. Cope[1], S. Gregory[2], A. Ashley-Koch[1].* 1) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC; 2) Duke Molecular Physiology Institute, Duke University Medical Center, Durham,NC.

Chiari Type I Malformation (CMI) is characterized by herniation of the cerebellar tonsils below the foramen magnum, resulting in significant neurologic morbidity. Some forms of CMI co-occur with connective tissue disorders (CTD) and may have a unique patho-mechanism due to cranio-cervical instability. Thus, CMI with CTD (CTD+) may have different genetic risk factors than CMI without a history of CTD (CTD-). Based on previous whole genome linkage analyses (Markunas et al. , 2013 and 2014), we prioritized 28 CMI-relevant candidate genes located under linkage peaks for deep sequencing to identify putative functional variants. The TruSeq Custom Amplicon assay (Illumina, San Diego, CA) was used to amplify coding and highly conserved regulatory regions of these genes. The resulting 1528 amplicons were sequenced on an Illumina MiSeq. Primers were trimmed from raw reads using FASTX-Toolkit and reads were aligned using BWA (Li and Durbin, 2009). Variants were called and quality filtered according to best practices in GATK (Van der Auwera et al. , 2013) and were annotated using ANNOVAR (Wang et al. , 2010). 801 variants remained after processing. To determine if rare variants (MAF < 5%) were associated with CTD status, we used the gene-based optimal sequence kernel association test (SKAT-O; Wu et al. , 2011; Lee et al. , 2012). CTD+ patients were more likely than CTD- patients to have rare variants in *LHX4* (OR=2. 6, p=0. 02) while CTD+ patients were less likely than CTD- patients to have rare variants in *MAFF* (OR=0. 24, p=0. 03). Restricting analysis to coding variants revealed an association with *NCOR1*, such that CTD+ patients were less likely than CTD- patients to have rare coding variants (OR=0. 15, p=0. 04). All three genes may play a role in CMI pathogenesis. Mutations in *LHX4* are associated with combined pituitary hormone deficiencies, a small sella turcica, and Chiari malformations (Machinis et al. , 2001; Tajima et al. , 2007; Castinetti et al. , 2008). *MAFF* is a transcription factor and the murine ortholog is expressed in cartilage and bone (Onodera et al. , 1999). *NCOR1* mediates ligand-independent transcription repression of thyroid-hormone and retinoic-acid receptors and is part of a complex which includes histone deacetylases and transcriptional regulators. In summary, we have preliminary evidence that rare variants in CMI linkage regions are differentially associated with CTD+ and CTD- forms of CMI. Replication is warranted to investigate these associations further.

1005T

Identifying genetic variants underlying self-reported composing and arranging. *J. Oikkonen[1,5], T. Kuusi[2], P. Peltonen[1], P. Raijas[3], L. Ukkola-Vuoti[4], K. Karma[2], P. Onkamo[5], I. Järvelä[1].* 1) Department of medical genetics, University of Helsinki, P. O. Box 63, FI-00014 HY, Finland; 2) Sibelius Academy, University of the Arts Helsinki, P. O. Box 30, FI-00097 Uniarts, Finland; 3) Conservatory of Joensuu, Rantakatu 31, 80100 Joensuu; 4) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, P. O. Box 20, FI-00014 HY, Finland; 5) Department of Biosciences, University of Helsinki, P. O. Box 56, FI-00014 HY, Finland.

Composing and arranging music are complex behavioral phenotypes. These creative activities in music have been shown to enhance certain brain areas including prefrontal cortex, premotor areas, amygdala and auditory cortex, but their molecular mechanisms are largely unknown. We wanted to study the molecular background of these activities with genome-wide linkage and linkage disequilibrium methods. We studied music-educated individuals who were asked if they had composed or arranged music. Our material consisted of families and unrelated individuals comprising totally 474 individuals. The participants were genotyped for over 700k SNPs. In the families, we estimated the heritability of composing as 58% and arranging as much as 70%. These heritabilities fall within the range of previously estimated heritabilities for music-related traits. The genome-wide linkage analysis using MERLIN showed evidence for creatively non-active phenotype at 18q21 (LOD score 3. 1), which contains for example cadherin genes. The creatively active phenotypes were only suggestively linked to any genetic region. At chromosome 4q, these regions did however overlap with regions previously linked with musical aptitude. Joint linkage and linkage disequilibrium analysis was performed as 2-stage analysis with families and unrelated individuals. First, the SNPs were pruned with simple association analysis to exclude unlikely markers. Then, the included SNPs were more comprehensively analyzed with Pseudomarker. The joint analysis showed no significant SNPs with the studied traits. However, pathway analysis of the suggestively identified SNPs showed overrepresentation of cerebellar memory functions. The cerebellum has been previously linked to music memory and memory functions have been proposed to be important in musical creativity. This study increases evidence about music phenotypes linked to chromosome 4, warranting more detailed analysis of the region. We also propose memory pathways to be important in creative activity in music.

1006F

Combined Linkage and Association Analyses Identify ANGPT2 Associated with Nocturnal Oxygen Saturation. *H. Wang[1], B. Cade[2], H. Chen[3], R. Saxena[4], K. Gleason[2], T. Feng[1], E. Larkin[5], S. Patel[2], S. Sunyaev[2], X. Lin[3], S. Redline[2], X. Zhu[1].* 1) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Division of Sleep Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 3) Department of Biostatistics, School of Public Health, Harvard University, Boston, MA; 4) Center for Human Genetics Research, Harvard Medical School, Harvard University, Boston, MA; 5) Vanderbilt University Institute for Medicine and Public Health, Vanderbilt Medical Center, Nashville, TN.

Although sleep disordered breathing (SDB) affects a large proportion of the population and is associated with significant cardiovascular and neuropsychiatric morbidity, its genetic determinants are not clear. Overnight oxygen desaturation is a feature of SDB which has significant clinical relevance and is relatively easily measured, and thus scalable for use in large samples. In this study, we performed combined linkage and association analyses of average nocturnal oxygen saturation level during sleep in a well-characterized sample of European American families collected from the Cleveland Family Study, followed up by gene-based association tests in an independent sample as a replication. We identified two rare haplotypes in *ANGPT2* that are significantly associated with the phenotype and account for 5% of phenotypic variation. *ANGPT2* encodes angiopoietin-2 (Ang-2), an endothelial growth factor that modulates vascular and inflammatory responses to lung injury. Our study provides the first evidence to date for variants in *ANGPT2* influencing nocturnal oxygen saturation level.

1007W

Association of asthma with innate immune genetic variants and interactions with early life viral infections. *L. Akhabir[1], A. Eslami[1], A. B. Becker[2], A. J. Sandford[1], P. P. Pare[1], D. Daley[1].* 1) The UBC James Hogg Research Centre, University of British Columbia, Vancouver, British Columbia, Canada; 2) The Children's Hospital Research Institute of Manitoba, University of Manitoba, Winnipeg, Alberta, Canada.

Background: Asthma is a complex disease with a major global burden in health, quality of life and economics. 2. 4 million Canadians (8% of the population) were afflicted with the disease in 2013. The heritability of asthma is high (48 to 79%), and a variety of factors contribute to susceptibility, including genetic predisposition, environmental exposures, lifestyle and interactions between these factors. Viral infections at an early age have been associated with asthma development and exacerbations. Innate immunity is paramount for sensing and fighting pathogens. We previously showed significant interaction of polymorphisms in innate immune genes with early life viral infections in children from a high risk birth cohort, the Canadian asthma primary prevention study (CAPPS) in asthma development at age 7. Objective: The objective of this study is to investigate the main genetic effects and interactions of the innate immune variants with asthma and related phenotypes (atopy and airway hyperesponsiveness (AHR)) in the CAPPS children at age 15. Methods: 434 CAPPS family trios were used in the analyses of 321 variants in 34 innate immunity genes. Genotyping data were obtained from dense genotyping in candidate genes studies. Family-based analyses were performed using general additive allelic likelihood ratio tests as implemented in the UNPHASED program. Case/control analyses for main genetic and viral effects and SNP-virus interactions were performed using logistic regression in R. Results: Genotypes were included in the analyses for samples with call rates >90% and <4 Mendelian inconsistencies. In the family-based analyses, a total of 29 SNPs showed nominally significant associations with asthma, atopy and AHR. *LBP* rs5741812 was significantly associated with both atopy and AHR (p=0. 0076 and 0. 0211 respectively). 9 SNPs showed significant SNP-virus interaction for atopy. The most significant was *IL1R2* rs4851519 with p=0. 00179. Permutation testing is ongoing to obtain empirical p values and confirm these findings. Case control analyses main effects and SNP-virus interactions in unrelated subjects are in progress. Conclusion: We identified significant interactions between SNPs in innate immune genes and viral infection during the first year of life for atopy at age 15. Interactions between *IL1R2* variants and viral infection seem persistent as we previously found a significant SNP-virus interaction for this gene using phenotypes from the CAPPS children at age 7.

1008T

Whole genome linkage study for caries risk in Guatemalans identifies novel genetic regions. *M. Govil[1], N. Mukhopadhyay[1], C. A. Sanchez[1], F. W. -B. Deleyiannis[2], K. Neiswanger[1], J. M. Resick[1], A. R. Vieira[1], A. M. Letra[3], R. M. Silva[3], M. L. Marazita[1].* 1) Craniofac/Dental Gen/Oral Biol, Univ Pittsburgh, Pittsburgh, PA; 2) Department of Surgery, Plastic and Reconstructive Surgery, University of Colorado School of Medicine, Denver, Colorado; 3) Department of Endodontics and Craniofacial Research Center, The University of Texas Health Science Center School of Dentistry at Houston, Houston, Texas.

Dental caries is a substantial global health risk for individuals of all ages. Both genetic and environmental factors contribute to caries risk. In this study, we sought to identify susceptibility genes for lifelong caries risk in families from rural Guatemala with mixed Caucasian and Native American ancestry. Genome-wide non-parametric linkage analysis was carried out for a novel age-adjusted quantitative caries phenotype. For subjects aged 2-60 years, we measured the number of decayed primary and permanent teeth excluding precavitated lesions, and missing or filled teeth. There were 52 individuals with no caries. Caries indices of affected individuals ranged from 1 to 28, with a mean of 7. 32 (SD=6. 36). Indices were adjusted for age using LOESS fitting with a second-degree polynomial. Linkage analysis included 636 individuals from 117 pedigrees, 474 with known phenotype. Genotyping was performed with the Illumina HumanHap 550 SNP panel. Multipoint genome-wide nonparametric quantitative linkage analysis was carried out using the Merlin-regress program. Prior to linkage analysis, SNPs were thinned out on the basis of linkage disequilibrium, and remaining SNPs analyzed as clusters based on a 0. 1 correlation threshold to account for any residual linkage disequilibrium. Significant linkage peaks were observed on chromosome bands 6p22. 1 (LOD=5. 86, near HLA region), 15q21. 1 (LOD=3. 25, *DOUX2*), 18q22. 1 (LOD=14. 71, *CDH7* and *CDH19*) and 22q11. 23 (LOD=9. 72, *MMP11*, *SMARCB1*, *SLC2A10*). The HLA region is associated with periodontitis, diabetes type 1, host defense and inflammatory outcomes. *DOUX2* plays a role in antimicrobial defense at mucosal surfaces. *CDH7* and *CDH19* are calcium dependent cell-cell adhesion glycoproteins. *MMP11* is involved in embryo development, *SLC2A10* belongs to the glucose transporter subfamily, and heterozygous mutations in *SMARCB1* cause MRD15 (mental retardation, autosomal dominant 15) with abnormal delayed dentition as a clinical trait. These findings are distinct from prior caries studies involving cohorts with ancestries different from the current work (Caucasians from Appalachia and Iowa, African-Americans from Pittsburgh and Iowa). This work was supported by NIH-NIDCR R00-DE018085, U01-DE018903, R21-DE016930, R01-DE016148. Visits to Guatemala were in collaboration with Children of the Americas. All analyses were conducted on a Centos 276-node cluster, Indy, supported by the School of Dental Medicine at the University of Pittsburgh.

1009F

Child's *HLA-DRB1* genotype increases maternal risk of systemic lupus erythematosus (SLE): results from the Mother-Child Immunogenetic Study in Autoimmunity (MCIS). *G. I. Cruz[1], X. Shao[2], H. Quach[2], K. Ho[3], K. Sterba[3], J. A. Noble[4], N. A. Patsopoulos[5], M. P. Busch[6], D. J. Triulzi[7], W. S. W. Wong[8], B. D. Solomon[8], J. E. Niederhuber[8], L. A. Criswell[3], L. F. Barcellos[1, 2].* 1) School of Public Health, University of California Berkeley, Berkeley, CA; 2) Genetic Epidemiology and Genomics Lab, California Institute for Quantitative Biosciences (QB3), University of California Berkeley, Berkeley, CA; 3) Rosalind Russell / Ephraim P. Engleman Rheumatology Research Center, Department of Medicine, University of California San Francisco, San Francisco, CA; 4) Children's Hospital Oakland Research Institute, Oakland, CA; 5) Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology, Brigham & Women's Hospital, Boston, MA; Division of Genetics, Department of Medicine, Brigham & Women's Hospital, Harvard Medical School, Boston, MA; 6) Blood Systems Research Institute, San Francisco, CA; 7) Institute of Transfusion Medicine and Department of Pathology, University of Pittsburgh, Pittsburgh, PA; 8) Division of Medical Genomics, Inova Translational Medicine Institute, Falls Church, VA.

SLE [MIM 152700] disproportionately affects women of reproductive age and pregnant patients are more likely to experience flares. Fetal microchimerism (FMC), or the persistence of a small population of cells in the mother, is a natural consequence of pregnancy. Risk of SLE is possibly increased through fetal HLA-antigen molecular mimicry. The causes of SLE are unknown but genetic and environmental factors, including Epstein-Barr virus (EBV) infection, are suspected. The strongest genetic association is with *HLA-DRB1* alleles **03:01, *15:01, *08:01*. We hypothesize that compared to controls, SLE cases are more likely to have children with a) *DRB1*-associated risk alleles and/or b) *DRB1*04:01* that encodes a homologous amino acid sequence to EBV. We investigated mother-child HLA relationships in 218 SLE and 349 control mothers (and their 881 children) from the Mother-Child Immunogenetic Study (MCIS). The MCIS is a study with over 9,000 individuals: cases were recruited at UC San Francisco; controls were recruited from the Blood Centers of the Pacific, the Institute for Transfusion Medicine at the University of Pittsburgh, and from studies at the Inova Translational Medicine Institute (ITMI). Comprehensive MHC region SNP genotyping was conducted using the Illumina MHC, ImmunoChip, and 660K arrays for MCIS participants and whole genome sequencing for ITMI controls. Classical two-field HLA alleles were imputed using SNP2HLA. Clinical data were abstracted from medical records. We selected mothers of European ancestry using multidimensional scaling and ancestry informative markers to minimize any impact of population stratification. We used logistic regression models to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for the association between SLE and having any children who carry *DRB1* risk alleles or *DRB1*04:01*. Initial findings reveal an increased risk of SLE among mothers with children who carry *DRB1*15:01* (OR 2. 26; 95% CI, 1. 39-3. 66) and **04:01* (OR 1. 73; 95% CI, 1. 14-2. 64), both adjusted for maternal genotype. Furthermore, we observed a stronger association between children who carry *DRB1*15:01* and the SLE complication lupus nephritis compared to controls (OR 2. 75; 95% CI, 1. 12-6. 76, n=383). These findings support the hypothesis that a child's genotype influences a mother's risk of disease, independent of the mother's genotype. This is the first study to demonstrate an association between a child's *DRB1* genotype and risk of SLE in the mother.

1010W

Gene-based analysis in a family-based sample using Generalized Least Square(GLS) approach identified novel IBD locus supported by strong eQTL evidence. *D. Li, T. Haritunians, X. Yan, S. Targan, D. McGovern.* Cedars-Sinai Medical Center, Los Angeles, CA.

More than 160 genetic loci have been identified in Inflammatory Bowel Disease (IBD), mostly via single SNP analysis. Gene-based analysis, which combines signals from all the SNPs in a gene, might provide additional insights to this complex disease. With the complicated and related data structure, family-based samples remain a challenge for gene-based analysis. Previously we proposed a Generalized Least Square (GLS) based approach for gene-based analysis in family-based samples to address this challenge, with its power validated in simulated GAW18 data. Here we applied this approach to an IBD cohort with complex family structure which consists of 3812 IBD cases and 7854 non-IBD controls, all genotyped by ImmunoChipv1. In addition to the known IBD genes such as IL23R and NOD2, we identified a novel IBD association with LRRC16A, which contains 26 common and rare SNPs with a gene-based p-value of 1. 68E-6, significant for the estimated ~25000 genes in the genome. The gene-based result is driven by multiple independent SNPs with weak effects in LRRC6A, and the p-value for the top SNP rs7752195 is 1. 07E-6. Bioinformatic analyses via established eQTL datasets indicate that the top SNP is a strong eQTL for BTN3A2, a gene about 700 kb downstream of LRRC16A. This eQTL is validated in multiple datasets from multiple tissues (In seeQTL, eQTL p = 5. 96E-51(monocyte) and 6. 76E-5(LCL); in SCANdb, eQTL p=8E-9 (LCL); in GeneVar, eQTL p=0. 0025 (LCL); in GTex, eQTL p=0. 004(Colon), 0. 008(blood) and 0. 001(adipose)). Interestingly, the gene BTN3A2 has a p-value of 1. 39E-4 in our gene-based analysis, again with multiple independent weak signals independent of the signals in LRRC16A. BTN3A2 codes for a subunit of CD277, which plays an important role in regulating adaptive immune response. Those results indicate that the proposed GLS approach can be a powerful way for gene-based analysis in family-based samples, and LRRC16A might contribute to IBD via long distance regulation of BTN3A2.

1011T

Analysis of Comorbidities of Multiple Sclerosis Patients using an Electronic Medical Record-linked DNA Biobank. *S. Frodsham[1], J. Denny[2], M. Davis[1].* 1) Department of Microbiology and Molecular Biology, BYU, Provo, UT; 2) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN.

Genome wide association studies (GWAS) have identified relationships between many different genes and diseases. GWAS studies scan whole genomes of many individuals and then associate genetic variants with disease that the individuals have. In contrast, a phenome wide association study (PheWAS), looks at the phenotypes of many individuals and associates those phenotypes with one or more genetic variants. Electronic medical records (EMR) linked to DNA biobanks provide both clinical and genetic data of patients. This study utilizes one such EMR-linked DNA biobank called BioVU from Vanderbilt University. The quantity and quality of data made available to us by BioVU are large enough to provide significant results when doing statistical analysis on the data. Data of insufficient quality were identified and removed using PLINK. Initially, we used ICD-9 billing codes and regular expressions to identify individuals diagnosed with MS and rheumatoid arthritis (RA), a common comorbidity. In this study we analyzed 1,003 patients (28 with RA) with 106 SNPs previously associated with a risk of developing MS were analyzed to see if any variants known to be risk factors for MS were actually associated with simultaneous development of both diseases. While no SNPs in the analysis passed the significance threshold, the number of SNPs in close proximity to the significance threshold warranted further investigation within a wider scope of phenotypes observed in MS patients. By studying a wider variety of phenotypes, we hope to identify diseases with genetic variants that strongly correlate to an increased risk of development of MS. PheWAS analysis was performed using a package within the statistical computing program, R. The study population consists of 1,003 individuals diagnosed with multiple sclerosis, the primary phenotype of interest in this study. ICD9 billing codes contained within the EMRs identified the comorborbidities or secondary phenotypes analyzed in this study. Using these secondary phenotypes, the PheWAS may identify significant genetic variants within this population of MS patients.

1012F

The Role of Mendelian Genes in Complex Disease Risk. *K. P. Shah[1], E. Gamazon[2], J. Denny[2], L. Bastarache[2], H. K. Im[1], D. Nicolae[1], A. Rzhetsky[1], N. J. Cox[2].* 1) Medicine, University of Chicago, Chicago, IL; 2) Vanderbilt Genetics Institute, Nashville, TN.

Mendelian diseases are typically severe and often involve multiple organ systems. Rare DNA mutations in approximately 2,000 genes are known to cause Mendelian diseases. Individuals with Mendelian diseases are also at risk for comorbid complex diseases later in life. Some genes have been found to provide a common genetic basis for Mendelian and complex diseases. Thus, by focusing Mendelian genes, we can gain insights into patterns of disease risk and the underlying biological mechanisms of disease. Using transcriptome data from the GTEx pilot project, we show that Mendelian genes tend to be both more broadly and more highly expressed across a wide range of tissues, but have less inter-individual variability in expression. Furthermore, we found that prediction of gene expression using genomic data performed equally well for Mendelian genes. Using a large electronic medical records database, BioVU, we found that lower predicted expression of Mendelian genes is associated with similar phenotypes to those found in patients with the same Mendelian disease. For example, individuals homozygous for mutations in *PEX19* are known to have Zellweger syndrome (ZS) resulting from disordered peroxisome biogenesis. ZS is characterized by hypotonia, distinctive facies, seizures, kidney failure, coagulopathies, boney stippling of patella and long bones, and early death. Similar to ZS patients, we found low predicted expression of *PEX19* in individuals from BioVU was associated with epilepsy (convulsions and partial epilepsy), kidney disease and failure, and fracture of the patella. Thus, even just decreased expression of the gene can result in phenotypic consequences similar to those observed for the Mendelian disease. Higher predicted expression of *PEX19* was associated with several cancer phenotypes, as would be predicted based on prior studies on genes regulating the expression of genes involved in peroxisome biogenesis. Our results suggest that even subtle changes in the expression of Mendelian disease genes have phenotypic consequences that are relevant to human health. Systematically annotating the phenotypes associated with varied expression of Mendelian genes will shed light on the continuum between Mendelian and complex diseases and allow us to gain more general insights into complex disease biology through this particularly well studied class of genes.

1013W

Study of the genetic architecture of diabetic retinopathy in African Americans from a de-identified medical records system. *N. Restrepo[1], R. Goodloe[2], E. Farber-Eger[2], D. Crawford[1,3].* 1) Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Vanderbilt Genomics Institute, Vanderbilt University, Nashville, TN; 3) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH.

Diabetic retinopathy (DR) is the leading cause of blindness in working age adults in the U. S. DR is caused by abnormalities in the microvasculature of the retina and presents in ~80% of type 1 and 40% of type 2 diabetics. Prevalence rates are higher in African American (AA) diabetics (36. 7%) compared to European American (EA) diabetics (24. 8%) which may be explained, in part, by higher rates of diabetes in AA. Still, the overall trend is higher even after accounting for differences in risk factors, suggesting that other population-specific variables are at play. Identification of genes that play a role in DR have been limited and most studies to-date have been performed in EA populations. Our study proposes to explore the genetic architecture of DR in AA, prior studies of which are altogether missing or uninformative. We have performed a genetic association study of DR in AA from the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study. AA DR cases (n=119) and controls (n=434) were identified in BioVU, the Vanderbilt University Medical Center DNA repository linked to de-identified EMR. Cases and controls were identified via a combination of International Classification of Diseases diagnostic codes, Current Procedural Terminology billing codes, and manual review of clinical records. Individuals were genotyped on the Metabochip, a targeted array. Single SNP tests of association were performed for common variants (MAF>0. 05) using logistic regression assuming an additive genetic model adjusted for age, sex, and mean blood serum glucose levels. We did not replicate associations previously described in other studies for SNPs within *AGER* and *TCF712*. Overall, no test was significant after Bonferroni correction. Of interest for future studies is rs7076968 on chr 10, located within the intergenic region of *ZCCHC24-EIF5AL1* (p=7. 60x10-6; OR=5. 54). In the general EAGLE BioVU AA population (n=11,521), this SNP has a MAF of 5. 5%. The MAF was lower in T2D controls (MAF=3. 5%) and notably higher in DR cases (MAF=9. 7%). Additionally, we identified several variants nominally associated at p<10-4 located in genes known to play a role in epithelial and endothelial tight cellular junctions (rs820626: p=5. 11x10-5; OR=2. 28 and rs17062682: p=3. 2x10-4; OR=2. 12). Although underpowered, this study highlights the need for further studies among AA to determine if lack of generalization is due to differences in LD or differences in the genetic risk factors for DR.

1014T

Bioinformatics Infrastructure for the Partners Biobank Initiative. *E. A. Tsai[1,2,3], J. Peirce[4], K. Embree[5,6], A. Brown[1], N. T. Boutin[5,6], S. S. Amr[1,3,7], S. T. Weiss[1,2,3], M. S. Lebo[1,3,7].* 1) Personalized Medicine, Partners Healthcare, Cambridge, MA; 2) Channing Division of Network Medicine, Brigham and Women's Hospital, Boston, MA; 3) Department of Medicine, Harvard Medical School, Boston, MA; 4) Customer Solutions, Illumina Inc. , San Diego, CA; 5) Partners Biobank, Partners Healthcare, Cambridge, MA; 6) Information Systems, Partners Healthcare, Cambridge, MA; 7) Department of Pathology, Brigham and Women's Hospital, Boston, MA.

The Partners HealthCare Biobank is a large research data and sample repository enrolling subjects who receive their clinical care at Massachusetts General Hospital or Brigham and Women's Hospital. IRB-approved institutional investigators have access to data from the patient's Electronic Medical Record as well as additional health information collected in a self-reported survey. A query tool called the Biobank Portal enables investigators to search the Biobank's database based on clinical data, survey data, sample attributes, and consent attributes. DNA samples from 25,000 consented subjects will be genotyped over two years and made available via the Biobank Portal. This resource will enable investigators to study how genetic background influences an individual's health or contribute to disease development and progression. DNA from peripheral blood is genotyped using Illumina Infinium Multi-Ethnic Genotyping Array (MEGA) with 1. 8 million SNPs including exome content (>50,000 loss of function variants) and ethnic variant SNPs; to date more than 5,000 individuals have been genotyped. Here, we describe the bioinformatics infrastructure that processes, stores, and delivers this genotype data to investigators. The iScan instrument and Illumina LIMS generate binary intensity data (IDAT) files, which are converted to binary Genotype Call (GTC) files. Our process 1) performs data quality control (QC), 2) stores data in a scalable Elasticsearch database, 3) delivers genotype data in PED format, 4) calls copy number variations (CNVs) based on the GTC file, and 5) imputes the genotype at additional sites of variation using the 1000 Genomes Phase 3 dataset. We will discuss how to assess the data quality for high-capacity biobank projects, including periodic inspection for trends and deviations and gender-based checks for inverted sample plates or mislabeled samples, and storage of metadata such as QC metrics as well as genotype and CNV data in database format for compression and easy retrieval. Imputation is performed using IMPUTE2, with increased power to impute SNPs in non-Caucasian cohorts due to the ethnic-specific SNP content. CNVs are called based on the intensity data using PennCNV. The genotype data will be delivered to the investigators for association or meta-analysis studies. This bioinformatics infrastructure is scalable to accommodate the large volume of data processing, storage, and retrieval.

1015F

Possibilities in Unbiased Population Based Research Using Neonatal Screening Biobanks. *J. Grauholm[1,3], C. H. Hansen[1,3], M. Bækvad-Hansen[1,3], J. B. Poulsen[1,3], D. M. Hougaard[2,3], M. V. Hollegaard[1,3].* 1) Neonatal Genetics, Statens Serum institute, Copenhagen, Capital Region, Denmark; 2) Danish Center for Neontal Screening, Statens Serum institute, Copenhagen, Capital Region, Denmark; 3) Department of Congenital Disorders, Statens Serum institute, Copenhagen, Capital Region, Denmark.

The Danish Neonatal Screening Biobank (DNSB) has been storing excess neonatal dried blood spots (neoDBS) samples from the neonatal screening since 1982. Currently DNSB holds more than two million individuals, or almost all Danes born since 1982. Denmark has a long tradition and goodwill from the population to collect all sorts of observations throughout life and to store them in large registries. As all Danes have a unique person identifier number (CPR), it is possible to connect registries and biobanks. This allows researchers to study the etiology of numerous diseases in extraordinary detail. The accessible amount of biological material on a neoDBS sample for research purposes is however limited, one to two 3. 2mm disks, which equals 3-6uL whole-blood. Consequently, it is important to adapt low input methods and design protocols that allows generation of high-quality data on low-concentrated samples. Here we present the genetic methods that we are currently using in DNSB studies of common and complex diseases. To gain recommended amounts of *DNA* for NGS and array genotyping, it is necessary to whole-genome amplify the neoDBS extracted DNA. While it is a robust process, it is critical to confirm technical consistency and accuracy in genetic sequencing and array studies. *DNA methylation profiling* of neonatal DBS samples offers an insight to early life epigenetics, and inherited epigenetics. In this early stage of life, environmental exposures to the child have been minimal. Collected without notably bias, and at proximately the same time of life, the neoDBS samples is close of being the perfect sample for studying methylation events of e. g. birth and birth related complications. *RNA* is an important intermediate between DNA and protein, with many functions. RNA array profiling and RNA-seq is possible on neoDBS. We present data that show that RNA is stable on neoDBS for decades, and usable for research. We also present evidence that the blood born markers of gender, as identified by the Genotype-Tissue Expression consortium, remains stable despite the spotting. In conclusion, this abstract shows how neoDBS from DNSB can be used for reliable DNA genotyping, DNA methylation- and RNA- profiling. The technological advancement, and the continuously expansion of neoDBS samples will, in the years to come, increase the number of diseases that can be studied, and hopefully help to better understand the etiology of disease.

1016W

Correlation Between Genetic Ancestry Proportions and Thousands of Common Disease States in Admixed Populations. *J. Jeff[1], D. Park[2], B. Glicksberg[3], G. Belbin[1,3], E. Bottinger[1,4], N. S. Abul-Husn[1,3,4,6], N. Zaitlin[5], E. E. Kenny[1,3,6,7].* 1) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY,; 2) Department of Bioengineering and Therapeutic Sciences Medicine, University of California San Francisco, CA; 3) Department of Genetics and Genomic Sciences,Icahn School of Medicine at Mount Sinai, New York, NY; 4) Department of Medicine,Icahn School of Medicine at Mount Sinai, New York, NY; 5) Department of Medicine, University of California San Francisco, CA; 6) Institute for Genomics and Multiscale Biology,Icahn School of Medicine at Mount Sinai, New York, NY; 7) Center for Statistical Genetics,Icahn School of Medicine at Mount Sinai, New York, NY.

Over 14,000 medical billing codes (ICD-9) are used to classify diseases, injures, health encounters, and procedures. ICD-9 codes are localized in electronic medical records (EMR) for billing purposes but have been recently used to define disease status for thousands of diseases. Phenome-Wide Associations Studies (PheWAS) are a hypothesis-free approach that tests several phenotypes with a single genetic variant. Such approaches have provided insight into novel pleiotropic effects and disease pathways of common diseases. Big data resources such as EMRs, with thousands of possible phenotypes, provide endless opportunities for PheWAS. The Bio*Me* Biobank at Mount Sinai, is a cohort of >32,000 multi-ethnic patients, including 3,705 African Americans (AA) and 5,104 Hispanic/Latinos (HL), with rich phenotype data from EMRs linked to DNA samples. The admixed AA and HL individuals contain varying proportions of African (AFR), European (EUR) and Native American (NA) genetic ancestry, and previous studies have linked genetic ancestry proportion with risk for several complex diseases. To investigate the relationship between ancestry proportion and disease risk we tested estimates of genetic ancestry proportion for an association with thousands of ICD-9 codes in Bio*Me*. Genetic ancestry proportion was estimated from >1 million SNPs and were consistent with published estimates for AFR, EUR, and NA ancestry in AA (82% AFR, 16% EUR, 1. 6% NA) and HL (28% AFR, 54% EUR, 18% NA). To define common disease states, we assigned disease status to each patient for ICD-9 codes where n ≥20, resulting in 1,728 and 1,263 ICD-9 codes in AA and HL, respectively. All tests were stratified by self-reported ancestry and each component of genetic ancestry was regressed on age, sex, BMI, and ICD-9 code. We identified 18 correlations at P≤8. 7E-5in AA or HL. We observed known correlations between AFR ancestry and anemia (b=0. 03, P=2. 7E-9, AA; b=0. 21, P=3. 18E-8, HL), sickle-cell trait (SCT) (b=0. 04, P=0. 01, AA; b=0. 21, P=1. 9E-8, HL) and uterine fibroids (b=0. 03, P=0. 003, AA; b=0. 05, P=3. 3E-8, HL). We also replicated association with NA ancestry and asthma (b=-0. 03, P=1. 8E-9) in HL. We identified significant pleiotropic effects between AFR ancestry and diseases with overlapping features such as acute venous thrombosis (b=0. 04,P=0. 008) and SCT (b=0. 03, P=0. 01). Replication of these findings in an independent cohort is underway, and may point to novel disease etiologies in these populations.

1017T

Discovery of disease-specific eQTLs aids interpretation of GWAS results in systemic lupus erythematosus. *A. Wuster[1,2], R. R. Graham[1], M. J. Townsend[3], T. W. Behrens[1], T. Bhangale[1,2].* 1) Department of Human Genetics, Genentech Inc, South San Francisco, CA; 2) Department of Computational Biology, Genentech Inc, South San Francisco, CA; 3) Department of Immunology Diagnostic Discovery, Genentech Inc, South San Francisco, CA.

A large proportion of GWAS disease associations can be attributed to variants affecting gene expression (expression quantitative trait loci or eQTLs). There is increasing evidence that many eQTLs are not only specific to tissues but also to the presence of certain stimuli, suggesting that some may only be seen in a stimulated or disease state. For example systemic lupus erythematosus (SLE), a complex autoimmune disease, may be regarded as a state in which cells are under stimulation by inflammatory cytokines such as IFN-α Discovering eQTLs specific to SLE and characterizing the eQTL differences between healthy and SLE blood may aid understanding of the disease process, uncover biomarkers, and suggest new drug targets. To this end, we identified eQTLs in whole blood samples from 83 SLE patients of European descent for which we have generated genotyping and RNA-sequencing data. As controls, we used whole blood eQTL data from the GTEx project. We developed an unbiased and computationally efficient approach to discover eQTLs that are specific to the case or control groups. Our algorithm does this by quantifying significant differences in eQTL signal strength between groups. We identified 307 cis-eQTLs that are specific to or significantly stronger in SLE patients compared to healthy controls (FDR < 0. 05), and whose SLE specificity was confirmed in a validation cohort. We also identified 60 eQTLs that were specific to or significantly stronger in controls. One possible mechanism for SLE-specific eQTLs could be differential activity of transcriptional regulators between cases and controls. Supporting this hypothesis, we found that regulators of the 307 SLE-specific eQTL genes, including MIFT, ARID5B, and SMC3, were differentially expressed between SLE patients and controls. Of 98 independent variants identified in SLE GWAS studies to date, 24 were significant (FDR < 0. 05) eQTLs for a gene other than the one reported by the GWAS, including ULK3, USF1, BTN3A2 and TMEM80. 11 of those eQTLs were specific to SLE patients. Our findings indicate that some eQTLs are only present in tissues from SLE patients, and that a sizable proportion of those eQTLs are relevant for the interpretation of other genetic data including GWAS results. Our approach is adaptable to the identification of subgroup-specific eQTLs in other conditions and diseases.

1018F

Enhanced methods for gene expression imputation from genetic variation data. *N. Mancuso[1], H. Shi[2], A. Gusev[*3,4], B. Pasaniuc[*1,2,5].* 1) Pathology & Laboratory Medicine, David Geffen School of Medicine, University of California at Los Angeles, Los Angeles, California, USA; 2) Bioinformatics Interdepartmental Program, UCLA, Los Angeles, California, USA; 3) Departments of Epidemiology and Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA; 4) Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA; 5) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, California, USA.

Genome-wide association studies (GWAS) have discovered thousands of variants that contribute risk to complex traits. Although a large proportion of GWAS risk variants overlap eQTL loci, the underlying mechanisms leading from genetic variation to gene expression to disease risk are still largely not understood. An attempt to bridge this gap is by estimating the aggregate contribution that eQTL SNPs have on complex traits. Ideally this would be accomplished through large-scale studies that measure genetic variation, gene expression as well as disease phenotypes in many individuals. However measuring gene expression in large numbers of tissues and individuals is prohibitive due to cost and specimen availability. An alternative approach is to computationally predict gene expression directly from genetic variation. In this work, we present methods for imputing gene expression from dense genotype data. In particular we explore the efficiency of the best linear unbiased predictor (BLUP), a Bayesian BLUP and various linear models while accounting for population structure. We present an ensemble predictor that incorporates multiple methods to improve gene expression prediction in simulations and real data analyses. *Contributed Equally.

1019W

Metabolomic profiles during an oral glucose challenge in 470 non-diabetic community residents: Associations with clamp-evaluated insulin sensitivity. *C. Nowak[1], J. Sundström1[1], A. Ganna[2], S. Salihovic[1], T. Fall[1], E. Ingelsson[1,3].* 1) Medical Sciences Department, Molecular Epidemiology, Uppsala University, Uppsala, Sweden; 2) Massachusetts General Hospital, Harvard Medical School and Broad Institute, Boston, Massachusetts, USA; 3) Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine, Stanford, California, USA.

Changes in blood metabolite levels during metabolic challenges such as the oral glucose tolerance test (OGTT) can reveal distinct biochemical alterations preceding type 2 diabetes. Our aim was to study 192 circulating metabolites measured using liquid chromatography/mass spectrometry in repeated blood samples during an OGTT in relation to insulin sensitivity. Metabolite profiling was done in 470 non-diabetic men (mean age 70. 6 ± 0. 6 years, BMI 26. 0 ± 3. 3 kg/m2) at time points 0, 30, and 120 minutes during an OGTT, respectively. Insulin sensitivity (M/I ratio, i. e. glucose disposal divided by mean insulin concentration) measured by a hyperinsulinemic-euglycemic clamp was assessed as the dependent variable in linear regression models with metabolite intensities at the three time points as independent variables. Analyses were adjusted for age and sample quality indicators (previous thawing, storage time, possible hemolysis, and whether plasma or serum samples were used). After Bonferroni correction for multiple testing, 33 out of 192 metabolites were significantly associated with M/I ratio. These included seven glycerophospho- and six glycerolipids, six unsaturated and one saturated fatty acid, four glycerophosphoethanolamines, three acylcarnitines, three bile acids, one monosaccharide, one steroid derivative, and propranolol. Graphical comparisons of confounder-adjusted metabolite trajectories following glucose challenge revealed distinct temporal patterns related to insulin sensitivity and metabolite class. Adjustment for BMI confirmed overall trajectory associations with clamp M/I, but altered time point-specific relationships. In particular, adjustment for adiposity reversed most positive associations between fasting lipid metabolite levels with insulin sensitivity to an inverse relationship (i. e. lower metabolite levels were associated with increased insulin sensitivity), but strengthened their statistical significance whilst weakening associations at 30 and 120 minutes after glucose ingestion. Associations of non-lipid metabolites remained largely unaffected by adjustment for BMI. Metabolite trajectories following an oral glucose challenge revealed distinct temporal patterns associated with insulin sensitivity that may help clarify early metabolic changes and biochemical alterations associated with diabetes pathology.

1020T

Assessing the Genetic Predisposition of Education on Myopia: a Mendelian Randomization Study. *G. Cuellar-Partdia[1], Y. Lu[1], P. F. Kho[1], A. W Hewitt[2], H. E. Wichmann[3,4], S. Yazar[5], D. Stambolian[6], J. E. Bailey-Wilson[7], R. Wojciechowski[8], J. J. Wang[9], P. Mitchell[9], D. A. Mackey[5], S. MacGregor[1].* 1) QIMR Berghofer Medical Research Institute, Brisbane, Australia; 2) School of Medicine, Menzies Research Institute Tasmania, University of Tasmania, Hobart, Australia; 3) Helmholtz Centre Munich, Institute of Epidemiology I, Neuherberg, Germany; 4) Institute of Medical Statistics and Epidemiology, Technical University Munich, Germany; 5) Centre for Ophthalmology and Visual Science, Lions Eye Institute, University of Western Australia, Perth, Australia; 6) Department of Ophthalmology, University of Pennsylvania, Philadelphia, PA, USA; 7) Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, MD, USA; 8) Wilmer Eye Institute, Johns Hopkins Medical Institutions, Baltimore, MD USA; 9) Centre for vision Research, Department of Ophthalmology, University of Sydney, Sydney, Australia.

Myopia is the largest cause of uncorrected visual impairments globally and its recent dramatic increase in the population has made it a major public health problem. Educational attainment has been consistently reported to be correlated to myopia. Nonetheless, correlation does not imply causation. Observational studies reporting the correlation of educational attainment and myopia do not tell us if education causes myopia or if instead there are confounding factors underlying the association. In this work, we use a two-step least squares instrumental-variable (IV) approach to estimate the causal effect of education on refractive error, specifically myopia. We used the results from the educational attainment GWAS from the Social Science Genetic Association Consortium to define a polygenic risk score (PGRS) and tested the PGRS in three cohorts of late middle age and elderly Caucasian individuals (*N*=4514). Using the PGRS as an IV, we estimated that each z-score increase in education (approximately 2 years of education) results in a reduction of 0. 91 ± 0. 27 diopters (*P*=6. 80x10-4) in the three cohorts combined. Our estimate of the effect of education on myopia was higher (*P*=0. 01) than the observed estimate in the three cohorts combined (0. 26 ± 0. 03 diopters reduction per education z-score [~2 years] increase). This suggests that observational studies may actually underestimate the true effect. Our Mendelian Randomization (MR) analysis provides new evidence for a causal role of educational attainment on refractive error.

1021F

Bone marrow mesenchymal stem cell molecular signatures in multiple sclerosis. *F. B. S. Briggs[1], M. Cameron[1], P. Wilkinson[1], M. Li[1], S. Planchon Pope[2], J. A. Cohen[2].* 1) Department of Epidemiology and Biostatistics, Case Western Reserve University, School of Medicine, Cleveland, OH; 2) Mellen Center for Multiple Sclerosis Treatment and Research, Cleveland Clinic, Cleveland, OH.

Background: The majority of multiple sclerosis (MS) patients present with relapsing remitting (RR) disease course and most will develop secondary progressive (SP), however factors precipitating transition are unknown. Adult mesenchymal stem cells (MSCs), including bone marrow (BM) MSCs, are immunomodulatory and neuroprotective; however, these properties may be diminished in MS patients. Wehypothesize that molecular signatures in BM-MSCs distinguish MS patients by disease course (RR vs SP) and from healthy controls (HCs). A robust BM-MSC molecular signature can dramatically transform diagnosis, treatment choice and treatment response prediction. **Methods:**BM-MSCs were collected, isolated and cultured using standard procedures. Subjects were non-Hispanic white. Cells from 4 age (+/- 5 yrs) and gender frequency-matched SP and HC (SP/HC) pairs were collected at passage 1. Cells from 3 frequency-matched RR and SP (RR/SP) pairs were collected at passage 3. RNA was isolated using the Qiagen AllPrep procedure. RNA-seq data was generated via an Illumina HiSeq 2500 (paired-end, 100 cycle, 50x106mapped reads/sample); and were aligned using*STAR aligner*and counted with*HTSeq*. Genes with at least 1 count/106for 3 subjects were analyzed using*EdgeR*. Pathway activity of top-ranking genes (by nominal p-values) was evaluated using gene set variation analysis (GSVA). **Results:**In the SP/HC pairs, there was evidence for differential pathway activation in interleukin binding, adaptive immune response, cytotoxic cell activity, apoptotic DNA fragmentation,p53 signaling, JAK/STAT signaling, and IL6, IL12, and IL3 pathways (*P*<0. 05) between SP and HC BM-MSCs. In RR/SP pairs differential pathway activity was observed for interferon-inducible genes, T cell proliferation, caspase regulation, apoptotic DNA fragmentation, and IL2, IL5, IL6, IL7, and IL12 pathways (*P*FDR<0. 001). Close analysis of distinguishing expression signatures reveals a reciprocal relationship between type I and type II interferon responses in RR vs SP BM-MCSs, with SP being more type II interferon response aligned (downregulated (*P*FDR<0. 001): IFI6, IFI27, IFI35, IFIT1, IFIT2, and IFIT3; upregulated (*P*FDR<0. 05): IL7 and IL18R1). This reciprocal relationship may further impact downstream processes, altering T cell activation, apoptosis signaling and effector responses (e. g. caspases) that would have a profound cumulative immunomodulatory effect which may contribute to disease course transition in MS patients.

1022W

A recurrent missense mutation in *CACNA1G*/Cav3. 1 S4 segment alters channel activity in autosomal dominant cerebellar ataxia. *M. Coutelier[1,2,3], I. Blesneac[4], A. Monteil[4], K. Roussel-Ando[5,6], ML. Monin[1], A. Brusco[7,8], A. Castrioto[9], C. Duyckaerts[5,6], A. Brice[1,10], A. Durr[1,10], P. Lory[4], G. Stevanin[1,2].* 1) Sorbonne Université, UPMC Univ Paris 06, UMR S 1127 and Inserm U1127 and CNRS 7225 and ICM, Equipe Brice, Paris, France; 2) Ecole Pratique des Hautes Etudes, Paris, France; 3) Laboratory of Human Molecular Genetics, de Duve Institute, Université Catholique de Louvain, Brussels, Belgium; 4) Institut de Génomique Fonctionnelle, CNRS UMR 5203, INSERM, U 1191, Université de Montpellier, LabEx 'Ion Channel Science and Therapeutics', Montpellier, France; 5) Laboratoire de Neuropathologie Escourolle, AP-HP, Hôpital de la Pitié Salpêtrière, Paris, France; 6) Alzheimer's and Prion Disease Team, UPMC Université Paris 06 UMR S 1127, Inserm, U 1127 and CNRS UMR 7225, ICM, Paris, France; 7) University of Torino, Department of Medical Sciences, Turin, Italy; 8) Città della Salute e della Scienza University Hospital, Medical Genetics Unit, Turin, Italy; 9) CHU de Grenoble, Département de Neurologie, Grenoble, France; 10) APHP, Department of Genetics and Cytogenetics, Groupe Hospitalier Pitié-Salpêtrière, Paris, France.

Inherited cerebellar ataxias (CA) are neurodegenerative disorders, clinically characterized by a cerebellar syndrome, often accompanied by other neurological or non-neurological signs during disease evolution. All transmission modes have been described; in autosomal dominant (AD) forms, more than 30 causative genes are implicated, but the molecular diagnosis remains unknow in about 40% of the patients. Implication of ion channels has long been an on-going topic in the genetics of CA and causative mutations in several channel genes have been recently described. In a large family with AD CA and mild pyramidal signs, without any mutation in previously known CA genes, we aimed at the identification of the causative variant. With a combined approach of linkage analysis and whole exome sequencing, we identified an arginine-to-histidine mutation in the segment S4 of the T-type Cav3. 1 channel protein, encoded by CACNA1G gene. Further screening of 483 index individuals allowed the description of two additional carriers of that same mutation. Electrophysiological experiments were performed in HEK293T cells to compare the properties of the R1715H mutant to the WT Cav3. 1 channels. These results revealed that R1715H channel showed a significant positive shift of current-voltage relationship and the steady-state activation curve, as well as a higher slope factor of the inactivation curve. Taken together, these data establish CACNA1G as a causative gene in AD CA. CACNA1G was a convincing candidate in the pathology as it is highly expressed in the cerebellum. Our study extends further our knowledge of the link between calcium channelopathies and progressive, not only episodic, CA. .

1023T

Identification of candidate genes for IQ discrepancy in extended families with autism using whole exome sequencing data. *A. Q. Nato[1], N. H. Chapman[1], H. K. Sohi[1], J. M. Viskochil[2], H. Coon[2], E. M. Wijsman[1,3,4].* 1) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA; 2) Department of Psychiatry, University of Utah, Salt Lake City, UT, USA; 3) Department of Biostatistics, University of Washington, Seattle, WA, USA; 4) Department of Genome Sciences, University of Washington, Seattle, WA, USA.

Autism spectrum disorders (ASDs) are neurodevelopmental disorders characterized by repetitive behavior and impairments in communication skills and social behavior. Performance IQ (PIQ) is often greater than verbal IQ (VIQ) in individuals with ASD. As a follow-up to an earlier analysis that identified a region of interest (ROI) on chromosome (chr) 15 and other regions in both a previous sample and the sample used here, we are focusing on identifying candidate genes within the chr 15 ROI using whole exome sequencing (WES) data. Our sample here consists of four extended high-risk families with ASD, which ranged from 72 to 202 individuals. The measures used to define IQ discrepancy are as follows: Wechsler block design and DAS pattern construction subtests (BD) to represent PIQ, and Wechsler vocabulary and DAS naming vocabulary and word definition subtests (VOC) to represent VIQ (BD-VOC). The ROI on chr 15 spans ~18 cM (79 cM - 97 cM). We obtained WES data (66 individuals), and identified 14,039 variants on chr 15. 2,588 variants in 229 genes are within the ROI. Within each family, we excluded indels and selected single nucleotide variants (SNVs) which have ≥ 2 copies, resulting in 421 SNVs (range: 254-342). We further selected SNVs that are nonsynonymous (from SeattleSeq 137 annotation) and have minor allele frequency < 0. 05 (from 1000 Genomes European genotype data). This filtration step resulted in 65 nonsynonymous SNVs (range: 12-27) in 41 genes (range: 11-24), where 13 SNVs and 18 genes with nonsynonymous SNVs are observed in 2-4 families. 7 of these SNVs are predicted by PolyPhen to be possibly- or probably-damaging. We used PBAP for file manipulation and then imputed genotype data for the remaining 390 individuals using GIGI based on the available sequence information and the inheritance vectors that were previously used in our linkage analysis. The dosages of both observed and imputed data are used as a measure of gene burden. Using gene burden as a covariate, we are performing variance components linkage analyses in SOLAR, using models allowing for additive effects with dominance in both the major gene and polygenic components, to identify genes that *as covariates* decrease the linkage signal on chr 15, thereby implicating them as the candidate loci. A similar approach will also be performed on other genomic regions with evidence for linkage in these families (1q, 2q, 3p, 7p, and 16q).

1024F

Language impairment in autism spectrum disorder: a sensitive indicator of functioning in genomic copy number changes? *C. W. Bartlett[1], J. F. Flax[2], A. D. Shindhelm[1], B. Green[2], S. Buyske[2,3], L. M. Brzustowicz[2].* 1) Battelle Ctr Mathematical Med, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) Department of Genetics, Rutgers University, Piscataway, NJ; 3) Department of Statistics, Rutgers University, Piscataway, NJ.

Introduction: Autism is a spectrum disorder (ASD) with behavioral, genetic, and medical characteristics varying in magnitude and complexity. Approximately 25% of individuals on the spectrum never achieve functional language. On the other end of the spectrum, individuals can be verbally fluent and lead relatively independent lives depending upon the degree of their social and behavioral issues. Recent research suggests that lower IQ is associated with a greater number of loss-of-function mutations and that these occurred in sporadic cases of ASD as opposed to multiplex families. In this family genetics study of ASD, copy number variants (CNVs) in individuals with ASD were compared between those with minimal language and those who were verbal at least at the phrase level of language. **Methods:** We divided 115 individuals with ASD into: 1) higher and lower functioning groups based on the Autism Diagnostic Observation Schedule (ADOS) and 2) clinical diagnosis of language impairment versus normal language. For ADOS classification, individuals who were at least 5 years old who had only very minimal language were considered lower functioning while children less than five years of age and all other individuals on the autism spectrum were considered higher functioning and received different versions of the assessment. All samples had CNV inferred from SNP array data using PennCNV. We assessed group burden using logistic regression analysis and followed up with linear regression on standardized language scores. To further assess the impact of CNVs on language we performed regression on an additional 294 family members. **Results**: The ADOS-based high/low groups did not show a difference in total numbers, total length, number of deletions or duplications. However, for the clinical diagnosis of language impairment versus language normal groups, total CNV length was different between the groups (P=0. 013). We then assessed if language was a surrogate for functioning in this population. Using all ASD subjects with quantitative language scores, both CNV total and CNV length predicted language (P=0. 012, P=0. 037, respectively). Follow-up using our total family set that includes quantitative language data on an additional 294 samples, the same conclusion held for CNV total (P=0. 037) but not for CNV length (P>0. 05). **Discussion:** Language impairment within autism spectrum disorder appears to be a sensitive indicator of functioning with regard to genomic copy number changes.

1025W

Two microdeletions that segregate independently in a family, suggesting three candidate genes for autism and motor delay at 12q12 and two candidate genes for cerebral palsy at Xp22. 13. *JDJ. Labonne[1,2], LC. Layman[1,2,3], HG. Kim[1,2].* 1) Section of Reproductive Endocrinology, Infertility & Genetics, Department of Obstetrics and Gynecoloy, Georgia Regents University, Augusta, GA 30912, USA; 2) Department of Neuroscience and Regenerative Medicine, Georgia Regents University, Augusta, GA 30912, USA; 3) Neuroscience Program, Medical College of Georgia, Georgia Regents University, Augusta, GA 30912, USA.

With the advent of high resolution microarray technologies, an increasing number of pathogenic copy number variations (CNVs) are being discovered in human patients. Microdeletions in the 12q12 region are rare, in part because of the close proximity to the centromere. We report a family with two microdeletions segregating independently at 12q12 and Xp22. 31. The patient (DGDP289), who possesses only the 12q12 microdeletion, is a 5-year-old boy displaying pervasive developmental disorder (PDD) and autism. He presents with learning disability, severe speech delay as well as gross and fine motor delay. He inherited the chromosomal anomaly from his father, who in turn inherited it from his mother. The patient's father suffers from depression and difficulties in mathematics, while the paternal grandmother has had lifelong learning difficulties. The Xp22. 31 microdeletion found in the patient's 3-year-old brother contains at least five genes, including *STS* and *VCX*. He displays ichthyosis and cerebral palsy. This Xp22. 31 genomic deletion was inherited from his healthy mother, who is a carrier. Cerebral palsy features may be caused by gene(s) deleted in this interval, possibly by *PNPLA4*. The *VCX3A* gene also located at Xp22. 31 has been previously shown to be associated with intellectual disability. It is possible that *VCX3A* may be involved in cerebral palsy due to a position effect if its expression is altered in the patient's brother. We refined the boundaries of the proximal and distal breakpoints of the 12q12 microdeletion and found it contains only four genes, including *SLC2A13*, *LRRK2*, *MUC19* and *CNTN1*. We compared the 12q12 microdeletion with previous reports and 8 CNV cases in the DECIPHER database. We also determined the transcript levels of three candidate genes namely, *LRRK2*, *SLC2A13* and *CNTN1*. The *LRRK2* gene is involved in Parkinson's disease; *SLC2A13* is predominantly expressed in the brain and *CNTN1* acts a ligand for Notch. We propose that the autistic features displayed by patient DGDP289 are caused primarily by *LRRK2*, with *SLC2A13* and *CNTN1* as possible additional contributors to the phenotype.

1026T

Integrative Analysis of Human Protein Complexes Reveals Biochemical Activities and Convergent Mechanisms of Action in Autism Spectrum Disorders. *J. Li[1], Z. Ma[1], R. Malty[4], H. Aioki[4], M. Shi[1], K. Jin[4,5], D. Wall[2], Z. Zhang[5], A. Urban[1,3], J. Hallmayer[3], M. Babu[4], M. Snyder[1].* 1) Genetics, Stanford University, Stanford, CA; 2) Pediatrics, Stanford University, Stanford, CA; 3) Psychiatry & Behavioral Sciences, Stanford University, Stanford, CA; 4) Department of Biochemistry, Research and Innovation Centre, University of Regina, Canada; 5) Banting and Best Department of Medical Research, Terrence Donnelly Center for Cellular and Biomolecular Research, University of Toronto, Toronto.

ASD is hallmarked by its extreme genetic heterogeneity. Recent studies have devised integrative and computational frameworks to infer molecular pathways in ASD. Ideally, however, the direct and physical organization of the genetic elements in ASD should be experimentally determined in a cellular context relevant to the disease. In this study, we systematically analyzed experimentally characterized human protein complexes implicated in ASD. By examining the ubiquitously expressed protein complexes, we identified the subunits associated with those implicated in ASD and observed that the co-complexed subunits operate in fundamental biological processes (e. g. chromatin remodeling intracellular transport) in human and their orthologs are involved in brain development in the mouse. This analysis also identified HDAC1/2 in the NuRD complex, which preferentially interact with and positively regulate the expression of ASD gene candidate orthologs in the embryonic mouse brain. With mass spectrometry analysis in neuronal culture, we purified and proteome-wide identified the subunits co-complexed with five proteins in idiopathic ASD (ANK2, CHD8, CUL3, DYRK1A, POGZ), FMRP in Fragile X syndrome, and HDAC1. The identified proteins formed an interactive network, displayed preferential expression in early fetal brain development, exhibited increased mutational burden for deleterious mutations in ASD patients and were strongly regulated by FMRP and MeCP2 causal for two ASD related phenotypes, Fragile X and Rett syndromes, respectively. Overall, this study reveals biochemically validated protein complexes associated with ASD, defines an ASD-associated protein complex network in an ASD relevant cell type, and suggests a shared molecular basis between the idiopathic and syndromic forms of ASD.

1027F

Effect of Vitamin D supplementation on NFE2L2 or Nrf2 gene expression in multiple sclerosis patients. *r. Amirinejad[1], m. a. sahraeian[2], b. mohamad soltani[1], m. behmanesh[1].* 1) genetics, tarbiat modares uiversity, tehran, tehran, Iran; 2) Tehran University of Medical Sciences.

INTRUDUCTION: Inflammation and oxidative stress are demonstrated a crucial role to promote tissue damage in multiple sclerosis (MS) in early stages. reports confirmed important obligation of anti-oxidative pathways for tissue protection in Inflammatory-nervous diseases like MS, particularly involving the transcription factor nuclear factor (erythroid-derived 2)-related factor 2 (Nrf2). Nrf2 is referred to as the "master regulator" of the antioxidant response, modulating the expression of hundreds of genes, including the familiar antioxidant enzymes, immune and inflammatory responds. Epidemiological studies showed that vitamin D has immune and immunomodulating effects on MS disease . Inverse relationship between exposure to UV and MS disease implicated this hypothesis that improvement in EDSS in RRMS patients treatment may attenuate oxidative stress and prevent the progression of multiple sclerosis through Nrf2 pathway. Mtherial and Method: Blood samples were collected from 20 relapsing remitting Multiple sclerosis patients and levels of vitaminD were measured. 15 vitamin D deficiency patients selected and treated with vitamin D supplemented. *Peripheral blood mononuclear cell*(*PBMC*)was *collected* from whole bloodby *ficoll*density gradient separation method from 15 patients before and 2 months after vitamin D therapy and 15 age and sex-matched controls obtained. RNA from PBMC was isolated using RNX- Plus(sinaclon) and *did*the*DNase treatment* for all RNA samples. three micrograms of total RNA was used in 20 ml of reverse transcription reaction,was used 2-steps RT-PCR Kit (sinaclon). Quantitative real time PCR was performed using evagreen master mix(TAKARA) in Applied Biosystems StepOne™ System. The melting curve of each sample was measured to ensure the specificity of the products. Data were normalized to the internal control GAPDH and analyzed using $\Delta\Delta$Ct method. Result: in this study we observed that the expression of nrf2 gene in RRMS patients after vitaminD therapy 1. 9 fold increased. improvement in Expanded Disability Status Scale was seen. Discussion: Our results implicates the positive role of vitamin D, on EDSS in patients with RRMS. the ability of vitamin D in reducing the oxidative stress performance by upregulation of nrf2 gene may be associated with the reduction of EDSS scores in MS patients. to further assess Nrf2 downstream target genes that have antioxidant response elements need to be studied.

1028W

Association analysis of rs823144 polymorphism of RAB7L1 gene in Parkinson's disease in Iranian population. *B. Emamalizadeh[1], M. Rahimi[1], H. Darvish[1], A. Khaligh[2].* 1) Department of Medical Genetics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 2) Iranian Research Center of Healthy Aging, Medical University of Sabzevar, Sabzevar, Iran.

Recent genome-wide association studies (GWAS) explored some new lociin association with Parkinson's disease. RAB7L1 is an important gene in one of neurological pathways, located in PARK16 locus, and its variants influence the risk for PD. We performed a case control study to examine association between rs823144 SNP in RAB7L1 gene and Parkinson's disease risk. **Methods:**blood samples of 980 patients and healthy controls were collected to analysis the frequency distribution of RAB7L1 polymorphisms (rs823144) using polymerase chain reaction-restriction fragment length polymorphism (PCR – RFLP) method. **Results:**Approximately 6. 3 percent of participants were C/C homozygous, 59. 1 percent were A/A Homozygous and 34. 6 percent were heterozygous. Significant association was found between presence of minor allele (C) and decreased risk of PD development ($P < 0.01$, OR = 1. 342). **Conclusion**: our dataconfirms the association between rs823144 and decreased risk of PD.

1029T

Mapping Genes using a Hidden Markov Model for Bipolar Affective Disorder in Consanguineous Families. *R. S. Harripaul[1,2,3], M. Ohadi[5], N. Moghimi[1,2,3], J. B. Vincent[1,2,3,4].* 1) Molecular Neuropsychiatry & Development (MiND) Lab, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada; 2) Neurogenetics, Centre for Addiction and Mental Dealth, Toronto, Ontario, Canada; 3) Institute of Medical Science, University of Toronto, Toronto, ON, Canada; 4) Department of Psychiatry, University of Toronto, Toronto, ON, Canada; 5) The Social Welfare and Rehabilitation Sciences University, Tehran, Iran.

Bipolar Affective Disorder (BD) is a psychiatric disorder characterized by transitions between depression and mania, including a high rate of suicide (6% over age 20) and self-harm (30-40%). This debilitating condition has no known cause, and both genetic and environmental factors contribute to its complex phenotype. We hypothesize that in rare cases, autosomal recessive mutations contribute to BD. To identify these genetic loci, 34 consanguineous Iranian families were genotyped with Affymetrix 5. 0 Single Nucleotide Polymorphism microarray chips. This genotype information was analyzed with the FSuite analysis pipeline and dCHIP to identify homozygous-by-descent (HBD) regions and a HBD Genome Wide Association study to identify novel recessive loci where risk variants may reside was performed. Whole Exome and Sanger sequencing were used to search for homozygous coding mutations within these HBD regions. Large runs of homozygosity have been identified in BD probands, including a 400 Kb loci that traverses the GRIK6 glutamate receptor gene. We have also identified 48 CNVs of interest that may disrupt candidate genes such as SYN3, SLC39A11 and S100A10. In addition, we looked for Copy Number Variations (CNVs) and 43 large HBD regions were identified. We identified large HBD regions such as the 56 Mb region on chromosome 8 (harboring candidate genes such as IMPA1, IMPAD1), a 10 Mb region on chromosome 17 (including SLC6A4) and a 7 Mb region on 5q35-2-qter including genes DRD1 and GRM6. Rare variants have been identified in more than one family for a number of genes. For instance, for ABCA13, one homozygous nonsense and one homozygous non-synonymous variant were identified in separate families. Some potential variants identified in this study have been previously implicated in Bipolar Affective Disorder, Schizophrenia and Depression while others have no known function and need further characterization.

1030F

Analysis of rs1572931 polymorphism of *RAB7L1* gene in Parkinson's disease in Iranian patients. *A. Khaligh[1], M. Rahimi[2], H. Darvish[2], B. Emamalizadeh[2].* 1) Iranian Research Center of Healthy Aging, Medical University of Sabzevar, Sabzevar, Iran. , Tehran, Tehran, Iran; 2) Department of Medical Genetics, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Genetic association studies have recently made increasing data about the genetic contribution to Parkinson's disease and this has led to significant developments in discovery of pathogenesis of the PD. A large number of differenttypes of polymorphisms, particularly single nucleotide polymorphisms, have been investigated in PD in different populations by association studies. Thers1572931polymorphism of*RAB7L1* gene, located in the promoter region, is recently studied and shown to be strongly associated with reduced risk of PD. **Subjects and methods***:*We examined the association ofrs1572931polymorphism with Parkinson's disease in 480 Patients and 480 normal controls by PCR-RFLP in Iraniansubjects. **Results***:*Asignificant difference in genotype and allele frequencies was observed between patients and controls (OR = 0. 71, 95% CI= 0. 56-0. 90, *p*= 0. 003). The TT genotype and the T allele were both significantly less frequent in PD cases. **Conclusion***:*Our results confirmed the protective effect of thers1572931SNP on Parkinson's disease and replicated the results of previous studies. We suggest further studies in other populations.

1031W

Small CNVs in Major Depressive Disorder. *DL. Nuñez Ríos[1], MC. Lattig Matiz[1], Y. Gómez Maquet[2].* 1) Departamento de ciencias Biológicas. Facultad de ciencias. Universidad de los Andes. Bogotá; 2) Departamento de psicología. Facultad de ciencias sociales. Universidad de los Andes. Bogotá.

Major Depressive Disorder (MDD) is a major public health problem in the overall global population with high prevalence in children and adolescents and a tendency to decrease life expectancy in affected individuals. Although the etiology of depression is still not fully understood, genetic factors play a powerful role in the development of MDD with a heritability of up to 40% as demonstrated by twin studies. Nonetheless, gene identification has been difficult probably due to the low penetrance and multifactorial mode of MDD. Copy Number Variations (CNVs) are chromosome structural variations that include one or more genes and have been associated with schizophrenia, intellectual disability, autism and some cases of MDD. Most studies in psychiatric disorders have reported large CNVs (>100kb) that include recurrent regions such as 15q13. 3, 16p11. 2, 22q11. 2. In the present study, nineteen patients with clinical diagnosis of MDD and seventeen healthy patients were screened by array Comparative Genome Hybridization (aCGH) using Agilent 4x180k platform. We did not find any of the recurrent CNVs reported to date in our cohort; however, we did detect three rare small CNVs (deletion in 4q22. 1, 5p13. 2 and 17q25. 1) that were not present in the control group. The deletion 4q22. 1 (46kb) was present in 11/19 patients (57,8%) and includes NUDT9 [OMIM606022] and SPARCL1 [OMIM606041] genes. SPARCL-1 (secreted protein, acidic, rich in cysteine)-like 1 also known as Hevin and SC1, is a protein associated with neuronal migration and formation, maturation and plasticity of excitatory synapses. Studies in human post-morten nucleus accumbens from depressed patients reported a decrease in SPARCL-1. The 5p13. 2 deletion (38kb) was found in 5/19 patients (26,3%) and involves the alanine glyoxylate aminotransferase 2 (AGXT2[OMIM 612471]) gene which has been associated with psychiatric disorders. The 17q25. 1 deletion (86kb) was detected in 12/19 patients (63,1%) and includes ACOX1 [OMIM 264470] and CDK3 [OMIM 123828] genes. ACOX1 (acyl-CoA oxidase 1, palmitoyl), is related with pathogenesis of neurological disorder. Most CNV studies suggest that genomic disorders of the nervous system are consequence of large CNV, however recent studies have associated small CNV (<100kb) to autism. Here we report a small MDD cohort from an understudied population such as Colombia with three small CNVs that affect genes related with neuronal activity, thus suggesting a role in the pathogenesis of MDD.

1032T

A novel neurogenetic disorder with *STARD9* mutation. *N. Okamoto[1], F. Miya[2], T. Tsunoda[2], M. Kato[3], S. Saitoh[4], M. Yamasaki[5], Y. Kanemura[6,7], K. Kosaki[8].* 1) Dept Medical Genetics, Osaka Med Ctr/Res Inst, Osaka, Japan; 2) Laboratory for Medical Science Mathematics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan; 3) Department of Pediatrics, Showa University School of Medicine, Tokyo, Japan; 4) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 5) Department of Pediatric Neurosurgery, Takatsuki General Hospital, Osaka, Japan; 6) Division of Regenerative Medicine, Institute for Clinical Research, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 7) Department of Neurosurgery, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 8) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan.

[**INTRODUCTION**]There are 45 mammalian kinesin family member (KIF) genes and the encoded proteins are classified into 15 kinesin families that are grouped according to the position of the motor domain in the molecule. The number of KIF genes involved in neurogenetic disorder is increasing. Two of the KIF is specifically enriched at the centrioles, Kif24 at the mother centriole and STARD9/Kif16a at the daughter centriole. STARD9 is a large protein with an N-terminal kinesin motor domain, an FHA phosphoprotein binding domain, and a C-terminal START lipid/sterol binding domain. During cell division, STARD9 is enriched at the daughter centriole. STARD9 is required for pericentriolar material (PCM) cohesion and bipolar spindle assembly. Torres et al. (Cell 2011) showed that STARD9-depleted cells have fragmented PCM, form multipolar spindles, activate the spindle assembly checkpoint, arrest in mitosis, and undergo apoptosis. We identified a mutation in the *STARD9* gene by WES in a patient with severe ID, microcephaly, epilepsy and cortical blindness. [**CLINICAL REPORT]**The 5-year-old female was the first child of healthy and non-consanguineous Japanese parents. She was hypotonic and her developmental milestones were markedly delayed. She experienced a generalized tonic seizure at 6 months old. Brain MRI revealed mildly enlarged lateral ventricles. Physical examination identified microcephaly and dysmorphic features including sparse eyebrow, epicanthal folds, flat nasal bridge, low set ears, short philtrum, high arched palate and prominent Cupid's bow. She showed axial hypotonia. Deep tendon reflexes were absent. Ophthalmological investigation revealed cortical blindness. She could no sit or crawl at 6 years old. She spoke no meaningful words. Her physical growth was also delayed. [**METHODS**]With the approval of our institutional ethics committee, the samples were analyzed using WES. Precise methods have been reported in another article (Okamoto et al. J Hum Genet 2014). [**RESULTS]** The patient was homozygous for a deletion in exon23 of the *STARD9* (c. 11760delC:p. L3920fs,). Her parents were heterozygous for the mutation. This mutation is predicted to eliminate the C-terminal START lipid/sterol binding domain. [**DISCUSSION**]There have been no reports on the association of the *STARD9* mutation and genetic disorders. We suppose that abnormality of *STARD9* causes a novel neurogenetic syndrome.

1033F

Copy Number Change (CNC) in chromosomal region 3p26. 3, is consistently associated with the ADHD phenotype in a large sibship. *N. Kapalanga[1,2], D. Kwiatkowski[1], T. Trudeau[1].* 1) Department of Pediatrics, Schulich School of Medicine, Western University London, Ontario, Canada; 2) Pediatrics/Genetics, GBHS, Owen Sound, Ontario, Canada.

The contribution of genetics to the etiopathogenesis of ADHD has been estimated to be between70 to 80 percent. The evidence for genetic susceptibility to ADHD has been gathered from genome-wide associationstudies, family studies and candidate gene study approaches. While the involvement of genetic factors in susceptibility to ADHD is incontrovertible specific ADHD genes have not been identified. Even in families with clear heritability of the ADHD phenotype specificgenes have not been identified. Only one confirmed candidate gene locationon chromosome 16 has been repeatedly linked to ADHD. In this study we report on a GAIN of 0. 46 Copy Number Change (CNC) in chromosomal region 3p26. 3, that consistently associated with the ADHD phenotype ina largesibship. Ina pediatric clinic, we routinely order array CGH and Fragile X testing in all individuals who present with a history of ADHD associated with other neurodevelopmental comorbidities, in at least two first degree relatives. A family presented at clinic consultationwith seven siblings who all had varying degrees of the ADHD phenotype. Two of the sibling were from a previous marriage and thus had a different father. Momalso had the ADHDphenotype. The Conners' Parent and Teacher Rating Scale – Revised (S) was administered to all seven siblings. Further, array CGH, Fragile X testing and routine CBC and SMA were ordered for each sibling. Six of the seven siblings fulfilled criteria for ADHD. Allsix siblingsfulfilling criteria for ADHD were found to have a GAIN of 0. 46 Copy Number Change (CNC) in chromosomal region 3p26. 3. Mom was was also found to have the same CNC. Thesibling who did not fulfill criteria for ADHD did not carry the CNC. The Fragile X test was negative in all siblings. The CBC and SMA were within normal limits in all siblings. This CNC is predicted to include the promoter of CNTN4. Disruption of this gene has been associated with developmental delay and certain clinical features of the 3p deletion syndrome. In view of the consistent association of this CNC with the ADHD phenotype in this family, it is not inconceivable that a gene in chromosomal region 3p26. 3 is involved in the etiopathogenesis of ADHD. A multigenerational study and testing of extended members of this family will be undertaken. Further, a candidate gene approach focusing on chromosomal region 3p26. 3 could be used to study other ADHD patients to identify an ADHD gene at this region. .

1034W

Missense mutations in *GPT2* are associated with autosomal recessive nonsyndromic intellectual disability. *T. Lobo-Prada[1,2,3,4], S. Uebe[4], H. Sticht[5], A. B. Ekici[4], S. Bogantes[2], A. Reis[4], A. Leal[1,3,4].* 1) School of Biology, University of Costa Rica, San Jose, San Jose, Costa Rica; 2) School of Medicine, University of Costa Rica, San Jose, San Jose, Costa Rica; 3) Neuroscience Research Center, University of Costa Rica, San Jose, Costa Rica; 4) Institut für Humangenetik, Universitét Erlangen-Nürnberg, Erlangen, Germany; 5) Institut für Biochemie, Universitét Erlangen-Nürnberg, Erlangen, Germany.

Nonsyndromic intellectual disability is a heterogeneous disease that affects the central nervous system impairing patients to develop independently. Despite that numerous candidate genes have been involved in the etiology of the disease it is noteworthy that for about half of the patients it is still not possible to establish the cause of the disease. Nevertheless, recently, with the combination of exome sequencing and microarray technology it has become possible to identify rare genetic variants associated with this disease especially in consanguineous families. We identified a large Costa Rican family with four adult siblings, three females and one male affected with nonsyndromic severe intellectual disability, spastic paraparesis and severe language delay whose parents were first degree cousins. After extensive search for common genetic variants, CNVs and microdeletions we performed whole exome sequencing and identified a novel missense mutation, p. G96R, c. G286A in glutamate pyruvate transaminase 2 (*GPT2)* in all affected individuals. *GPT2* encodes a pyrodoxal enzyme involved in the production of neurotransmitter glutamate and is highly expressed in the white matter of brain and cerebellum. This variant is highly conserved in evolution and all in silico prediction algorithms used indicated that it is pathogenic. Protein modeling of GPT2 predicted that the mutation is localized in a loop where the substrate binds to the active site of the enzyme, therefore, suggesting that the catalytic activity is impaired. Recently, a further single distantly consanguineous family with three toddler siblings, two males and one female, with static developmental encephalopathy associated with severe intellectual disability, carrying a homozygous mutation in *GPT2,* p. S153R, c. C459G was reported. With our report of a second case we confirm *GPT2* as a novel cause of autosomal recessive nonsyndromic intellectual disability and support the premise that *GPT2* is highly important for the neurodevelopment of the central nervous system.

1035T

Effects of *DCDC2* and *KIAA0319* in a case-control sample for reading disability in African American and Hispanic American children. *D. T. Truong[1], N. R. Powers[1,2], J. D. Eicher[2], J. R. Gruen[1,2,3], Genes, Reading, and Dyslexia Study.* 1) Pediatrics, Yale University School of Medicine, New Haven, CT; 2) Genetics, Yale University, New Haven, CT; 3) Investigative Medicine, Yale University School of Medicine, New Haven, CT.

Reading disability (RD) is the most prevalent neurodevelopmental disorder diagnosed in school-aged children, and is characterized by a specific and unexpected difficulty in reading ability despite normal intelligence. The etiology of RD is complex with a strong genetic component reflected by heritability estimates ranging from 60% to 87%. Several genomic regions have been associated with RD, but *DCDC2* and *KIAA0319* are the most replicated candidate risk genes. Previously, our lab identified a highly polymorphic compound short tandem repeat within the breakpoints of a 2,445bp microdeletion in intron 2 of *DCDC2* (READ1) that mediates *DCDC2* enhancer activity, and is also strongly associated with RD in clinical studies. In *KIAA0319*, a putative functional SNP (rs9461045) is linked to reduced expression and impaired reading performance. Furthermore, there is evidence of epistasis between variation tagging rs9461045 and alleles 5 and 6 of READ1, independently, that yields non-additive reductions in reading and language performance. These studies have contributed to our understanding of the genetics of RD, but they primarily represent populations of European (EUR) ancestry, with Hispanic American (HA) and African American (AA) populations largely underrepresented. Therefore, it is unclear if genetic variants previously identified in EURs are informative in HA and AA groups. Thus, the purpose of this study is to examine READ1, *KIAA0319* (rs9461045), and their epistatic effects on reading behavior in the Genes, Reading, and Dyslexia (GRaD) study—a unique case-control cohort for RD comprised of 540 African American and 951 Hispanic American children. We performed a single marker analysis examining READ1 alleles previously associated with RD and language impairment in EURs (alleles 3, 5, 6, and deletion) and rs9461045 in *KIAA0319*. We found a significant association with allele 6 (p=3. 85×10-3) and rs9461045 (p=2. 77×10-3) on the Test of Word Reading Efficiency (TOWRE) in AA children. No associations with reading-related performance were observed in HA children. Examination of epistasis between READ1 alleles and rs9461045 on reading and language-related performance in the GRaD study is forthcoming. Results from the single marker analysis indicate overlapping genetic variation associated with reading-related performance across AA and EUR populations, and further support *DCDC2* and *KIAA0319* as risk genes for RD in this sample. .

1036F

Genetic susceptibility loci for late-onset Alzheimer's disease and the prevalence of subjective memory loss in healthy, older adults. *D. A. Carere[1,2], A. Bhushan[3,4], L. B. Chibnik[3,4,5], R. C. Green[2,4], W. C. Willett[3], P. Kraft[1].* 1) Program in Genetic Epidemiology and Statistical Genetics, Harvard T. H. Chan School of Public Health, Boston, MA; 2) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Boston, MA; 3) Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA; 4) Harvard Medical School, Boston, MA; 5) Department of Neurology, Brigham and Women's Hospital, Boston, MA.

Background: Subjective memory complaints (SMCs) in healthy adults have been linked to later development of Alzheimer's disease (AD). We investigated whether*apolipoprotein-E*(*APOE*) genotype and 19 AD-associated single nucleotide polymorphisms (AD-SNPs) are associated with SMCs. **Methods**: Analyses were restricted to Health Professionals Follow-up Study (HPFS) and Nurses' Health Study (NHS) participants with SMC and genotype data. SMCs were measured by 6 yes/no survey questions (e. g. , Do you have trouble finding your way around familiar streets?) in 2008 or 2012, and we computed a count of complaints (SMCscore, 0-6). Non-genetic AD risk factors (e. g. , depression) were surveyed at the same time. Genotyping and genome-wide imputation of HPFS/NHS participants had previously been performed. *APOE*genotype was inferred from 2 SNPs; weighted AD-SNP risk scores were computed from 19 other AD-associated SNPs (SNPscore). We performed linear regression of SMCscore on*APOE*genotype and SNPscore, adjusted for age and ancestry. In secondary analyses, we adjusted for non-genetic AD risk factors; evaluated effect modification by*APOE*genotype (SNPscore only), age, and gender; and excluded participants recruited for genotyping to studies of cardiovascular disease. **Results**: Among 5,022 HPFS participants (male; mean age = 76. 4±7. 7, standard deviation; SMCscore = 0. 84±1. 16) and 7,347 female NHS participants (female; age = 78. 7±6. 5; SMCscore = 0. 81±1. 23), 53. 3% reported no SMCs, and <5% reported ≥4 SMCs. *APOE*genotype was significantly associated with SMCscore, but the effect was limited to the*e3e4*(b = 0. 11; 95% confidence interval = 0. 06, 0. 15) and*e4e4*genotypes (b = 0. 38; 0. 21, 0. 56). There was no significant association between SNPscore and SMCscore, with or without adjustment for*APOE*genotype and non-genetic AD risk factors (b = 0. 05; -0. 02, 0. 12). The effect of*APOE*genotype was non-significant among women <75 years (b = 0. 06; 95% CI = -0. 06, 0. 12) but significant among women ≥75 years (b = 0. 18; 95% CI = 0. 10, 0. 25). Excluding cardiovascular disease study participants, the effect of SNPscore on SMCscore was similar to that of*APOE*genotype in our primary analysis (b = 0. 11; 0. 06, 0. 18). **Conclusion**:*APOE*status is associated with SMCs in older adults. The association between AD-SNPs and SMCs among non-cardiovascular disease study participants warrants further investigation given the complex relationship between AD, vascular dementia, metabolic disease, and*APOE*status.

1037W

Aging-related miRNAs are associated with schizophrenia and verbal memory impairment. *E. Quillen[1], M. A. Carless[1], J. Neary[1], R. C. Gur[2], M. F. Pogue-Geile[3], J. Blangero[4], H. H. H. Göring[4], V. L. Nimgaonkar[5], R. E. Gur[2], L. Almasy[4].* 1) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 2) Departments of Psychiatry, Neurology and Radiology, University of Pennsylvania Medical Center, Philadelphia, PA, USA; 3) Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA; 4) South Texas Diabetes and Obesity Institute, University of Texas Health Science Center at San Antonio, Regional Academic Health Center, Harlingen, TX, USA; 5) Department of Psychiatry and Human Genetics, University of Pittsburgh Medical Center, Pittsburgh, PA, USA.

Mounting evidence supports a role for aging-related microRNAs (miRNAs) in a wide variety of disorders including heart disease, the metabolic syndrome, muscle disorders, and neurodegenerative diseases. Altered expression may be, in effect, prematurely aging the tissue and contributing to the onset of symptoms. To investigate a role for aging-related miRNAs in schizophrenia (SCZ), fourteen miRNAs previously shown to alter activity of canonical aging pathways were selected for analysis in 419 individuals from the Multiplex-Multigenerational Genetic Investigation of Schizophrenia (MGI). MGI is a family-based study of SCZ and related endophenotypes including measures of disordered cognition. SCZ has a heritability of 0. 92 in these families. Reduced verbal memory (VMEM) is an important endophenotype for SCZ from a clinical perspective because it is a byproduct of SCZ that is resistant to current treatments and a useful endophenotype due to its relatively large heritability (h2 =0. 52) and negative genetic correlation with the disease, suggesting overlapping genetic influences. miRNAs were extracted from lymphoblastoid cell lines and sequenced on the Illumina GA*IIx* utilizing TruSeq small RNA technology. All observed mature human miRNAs were normalized to reads per million (RPM) and aligned with 100% identity and a seed of 12 bases to mirBase version 20 using NextGENe software. Association analysis was performed in SOLAR to assess the age- and sex-adjusted relationship between miRNA expression and VMEM and SCZ while controlling for pedigree structure. Of the fourteen miRNAs considered, seven were nominally associated ($p < 0.05$) with SCZ and/or VMEM demonstrating that aging-related miRNAs are associated with SCZ more often than expected by chance. After Bonferroni correction, miR-206 is significantly associated with VMEM ($p = 0.0008$) and let-7 is significantly associated with SCZ ($p = 0.0015$). miR-206 targets the insulin-like growth factor 1 pathway which has been shown to play a major role in aging in multiple model organisms. miR-206 expression has also been shown to regulate the brain-derived neurotrophic factor in Alzheimer's disease, but associations with SCZ have been mixed. Similarly, let-7, which is known to target DAF-12 signaling has been associated with both decreased lifespan and Alzheimer's in model organisms. Taken together, these findings support a role for early-life changes in miRNA normally associated with aging in schizophrenia and verbal memory.

1038T

Neurodevelopmental Copy Number Variants and Clinical Risk: a Pediatric Medical Record Population Study. *K. Ahn[1], F. Mentch[2], C. Kao[2], H. Hakonarson[2], J. Rapoport[1].* 1) Child Psychiatry Branch, National Inst Mental Health, NIH, Bethesda, MD; 2) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA.

Background: Chromosomal copy number variants (CNVs) associated with schizophrenia have been shown to have a broad etiological risk for neurodevelopmental disorders such as autism or intellectual deficiency and some involve other categories of disease. Their overall penetrance more broadly defined remains unknown within pediatric populations. **Method:** Selected CNVs were screened utilizing Taqman within a population sample of 60,000 pediatric patents. CNV carriers were examined with respect to pre-scored pediatric records and compared to disorders in 5:1 matched non carrier group for each individual CNV. Prevalence for 14 disorder categories were estimated. **Results:** Five of these CNVs (MYT1L (dup), 22q11 dup, 16p11. 2 (deletions and duplications), and 15q11. 3 (deletions and duplications) were detected in the pediatric population at expected rates derived from large control population studies. MYT1L dup did not contribute to increased risk of any disorder category. Having any of the other CNVs showed an increase for four of the 14 disease categoriesafter Bonferroni correction: congenital deficit, surgery, digestive system disorder or mental disorder (preimarily developmental delay). When individual CNVs were examined in relation to their matched non carriers, significant associations were found between 22q11duplication, and gastro intestinal reflux disorder which was more likely to be present in the presence of developmental delay; 16p11. 2del was associated with both mental and nervous system disorders. 15q13. 3 trend for association with mental disorders. Notably, pediatric records did not indicate increased risk for subjects with MYT1L duplication. . **Conclusion:** A broader concept of overall clinical penetrance is of importance for both genetic counseling and understanding the pathophysiology of these disorders. Psychiatric follow-up is needed to complete this study and is ongoing. . This is also the first report of a CNV association (22q11 duplication) for pediatric GERD, and may represent a delay in vagal nerve maturation.

1039F

Using a tiered cross-disorder approach to identify and confirm candidate genes for brain disorders. *A. Gonzalez-Mantilla[1], A. Moreno-De-Luca[1,2,3], D. H. Ledbetter[1,2], C. L. Martin[1,2].* 1) Autism & Developmental Medicine Institute, Geisinger Health System, Danville , PA; 2) Genomic Medicine Institute; 3) Department of Radiology, Geisinger Health System.

Developmental brain disorders (DBD) are a group of clinically and genetically heterogeneous disorders characterized by high heritability. We used a tiered, cross-disorder approach to identify novel candidate genes and provide additional evidence for genes previously associated with DBD. We based our analyses on studies showing that the identification of two independent *de novo* loss-of-function variants (LoF) in the same gene among unrelated cases with a DBD is a powerful method to reliably identify disease causative genes. Initially, we conducted a phenotype-based review of studies that used whole genome or exome sequencing, chromosomal microarrays, and/or targeted sequencing to evaluate large cohorts of cases with intellectual disability, autism, epilepsy, schizophrenia, attention deficit hyperactivity disorder, or bipolar disorder and annotated all LoF variants and associated phenotypes. We selected genes with two or more LoF variants (which also included intragenic/single gene deletions) and conducted a genotype-based literature review focused on each of the candidate genes to identify additional LoF variants and phenotypes from smaller cohort studies and case reports to compile a comprehensive cross-disorder, genotype-phenotype knowledge base. We ranked the candidate genes into four tiers based on the level of evidence: tier one includes genes with three or more *de novo* LoF variants; tier two, genes with two *de novo* LoF variants; tier three, genes with at least one *de novo* LoF variant; and tier four, genes with only inherited LoF variants. The phenotype-based literature review identified 1,807 cases with LoF variants involving 973 genes from 108 studies. Filtering our results to genes with two or more LoF variants identified in at least two unrelated cases, resulted in 232 genes. The genotype-based review identified 358 additional cases with LoF variants in 68 candidate genes from 147 studies, for a total of 232 genes where LoF variants were found twice or more in a sample of 1,417 cases. Forty-nine genes were ranked in tier one, 47 in tier two, 67 in tier three, and 69 in tier four. We identified 6 novel candidate genes and provided additional evidence for 5 novel putative candidate genes. By integrating genomic data from structural and sequence LoF variants and phenotype data from six apparently distinct DBD, the yield of gene discovery increased over what would be obtained if each disorder and type of genomic variant were analyzed independently.

1040W

Association analysis of polymorphisms in the dopaminergic pathway and cognitive measures in the Diabetes Heart Study. *S. E. Martelle[1, 2], L. M. Raffield[1], A. J. Cox[3], B. I. Freedman[4], C. E. Hugenschmidt[5], J. D. Williamson[5], D. W. Bowden[2].* 1) Integrative Physiology and Pharmacology, Wake Forest School of Medicine, Winston Salem, NC; 2) Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston Salem, NC; 3) Molecular Basis of Disease, Griffith University, Southport, Brisbane, QLD, Australia; 4) Department of Internal Medicine, Nephrology, Wake Forest School of Medicine, Winston Salem, NC; 5) Department of Internal Medicine, Gerontology and Geriatric Medicine, Wake Forest School of Medicine, Winston Salem, NC.

A wealth of epidemiological and imaging studies have established an association between type 2 diabetes (T2D) and accelerated cognitive decline. It is less clear, however, the exact mechanism linking T2D and reduced cognitive function. The monoamine system can be altered by T2D and is extensively involved in cognition. The present study tested candidate polymorphisms in genes of the DA pathway (e. g. *DDC*, *TH*, *DRD2*, *DRD3*, and *DAT1*) for association with variation in cognitive performance in a large and well-described familial cohort of T2D patients and related controls of European ancestry (n = 893). Previous analyses in the cohort, known as Diabetes Heart Study (DHS) – Mind, showed that cognitive measures are significantly heritable. Thus, we hypothesized that sequence variants in genes linked to DA metabolism are associated with cognitive function as assessed by (1) the Digit Symbol Substitution Task; (2) the Modified Mini-Mental State Examination; (3) the Stroop Task; (4) the Rey Auditory-Verbal Learning Task (RAVLT); and (5) the Controlled Oral Word Association Task for Phonemic and Semantic Fluency in this cohort. From a total of 13 SNPs (7 genes/regions) selected to form an a priori set of candidate variants, one coding variant from the dopamine decarboxylase gene (rs10499695) was associated with poorer performance on a subset of RAVLT measuring retroactive interference (p = 0. 001, b = -0. 45). Furthermore, the well-documented DAT1 3'-UTR VNTR was genotyped in the cohort and showed no evidence of association with any cognitive task. Additional analyses of loci of interest were performed using data from an Affymetrix® Genome-Wide Human SNP Array 5. 0 and 1000 Genomes imputation from this array in a subset of the cohort (n = 572). 499 SNPs were extracted from the *DAT1*, *DDC*, *DRD2*, *DRD3*, and *TH* genes. Nominal associations (p < 0. 005) in *DAT1*, *DDC* and *TH* with Stroop and RAVLT performance were revealed, but rs10499695 failed to reach significance (p=0. 039) in this subset of individuals. In conclusion, these data suggest a possible role of dopaminergic genes in cognitive performance in T2D.

1041T

Epigenetics of serotonin 1B receptor: analysis of suicidal behavior in mood disorder and schizophrenia. *A. Bani-Fatemi[1,2], A. Howe[1], N. Hettige[1,2], C. Zai[1], J. Kennedy[1,2,3], V. De Luca[1,2,3].* 1) Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Institute of Medical Science, University of Toronto, Ontario, Canada; 3) Department of Psychiatry, University of Toronto, Ontario, Canada.

Background: Several studies have suggested that suicidal behavior is partially determined by genetic factors supporting a search for candidate genes involved in the neurobiology of serotonin system. HTR1B has been highlighted in the literature as being involved in suicidal behavior. We analyzed parent-of-origin effect (POE) in suicide attempters and differential expression of HTR1B rs6296 alleles in suicide victims. **Methods:** We compared the C861G (rs6296) allele-specific mRNA levels in the frontal cortex of suicide (n=13) and non-suicide victims (n=13) from the Stanley Medical Research post-mortem brain collection. We also performed a family-based association study and QTDT analyses of C861G rs6296 polymorphism in 162 nuclear families with at least one subject affected by bipolar disorder with suicidal behavior and then we tested the rare and common variants in the HTR1B promoter for linkage disequilibrium with the C861G in an unrelated sample of suicide attempters and non-attempters. **Results:** We observed no alterations in the C/G expression ratio in suicide victims compared to controls (p=0. 370); however, individuals with comorbid alcohol abuse (p=0. 009) and substance abuse (p=0. 03) were found to have a greater C/G ratio. There was no preferential transmission of C861G variant and suicidal behavior severity scores in both maternal and paternal meiosis (p>0. 05). When comparing the linkage disequilibrium between the C861G and the cis-acting SNPs, we did not find any difference in suicide attempters and non-attempters. **Conclusions:** The novelty of the present study was to combine two different molecular strategies to understand the effect of the promoter variants in differentiating the expression between the maternal and paternal chromosomes. We found that the promoter variants are likely to influence the different expression of the rs6296 SNP. No evidence of allelic imbalance or POE of rs6296 in suicidal behavior was found. Further research is required to assess these main effects in larger cohorts.

1042F

Exploration of a multiplex family with synesthesia and autism using whole genome sequencing. *A. Mathieu[1,2,3], A. Maruani[1,2,3,4], S. Calderari[1,2,3,5], F. Amsellem[1,2,3,4,5], L. Bouvet[6], J. Buratti[1,2,3,5], T. Bourgeron[1,2,3,5], L. Mottron[7], R. Delorme[1,2,3,4].* 1) Human Genetics and Cognitive Functions Unit, Institut Pasteur, Paris, France; 2) CNRS UMR 3571 : Genes, Synapses and Cognition, Institut Pasteur, Paris, France; 3) Université Paris Diderot, Sorbonne Paris Cité, Human Genetics and Cognitive Functions, Paris, France; 4) Assistance Publique-Hôpitaux de Paris, Robert Debré Hospital, Department of Child and Adolescent Psychiatry, Paris, France; 5) FondaMental Foundation, Créteil, France; 6) Unité de recherche interdisciplinaire OCTOGONE, Université Toulouse Jean Jaurès, Toulouse, France; 7) Centre d'excellence en trouble envahissants du développement de l'Université de Montréal (CETEDUM), Hôpital Rivière-des-Prairies et Département de Psychiatrie, Montréal, Canada.

Autism occurs in 1% of the population and is characterized by impaired social communication as well as by restricted interests, and repetitive behaviors. Synesthesia affects approximately 1-4% of the non-autistic population and occurs when perceiving a class of elements within one sensory modality triggers the perception of another class of elements in another modality. Synesthesia is reported more frequently in individuals with autism compared with general population. However, the genetic factors increasing the risk of synesthesia remain unknown. Here, we analyzed a three-generation multiplex family including members with autism, synesthesia or both using whole genome sequencing. The family is composed of 46 individuals (26 males and 20 females), among which two are autistic individuals without synesthesia, 3 are autistic and synaesthetes autism (including one with savant syndrome, PMID 24600416), and 10 non autistic synaesthetes only. We first performed a SNP genotyping using 1 Omni and 2. 5 Omni Illumina BeadChip arrays for 29 individuals. Then, whole genome sequencing was performed on 23 patients. For linkage analysis, we used Alohomora software to combine the pedigree information, allele frequencies and map position of the SNPs. We used Merlin to perform the non parametric (NPL) and parametric linkage analysis. The haplotype segregating with synesthesia was ascertained by Merlin and visualized by HaploPainter. For ASD or for the combination of ASD and synesthesia, linkage analysis revealed no significant LOD score. Concerning synesthesia only, a single region showed a LOD score of 2. 5 and NLP score of 1. 6 on chromosome 10. The minimal interval of 1. 89 Mb (chr10 135656 to 2027751) includes 17 genes. We could not detect copy number variant (CNV) associated with the trait. Among the variants identified using whole genome sequencing, a stopgain mutation of the isopentenyl diphosphate:dimethylallyl diphosphate isomerase type 2 *(IDI-2)* was segregating with synesthesia. In summary, this pilot linkage and sequencing study revealed a new candidate genomic region for synesthesia. Further analyses are in progress to replicate this finding and to validate mutations of IDI-2 as risk factors for this complex trait.

1043W

Targeted neurogenesis pathway-based gene analysis identifies *ADORA2A* associated with hippocampal volume in mild cognitive impairment and Alzheimer's disease. *E. Horgusluoglu[1], K. Nho[2], S. Risacher[2], AJ. Saykin[1,2,3,4].* 1) Medical and Molecular Genetics, Indiana University, School of Medicine, Indianapolis, IN; 2) Center for Neuroimaging, Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN; 3) Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN; 4) Indiana Alzheimer's Disease Center, Indiana University School of Medicine, Indianapolis, IN.

Background: New neurons are generated throughout adulthood in the olfactory bulb and dentate gyrus of the hippocampus, and are incorporated into hippocampal networks during construction and maintenance of neural circuits and in turn contribute to learning and memory. Numerous intrinsic and extrinsic factors such as growth factors, transcription factors, and cell cycle regulators control neural stem cells proliferation, differentiation, and maintenance into mature neurons. However, the genetic mechanisms controlling adult hippocampal neurogenesis remain unclear. To address this issue we performed a gene-based association analysis of neurogenesis pathway-related candidate genes using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI). Methods**:** Neurogenesis-related genes were curated from existing databases (Qiagen RT2 Profiler PCR Arrays and Mammalian Adult Neurogenesis Gene Ontology (MANGO)). The gene list was filtered by AD susceptibility genes from the Alzgene database (http://www. alzgene. org/) and large-scale GWAS (Lambert,et al. 2013, *Nature*). Caucasian non-Hispanic individuals (N=1,525) with AD or mild cognitive impairment (MCI) and cognitively normal older adults from the ADNI cohort with MRI and genotyping data were included. Gene-based association analysis of neurogenesis pathway-related candidate genes was performed. Baseline bilateral hippocampus and hippocampal subfield (CA4 and dentate gyrus (CA4DG)) volumes were extracted from MRI and served as phenotypes. Gender, age, intracranial volume, MRI field strength, and diagnosis at scanning were entered as covariates. The empirical *p* value from permutation testing for each gene wascalculated to adjustfor the number of significant SNPs in each gene. Results: *ADORA2A* was significantly associated with total hippocampal volume and CA4DG ($p<0.001$). For the most significant SNP (rs9608282) in *ADORA2A,* dosage of the minor allele (T) increased hippocampal volume. rs9608282 was also associated with composite memory score ($p= 0.0076$). We replicated the association between rs9608282 (*ADORA2A*) and total hippocampal volume in the independent AddNeuroMed cohort. Conclusion: *ADORA2A*-mediated control of neuroinflammation modulates adult neurogenesis and the inhibition of *ADORA2A* prevents Aβ-induced neurotoxicity. Targeted pathway-based genetic analysis combined with brain imaging endophenotypes appears promising to help elucidate disease pathophysiology and identify potential therapeutic targets.

1044T

Don't lose the forest for the trees- new insights into Juvenile Psychosis from chromosomal microarrays. *C. A. Brownstein[1], P. Agrawal[2], M. C. Towne[2], V. Lip[4], A. H. Beggs[1], J. Gonzalez-Heydrich[3].* 1) Division of Genetics and Genomics, The Research Connection, The Manton Center for Orphan Disease Research, and the Developmental Neuropsychiatry Program, Boston Children's Hospital, Boston, MA; 2) Division of Genetics and Genomics, The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA; 3) The Developmental Neuropsychiatry Program, Boston Children's Hospital, Boston, MA; 4) Claritas Genomics, Cambridge, MA.

New candidate genes and regions for schizoaffective disorders including schizophrenia are being discovered at a rapid rate. However, this family of illnesses are still considered to be highly heritable neuropsychiatric disorders of complex genetic etiology. Copy number variation (CNV) is a highly topical area of research in schizophrenia, but the clinical relevance is uncertain and the translation to clinical practice is under-studied. However, genome overlap between psychiatric disorders means that when a CNV is implicated in a related mental health disorder, the evidence becomes more convincing. Regions that have been implicated in autism have been also implicated in schizophrenia. Here, we describe a cohort of juvenile patients with auditory and visual hallucinations at a tertiary children's hospital. Forty-one families have been identified with having a child showing signs of psychosis before 11 years of age. 28 families have been consented thus far. 13/21 patients have had abnormal chromosomal microarray results (8 deletions, 4 duplications, 1 LOH). Genes affected by the copy number variations include ABCC6, NOMO1, NPIP, NTAN1, RRN3, PDXDC1, FLJ00285, FLJ00285, LOC100288332, DQ599532, MPV17L, C16orf45, KIAA0430, NDE1, MYH11, FOPNL, ABCC1, DDHD2, CPPED1, SMARCA4, CTSB, GATA4, NEIL2, BLK, TDH, MTMR9, XKR6, PINX1, RP1L1, UNQ9391, MSRA, TNKS, PPP1R3B, ERI1, CLDN23, PRAGMIN, DEFB109, HE2, SPAG11A, SPAG11B, DEFB107B, LOZ93432, MOXD2P, TRYX3, PRSS58, LOC730441, TRY2P, and TCRBV,Genes that have been identified as candidates in other mental health disorders include NTAN1, NDE1, SMARCA4, and ABCC1. Boston Children's Hospital is developing the infrastructure needed for large-scale psychiatric research and treatment discovery with the creation of the Developmental Neuropsychiatry Program (DNP). The DNP is investigating therapeutics to prevent the development of schizophrenia in at-risk children by working with the youngest patients (some of which are described here), creating neuronal cell cultures to identify cell autonomous effects, and constructing models of neural networks. The study of rare Mendelian forms of Juvenile Psychosis are an effective way to discover new candidate genes for the condition. Syndromic patients with very early onset psychosis likely increases the chances of finding a CNV with CMA. Don't lose the forest for the trees-CMAs are less expensive than exome and still yield valuable insights.

1045F

Genetic Association Study Suggests FOXA1 Could Be a Common Regulator of Epilepsy Electroencephalogram Endophenotype and Reading Disability Comorbidity. *N. Panjwani[1], J. Crosbie[3,4], P. Arnold[1,3,4], R. Schachar[1,3,4], D. K. Pal[5], L. J. Strug[1,2].* 1) Genetics and Genome Biology, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 2) Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 3) Neurosciences and Mental Health Program, Research Institute, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 4) Department of Psychiatry, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 5) Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK.

IntroductionRolandic epilepsy (RE) is a neurodevelopmental disorder characterized by seizures and centrotemporal sharp waves (CTS) on electroencephalogram. Patients and relatives have high risk of literacy impairments (RDG). We have identified linkage of CTS in RE to 11p13 and association with markers in the *ELP4-PAX6* locus (Strug et al. 2009), although this association was not replicated (Reinthaler et al. 2014). We have also found linkage of RDG in RE to the 1q42 locus (Strug et al. 2012), but there have been no replication studies. We therefore aimed to clarify the evidence for the two loci in RE-associated CTS and RDG in an independent collection of RE families and population-based controls. **Methods**RE families from North America and Europe (214 individuals with CTS, 140 with RDG) and 749 sex and ethnicity matched controls from the Ontario Science Centre (N=749) were genotyped on the HumanCoreExome and HumanOmniExpress BeadChips. We used generalized estimating equations with a logit link and independence correlation structure for association analysis, corrected for sex and population structure. We fine-mapped the loci using imputation for significant regions, defined as –log10P=2. 514 for the *ELP4-PAX6* and –log10P=4. 445 for 1q42. **Results**One genotyped SNP (rs662702) reached regional significance at the *ELP4-PAX6* locus (OR=2. 07; –log10P = 2. 79). Results remained consistent when patients were geographically grouped: North Americans OR=2. 11, –log10P=2. 12; and Europeans OR=1. 92, –log10P=1. 68. Similarly, one genotyped SNP reached regional significance for RDG, rs1495855, –log10P=5. 08, OR=2. 28, along with 10 nearby imputed SNPs in linkage disequilibrium (LD). The CTS associated rs662702 is in the 3'-UTR of *PAX6* while rs1495855 for the RDG is intergenic. Interestingly, both SNPs are located at or in complete LD with reported FOXA1 transcription factor binding regions while histone marks suggest enhancer activity in neurons (Ward and Kellis, 2011; Boyle et al 2012). **Discussion**We report replication of both the 11p13 CTS locus and 1q42 RDG locus. Changes in the 3'UTR of *PAX6,* an important mediator of forebrain development (Georgala et al 2011), may influence PAX6 levels. FoxA factors regulate regional brain and neuronal specification and differentiation (Lin et al. 2009). Furthermore, histone modifications suggest enhancer activity in early neurodevelopment. These results stimulate new mechanistic hypotheses in common neurodevelopmental disorders.

1046W

Exome sequencing in a cohort of patients with microcephaly and related conditions – identification of known and novel disease genes and the lessons learned. *A. T. Pagnamenta[1], M. F. Howard[1], N. Popitsch[1], S. J. L. Knight[1], N. Galjart[2], A. Goriely[3], D. A. Keays[4], J. C. Taylor[1], U. Kini[5].* 1) National Institute for Health Research Biomedical Research Centre, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK; 2) Department of Cell Biology, Erasmus Medical Center, 3015 GE Rotterdam, Netherlands; 3) Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, United Kingdom; 4) Institute of Molecular Pathology, Vienna 1030, Austria; 5) Department of Clinical Genetics, Oxford University Hospitals NHS Trust, Oxford OX3 9DU, UK.

Background: Development of the brain requires a delicate balance between cellular proliferation, differentiation, organisation, migration and apoptosis. When this process is disrupted, structural abnormalities may result. Due to significant levels of genetic heterogeneity and the complexity of the processes involved, we assessed the diagnostic value of whole-exome sequencing (WES) for patients with microcephaly and related conditions. **Methods:** Over 200 subjects with brain malformations were recruited. After excluding likely causal CNVs, 19 families with microcephaly and related conditions were selected. 7 of these families were reported to be consanguineous and for these, SNP arrays were used to identify regions of autozygosity. Exomes were captured from patient DNA samples using commercially available kits and sequenced on the Illumina HiSeq platform. **Results:** Likely causative variants were identified in 6 genes which were novel at the time of discovery. ENCODE data helped us identify a novel variant located 4bp upstream of the *PIGY* transcription start site which cosegregated with microcephaly and disrupts a consensus SP1 binding site. Using qPCR and deep Ion PGM sequencing, we showed that this variant leads to reduced *PIGY* expression. In another affected sib pair, we identified a *CLASP1* p. G1220S mutation located within a 26Mb region of shared autozygosity for which protein interaction studies are currently underway. A boy with microcephaly and cerebellar vermis atrophy harbored a *de novo* p. C29* variant in *SOX11,* a gene recently associated with Coffin-Siris syndrome*.* A girl with progressive microcephaly was found to harbor a 7bp deletion in *GPT2* within a 34Mb region of autozygosity. This variant was overlooked initially as it was in 73% of reads and erroneously called heterozygous, emphasizing the benefits of having a flexible filtering strategy. Mutations in *GPT2* have recently been reported in developmental encephalopathy. In 2 other families we identified *PGAP3* and *PI4KA* as novel disease genes, which we have described elsewhere. Presumed causative variants were also detected within well-established disease genes including *de novo* mutations in *CASK*/*PCDH19*, compound heterozygous variants in *CDK5RAP2*/*ASPM* and a homozygous *TBC1D24* substitution. **Conclusions:** Likely pathogenic variants were identified in 63% of cases, in both known and novel disease genes, highlighting the benefits of WES over gene panel approaches in this heterogeneous group of patients.

1047T

Gene-gene interaction analysis of mitochondria-related genes in the risk of Parkinson's disease and Alzheimer's disease. *S. J. Chung[1], J. Kim[1], Y. J. Kim[1], S. You[2], M. Kim[3], S. Y. Kim[4], J. Lee[1].* 1) Neurology, Asan Medical Center, Seoul, Seoul, South Korea; 2) Department of Neurology, Dongsan Medical Center, Keimyung University, Daegu, Korea; 3) Department of Neurology, Bobath Memorial Hospital, Seongnam, Republic of Korea; 4) Department of Psychiatry, Asan Medical Center, University of Ulsan College of Medicine, Seoul, South Korea.

Objectives: To investigate the role of genetic interaction among mitochondria-related genes in the development of Parkinson's disease (PD) and Alzheimer's disease (AD). **Background:** Mitochondrial dysfunction in the substantia nigra has been documented in patients with PD. It is also recognized as major contributor in Aß-mediated neurotoxicity in AD. The importance of mitochondria to the cause and pathogenesis of PD and AD has been reinforced by the identification of specific mutations in the mitochondria-linked genes. **Methods:** The mortalin (*HSPA9*), *parkin*, *PINK1*, *DJ1*, and *COQ2* genes were selected. For *HSPA9*, direct sequencing using an automated DNA sequencer (ABI 3730, Applied Biosystems, CA, USA) for the entire coding region (17 exons) and exon-intron boundary of 50 bp was performed in 24 AD patients, 24 PD patients, and 24 controls. The genetic variants with minor allele frequency (MAF) > 1% from sequencing data were selected for the second-stage genotyping. Common (MAF > 5%) and low-frequency (MAF = 1 – 5%) genetic variants were also selected using the HapMap and 1000 Genomes Project JPT and CHB samples to select variants in the *HSPA9*, *parkin*, *PINK1*, *DJ1*, and *COQ2* genes. We genotyped genetic variants (8 in *HSPA9*, 10 in *parkin*, 4 in *PINK1*, 6 in *DJ1*, and 2 in *COQ2*) in 400 AD cases, 500 PD cases, and 500 controls, using the Fluidigm high-throughput platform. Logistic regression analysis with additive coding schemes as a primary analysis was performed. **Results:** The high-risk group carrying more than 4 risk alleles in the *HSPA9* SNP rs41295739, *parkin* SNP rs9356040, *PINK1* SNP rs76795760, *DJ1* SNP rs2235733, and *COQ2* variant M128V had significantly increased risk of PD compared with the low-risk group carrying 0-2 risk alleles (= 0. 01 for 4 vs 0-2 risk alleles). The high-risk group carrying more than 3 risk alleles in the *HSPA9* SNP rs41295739, *parkin* SNP rs1784590, *PINK1* SNP rs1043424, *DJ1* SNP rs34124834, and *COQ2* SNP rs148156462 had significantly increased risk of AD compared with the low-risk group carrying 0-2 risk alleles (= 0. 039 for 4 vs 0-2 risk alleles). **Conclusions:** Our results suggest that genetic interaction among mitochondria-linked genes may play a role in the development of PD and AD.

1048F

Deletion in *ABCA7* associated with Alzheimer's disease in African Americans. *H. N. Cukier[1], B. W. Kunkle[1], S. Rolati[1], K. L. Hamilton-Nelson[1], M. A. Kohli[1], B. A. Dombroski[2], B. N. Vardarajan[3], P. L. Whitehead[1], D. Van Booven[1], R. Lang[4], E. R. Martin[1,5], G. W. Beecham[1,5], L. A. Farrer[6], M. L. Cuccaro[1,5], J. M. Vance[1,5], R. P. Mayeux[3], J. R. Gilbert[1,5], R. M. Carney[1,5], G. S. Byrd[4], J. L. Haines[7], J. D. Schellenberg[2], M. A. Pericak-Vance[1,5], Alzheimer Disease Genetics Consortium (ADGC).* 1) John P. Hussman Institute for Human Genomics (HIHG), University of Miami, Miller School of Medicine, Miami, FL; 2) Department of Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 3) The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Gertrude H. Sergievsky Center, Departments of Neurology, Psychiatry, and Epidemiology, College of Physicians and Surgeons, Columbia University, New York, NY; 4) Department of Biology, North Carolina A&T State University, Greensboro, NC; 5) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miller School of Medicine, Miami, FL; 6) Departments of Medicine, Neurology, Ophthalmology, Genetics & Genomics, Epidemiology, and Biostatistics, Boston University, Boston, MA; 7) Department of Epidemiology and Biostatistics, Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, OH.

The *ATP-binding cassette, sub-family A (ABC1), member 7* (*ABCA7*) gene has been implicated as a risk factor in Alzheimer's disease (AD) in both African American (AA) and Caucasian (CA) populations, each with distinct polymorphisms of differing effect sizes. The effect in AA is significantly higher, comparable to that found in CA for *APOE ε4*. The AA risk allele, rs115550680, is monomorphic in CA (1,000 Genomes) and located 13. 2 kb upstream of the CA risk allele, rs4147929. Despite numerous studies implicating *ABCA7*, the underlying damaging alleles conveying the genome-wide signals have yet to be revealed. We performed custom capture sequencing of a 150 kb region encompassing the *ABCA7* region on 40 AA individuals with AD and 37 control AA carrying the AA risk allele. Data were processed via GATK and BWA, and deletions identified with Pindel. 1,120 SNVs were detected by sequencing. 11 variants had different frequencies in cases and controls (p<0. 1). In addition, a 44 base pair deletion (rs142076058) was identified in all 77 risk genotype carriers, signifying that it could be in high linkage disequilibrium with the AA risk allele. This deletion was further assessed in two AA cohorts – HIHG (cases:494, controls >65 yrs:494) and ADGC (cases:521, controls >65 yrs:890). Analyses for each cohort were adjusted for gender, age, and *APOE* status. In the HIHG cohort, the deletion was present in more cases (14%) compared to controls (10%) and correlated significantly with disease (p=0. 0002, OR=2. 1 [CI:1. 4-3. 1]). The AA risk allele is similarly significant (p=0. 0008, OR=2. 0 [CI:1. 3-3. 0]). There were 3 discordant individuals with the deletion alone; all 3 presented with AD, demonstrating that the deletion could be the underlying, causative variant. The association was then replicated in the ADGC cohort (p=0. 0097, OR=1. 6 [CI:1. 1-2. 3]). Joint analysis of the AA cohorts resulted in an increased effect size (p=0. 0003, OR=1. 6 [CI:1. 3-2. 1]). The deletion was found in only 0. 09% of our CA cohort. Reverse transcription-PCR from RNA extracted from the blood of AA individuals with and without the deletion demonstrated that the deleted allele produces a stable, detectable RNA strand. The deletion falls in the 14th exon of *ABCA7* and is predicted to result in a frameshift and truncating mutation (p. Arg578Alafs) that could interfere with protein function. This *ABCA7* deletion could represent an ethnic-specific, pathogenic alteration in Alzheimer's disease.

1049W

Towards the identification of new genes in recessive Parkinson's disease. *V. Drouet[1], A. Nicolas[1], M. Jacoupy[1], F. Cormier-Dequaire[1,2], M. Sabahtou[3], S. Liebau[4], E. Lohmann[5,6], M. Tazir[7], A-L. Leutenegger[8], A. Singleton[9], O. Corti[1], S. Lesage[1], A. Brice[1,10] for the French PDG group and the IPDGC consortium.* 1) Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Inserm U 1127, CNRS UMR 7225/NEB, Institut du Cerveau et de la Moelle épinière, ICM, Paris, France; 2) Centre d'Investigation Clinique Pitié Neurosciences CIC-1422, Paris, France, Paris, France; 3) Fondation Jean Dausset-CEPH, Paris, France; 4) Institute of Neuroanatomy, Eberhard Karls University Tübingen, Tübingen, Germany; 5) Movement Disorders Unit, Department of Neurology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey; 6) Department of Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany; 7) Service de neurologie CHU Mustapha, Alger, Algérie; 8) Inserm U946, Université Paris Diderot, Institut Universitaire d'Hématologie, UMR946, Paris, France; 9) Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD 20892, USA; 10) AP-HP, Hôpital de la Salpêtrière, Department of Genetics and Cytogenetics, Paris, France.

Autosomal recessive early-onset Parkinson's disease (PD) is clinically and genetically heterogeneous: mutations in three genes, *PARK2*, *PINK1* (PARK6) and *DJ-1* (PARK7), cause phenotypes similar to idiopathic PD with a good and prolonged response to dopaminergic therapy. Mutations in other genes, *ATP13A2* (PARK9), *PLA2G6* (PARK14), *FBXO7* (PARK15), *DNAJC6* (PARK19), and *SYNJ1* (PARK20), cause more severe disease, a poor response to levodopa and additional clinical signs: dystonia, cognitive impairment, neuro-behavioural abnormalities, pyramidal signs, ophthalmoparesis, and autonomic dysfunction. In approximately 50% of the familial cases with early-onset PD, mutations in these genes are not detected, suggesting further genetic heterogeneity with the involvement of one or, more probably, several other genes. To identify new genes implicated in this disorder, we selected 140 early-onset (<55 years) PD families or isolated cases with known or suspected consanguinity, excluded for *PARK2*, *PINK1*, *PARK7*, and *LRRK2* G2019S mutation. We performed genome-wide SNP genotyping to detect genomic rearrangement and regions of homozygosity. Then, exome sequencing in 62 cases with confirmed consanguinity (inbreeding coefficient F≠0) were carried out. We prioritized homozygous variants predicted to produce frameshifts, stop codons or splicing defects, which: 1) would disrupt protein function; 2) were rare (allele frequency <1%) or novel but absent in the homozygous state in 530 Caucasian controls and 6503 exomes in the EVS database; 3) fell within homozygous intervals. Preliminary analyses of the exome data led to the identification of deleterious mutations in *ATP13A2*, *FBXO7* in 2 affected sibpairs, and a list of 44 new candidate genes. Functional validation of the best candidate genes is in progress. The identification of new genes involved in early-onset autosomal recessive PD will improve diagnosis and genetic counselling in patients and their relatives. It will also open new avenues of research on the mechanism of neurodegeneration in PD and allow the development of currently unavailable, innovative, mechanistically-based treatments to cure or slow the progression of the disease.

1050T

Molecular study of PMP22 gene duplication by MLPA and point mutations of MPZ and MFN2 genes in Charcot–Marie –Tooth patients in northwest of Iran. *P. Aob[1], A. Biglari[1], M. r. Ranjouri[1], S. Mansouri derakhshan[2], M. Shekari khaniani[2].* 1) Medical Genetic Department, Faculty of Medicine, Zanjan university of medical science, Zanjan, Iran; 2) Medical Genetic Department, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran.

Back ground: Charcot-Marie-tooth (CMT) is one of the sensory and motor neuropathies with a frequency of 1 in 2500. Clinical manifestations of CMT include slowly progressive distal weakness, sensory loss and foot deformities, which spreads proximally as the disease progresses. The syndrome represents a genetically and clinically heterogeneous group of inherited disorders classified into demyelinating form (CMT1) and axonal form (CMT2). This disorder has been related to mutations in more than 45 genes affecting glial or neuronal functions. **Objective:** Since there is no information on the frequency of CMT1A, CMT1B and CMT2A neuropathies in northwest of Iran, the aim of the present study was to estimate the frequency of genetic subtypes of CMT in this population. **Material and Methods:** From whole peripheral blood of 30 unrelated patients with a suspected diagnosis of CMT total genomic DNA was extracted by salting-out. DNA samples were analyzed by MLPA kit for detection of PMP22, COX10 and TEKT3 (located in the CMT1A/HNPP region) duplication, and KIF1B mutation. MLPA data analysis were performed using GeneMarker V2. 6. 3. Furthermore, PCR amplification and direct sequencing methods were used for the point mutations of MPZ and MFN2 genes in samples negative in MLPA analysis. **Results**: The MLPA method identified 6 patients with CMT1A/HNPP region rearrangements (including PMP22, COX10 and TEKT3 genes), 2 patients with only PMP22 duplication and 1 patient with COX10 duplication. No mutation were found in KIF1B gene. Also, the point mutation in *MPZ* and *MFN2* were found by direct sequencing in 2 and 1 patients, respectively. In total, the frequency of CMT1A, CMT1B and CMT2A were estimated 30%, 6. 6% and 3. 3%, respectively. **Discussion:** The data presented here provide an overview of frequencies of genetic subtypes of CMT patients in the Northwest of Iran for the first time. We analyzed the most frequently mutated genes associated with CMT in a selected cohort of patients and were able to achieve a molecular diagnosis in 40 % of patients (12/30). Finally, the screening of PMP22, MPZ and MFN2 mutations, which are the genes involved in 90% of CMT cases, showed that they have a low frequency in the population we studied. Therefore, it is possible that the rare mutations known to play a part in CMT are the cause of the disease in the population. Hence, whole genome sequencing is suggested to achieve an exact molecular diagnosis.

1051F

A *LRRK2* haplotype provides protection for Korean multiple system atrophy. *E. Scott[1], I. Guella[1], S. Bortnick[1], B. S. Jeon[2], M. J. Farrer[1].* 1) Djavad Mowafaghian Centre for Brain Health, Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 2) Department of Neurology, Clinical Research Institute and Movement Disorder Center, Seoul National University Hospital, Seoul, South Korea.

Multiple system atrophy (MSA) is a rare, sporadic, adult-onset neurodegenerative disorder that is similar in both clinical and pathological presentation to Parkinson's disease (PD). MSA is characterized by progressive autonomic dysfunction in addition to parkinsonian features (MSA-P subtype) or cerebellar features (MSA-C subtype). In PD, mutations in the *LRRK2* gene are recognized as one of the most common genetic causes to date. Polymorphic variants in *LRRK2* have been previously associated with PD susceptibility in both Caucasian and Asian PD patients. Recently, a protective haplotype in *LRRK2* was implicated in a Caucasian MSA cohort. Herein, we evaluate the association between common *LRRK2* exonic variants and MSA in the Korean population. 12 SNPs selected from previous *LRRK2* MSA and PD studies were genotyped by Sequenom iPLEX in 312 MSA patients and 477 controls. No significant nominal associations for the 12 SNPs were detected after correcting for multiple testing. However, non-significant trends towards protective association were observed for rs7308720, rs7133914, and rs11175964. After haplotype analysis, one large, significant protective haplotype that spans the majority of the coding region of *LRRK2* was observed (OR = 0. 65, p = 0. 027). This protective haplotype exists only when the A alleles are found at both rs7133914 and rs11175964, such that if the G alleles are present at these two SNPs, the haplotype becomes a risk haplotype. In conclusion, our study replicates, in the Korean population, the association observed between *LRRK2* and MSA. Moreover, our findings narrow down the previously-reported association to two variants (rs7133914 and rs11175964) within the ROC domain of *LRRK2*, suggesting that these variants confer the protection against MSA. The GTPase activity of LRRK2 is localized to the ROC domain, thus it is suggested that rs7133914 and/or rs11175964 alter the GTPase activity to provide protection. In addition, our findings suggest a possible association between MSA and Crohn's disease through a shared variant in *LRRK2* (rs3761863), such that individuals with MSA may be more susceptible to developing Crohn's disease. Further studies are warranted to confirm the roles of rs7133914 and rs11175964 in protection as well as the association of rs3761863 in development of Crohn's disease in MSA patients.

1052W

Parkinson's disease and genetic variability in a French Canadian isolate. *L. Sellmer[1], E. M. Bernhardi[1], I. Guella[1], M. McKenzie[1], E. Poucher[2], M. J. Farrer[1].* 1) Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; 2) Centre Medical Berger, 1000 Chemin Sainte-Foy, Quebec City, QC, Canada.

Parkinson's disease (PD) is a progressive neurodegenerative disorder with a global prevalence of about 4% by age 80. Even though most cases of PD are idiopathic, over 10% of patients report a positive family history of PD. Several genes have been implicated in autosomal-dominant PD, however, a larger portion of familial cases remain unexplained. Recently, we were able to obtain samples from 22 families of a French Canadian community; members of this community descend from *Les filles du roi.* The King's Daughters consisted of approx. 800 French women from northern and western France who immigrated to what is today called Quebec in the 17th century. The French Canadian community investigated in this study shows a seemingly very high prevalence of PD for which no genetic cause has been described. In order to investigate the underlying genetic components in this population isolate, we investigated sequence variability and exon dosage in genes known to cause PD. Subsequently, we performed Whole Exome Sequencing (WES) of 25 members of selected pedigrees. In addition, informative samples of each pedigree were genotyped on the Illumina MEGA SNP array. Discriminant Analysis of Principal Components (DAPC) performed in various Caucasian populations indicates that French Canadians are distinct in their genetic profile from the Faroese, another population isolate, Norwegians and other Caucasians. Mendelian sharing ratios above 0. 41 for unrelated individuals show that supposedly unrelated individuals from this French Canadian community are more closely related than to be expected, confirming a common founder. Large runs of homozygosity shared across multiple families were imputed from genotyping on the MEGA platform and further support relatedness and a common underlying genetic cause for disease. However, no genetic variability in known PD genes was observed in any of the investigated samples, suggesting that novel, population-specific variability could underlie disease in this population isolate. Herein, we have performed a comprehensive analysis of the genetic profile of an inbred community with a high prevalence of PD. Mutations in known genes linked to PD cannot explain this occurrence and further exome sequencing and comparative analysis within and among pedigrees with multi-incident PD are on-going.

1053T

Association between serotonin transporter VNTR and depression at 12 months post aneurysmal subarachnoid hemorrhage. *A. Stanfill, P. Sherwood, E. Crago, S. Poloyac, Y. Conley.* University of Pittsburgh, Pittsburgh, PA.

The VNTR in the second intron of the serotonin transporter gene (*SLC6A4*) has been associated with the development of post stroke depression (PSD) in ischemic stroke; however it has not been investigated with development of PSD in aneurysmal stroke. Aneurysmal subarachnoid hemorrhage (aSAH) patients are also likely to experience depression, which can affect mental acuity, lengthen recovery times, and can result in reductions in quality of life. The 9/12 and 12/12 genotypes are high risk, the 10/12 genotype is intermediate risk, and the 10/10 genotype is low risk in studies of PSD following ischemic stroke. The purpose of this study is to determine whether the association between this VNTR and PSD holds for aSAH patients. Blood samples were collected from 135 aSAH patients at time of admission. The DNA template was extracted and the samples were genotyped using PCR-RFLP. Subjects were divided into depression risk categories (high, intermediate, low) based upon VNTR genotype. Demographic information was obtained from hospitalization records, and depressive symptoms were measured at twelve months post stroke using the Beck's Depression Inventory II (BDI-II). This instrument is a 21 item self-report measure on which subjects rate depressive symptoms and attitudes on a 5-point Likert scale, and has established validity with clinically diagnosed depression. The sample was 80% female, 92% Caucasian, and 55. 08 ± 10. 31 years. Mean Fisher score at time of admission was 2. 65 ± 0. 68 and the mode was 2. Significant associations with PSD were found between all 3 genotype risk categories (ANOVA, p=0. 05), and between high risk and low risk genotypes (point biserial correlation r=0. 23, p=0. 04). For exploratory and hypothesis generating purposes, a multiple regression model with genotype, age, race, gender, and Fisher score at time of admission was significant to predict PSD (R2=0. 16, p=0. 004). These results support the association found between VNTR genotype and PSD for subarachnoid hemorrhage. This information could be used in the future to identify patients most at risk for PSD, which would then allow for targeted medical intervention prior to the development of severe symptoms.

1054F

Childhood Maltreatment Associates with Long Non-Coding RNAs Differentially Expressed in the Rostral Anterior Cingulate Cortex. *Y. Zhou, P. E Lutz, G. Turecki.* McGill Group for Suicide Studies, McGill University, Montreal, Quebec, Canada.

Background: Childhood maltreatment (CM) has been strongly associated with the dysregulation of key genes underlying various stress response systems. Changes incurred during early life may lead to lifetime negative consequences contributing to suicide etiology. Recently, long non-coding RNAs (lncRNAs) have been implicated in brain development as well as in various neurodegenerative diseases and neuropsychiatric disorders. Whether lncRNAs play a significant role in translating the effects of CM into biological programs of gene regulation remains to be investigated. Our **objectives** are (i) to investigate whether CM associates with changes in lncRNAs expression in specific brain regions, and (ii) to test whether these lncRNAs regulate the expression of other genes underlying important cellular processes. **Methods:** We performed high throughput RNA-sequencing (RNA-seq) in the rostral anterior cingulate cortex (rACC), a brain region implicated in emotional regulation. We compared 26 suicide completers with a history of CM and 24 age-matched controls free of psychiatric disorders who died by accident or naturally, and with no history of CM. Then, we used reverse transcription quantitative polymerase chain reaction (RT-qPCR) to validate the expression of one of the differentially expressed lncRNAs in the original two groups of subjects as well as in a third group of 25 suicide completers without a history of CM. **Results:** 8 differentially expressed lncRNAs were identified from the RNA-seq results. There was a significant increase in the expression of the lncRNA RP11-273G15. 2 in suicide completers with a history of CM compared to controls. Validation using RT-qPCR showed a similar pattern of expression whereby RP11-273G15. 2 expression was higher in suicides exposed to CM compared to both controls and suicides not exposed to CM. Furthermore, the expression of RP11-273G15. 2 was significantly correlated with 7 nearby genes; 4 were lymphocyte antigen 6 (LY6) genes and 2 were genes encoding zinc finger proteins. **Conclusions:** Exposure to CM may increase the expression of RP11-273G15. 2 in the rACC. Possible molecular mechanisms underlying the regulation of lncRNA expression, such as epigenetic modifications, remain to be investigated. Additionally, RP11-273G15. 2 may regulate the expression of a subset of LY6 genes and zinc finger proteins, which may underlie important cellular processes that contribute to suicide etiology.

1055W

Follow-up of the SNP rs1156026 association with schizophrenia in a Quebec sample reveals a complex interplay with DNAJC15 deletions in predicting schizophrenia and bipolar disorder. *A. Bureau[1,2], Y. C. Chagnon[2,3], J. Croteau[2], A. Fournier[2], M. -A. Roy[2,3], T. Paccalet[2], C. Mérette[2,3], M. Maziade[2,3].* 1) Médecine sociale et préventive, Univ Laval, Québec, PQ, Canada; 2) Centre de recherche du Centre intégré universitaire en santé et services sociaux de Québec, PQ, Canada; 3) Psychiatrie et neurosciences, Univ Laval, Québec, PQ, Canada.

Single nucleotide polymorphism rs1156026 at 13q13-q14 was previously associated to schizophrenia (SZ) (Bureau et al. Biol Psychiatry 74(6): 444–450, 2013). Sequencing of 5 Mb of surrounding genomic DNA was undertaken in 12 subjects from 8 Eastern Quebec kindreds carrying the rs1156026 associated T allele. Genomic DNA was captured using custom microchips (Nimblegen, Roche; IBIS Laval University) and sequenced on a Illumina HiSeq sequencer (Génome Québec). Deleted forms of the deletion/insertions rs10569822, rs113886253 and rs146020684 were found on a haplotype segregating with the major psychosis (MP) phenotype comprising SZ, bipolar disorder (BP) and schizoaffective disorder in kindreds linked to the 13q13-q14 region (Maziade et al. EJHG, 17(8): 1034-1042, 2009). These deletions located in introns of the *DNAJC15* gene were genotyped in 1180 available subjects from 48 kindreds using custom specific PCR primers and DNA sequencers (LiCor). The three deletions were in strong linkage disequilibrium (r2 > 0. 9) and were significantly associated to BP (odds ratio (OR) = 1. 79, p = 0. 038 for rs10569822), but not to SZ or MP when compared to non-affected adult relatives in the 10 kindreds linked to the 13q13-q14 region (N=228). We next examined the association of the deletions and rs1156026 jointly, using a dominant model. We observed that the deletions were associated to BP among carriers of the T allele of SNP rs1156026 in the linked kindreds (OR = 3. 94, p=0. 012, for rs10569822). This association was not observed in subjects with the CC genotype of SNP rs1156026. We also observed that the odds ratio for SZ between carriers and non-carriers of the T allele of rs1156026 was independent of the deletions. These results reveal a complex interplay between genetic variants at 13q13-q14 and the risk of psychotic disorders, with some genotype combinations conferring risk of SZ and others conferring risk of BP.

1056T

Genetic and functional analysis of neurodevelopmental genes associated with schizophrenia in Indian population. *A. Jajodia[1], H. Kaur[1], K. Kumari[2], R. Baghel[1], M. Sood[2], R. Chadda[2], S. Jain[3].* 1) Genomic and Molecular Medicine Unit, CSIR-Institute of Genomics and Integrative Biology, Delhi, India; 2) Department of Psychiatry, All India Institute of Medical Sciences, Ansari Nagar, New Delhi 110029, India; 3) Molecular Genetic Laboratory, Department of Psychiatry, National Institute of Mental Health and Neuro Sciences, Hosur Road, Bengaluru 560029, India.

Schizophrenia is a complex psychiatric disorder with lifetime prevalence of about 1% worldwide. Several genotyping studies have suggested a strong genetic component to the disease. The present study was conducted to identify the genetic markers that are differentially associated with the disease in two ethnically diverse Indian populations, belonging to Indo-European and Dravidian ancestry. The genotyping study was performed using a custom panel of 1536 SNPs in 840 schizophrenia cases and 876 controls (351 patients and 385 controls from North India; and 436 patients, 401 controls and 143 familial samples from South India). Meta-analysis of these populations identified three strongly associated variants: STT3A (rs548181, p = 1. 47 × 10–5), NRG1 (rs17603876, p = 8. 66 × 10–5) and GRM7 (rs3864075, p = 4. 06 × 10–3). Further, this data was compared with the data from the Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC-SCZ) using meta-analysis, which supported rs548181 (p = 1. 39 × 10–7) variant in STT3A. In addition, combined analysis of sporadic case–control association and a transmission disequilibrium test in familial samples from South Indian population identified three associations: rs1062613 (p = 3. 12 × 10–3), a functional promoter variant of HTR3A; rs6710782 (p = 3. 50 × 10–3), an intronic variant of ERBB4; and rs891903 (p = 1. 05× 10–2), an intronic variant of EBF1. All these genes are involved in various neurodevelopmental processes. Variant rs1062613 present in the gene coding for serotonin receptor was further studied for its functional role. Our in silico simulation and in vitro gel shift assays revealed that the transcription factor CTCF (CCCTC-binding factor) binds at the polymorphic site rs1062613 (C/T). The polymorphism affects the DNA binding efficiency of CTCF, with a stronger affinity towards the wild-type C-allele. Moreover, the binding was found to be dependent on methylation at the C-allele. CTCF was found to bind with a greater strength to methylated cytosine. Our results suggest the role of epigenetic mechanisms in the development of schizophrenia by modulating transcription factor binding. The results support the risk variants observed in the earlier published work and suggest a potential role of neurodevelopmental genes in conferring risk towards schizophrenia.

1057F

Follow-up of linkage signals in the Quebec Kindred sample of schizophrenia and bipolar disorder using RNA expression and large scale SNVs analysis reveals potential candidate genes. *Y. Chagnon[1,3], A. Bureau[1,2], J. Croteau[1], A. Fournier[1], M-A. Roy[1,3], T. Paccalet[1], M. Maziade[1,3], C. Mérette[1,3].* 1) Centre de Recherche, Centre de recherche du CIUSSS Quebec, Quebec, PQ, Canada; 2) Médecine sociale et préventive, Univ Laval, Québec, PQ, Canada; 3) Psychiatrie et neurosciences, Univ Laval, Québec, PQ, Canada.

Introduction. Schizophrenia (SZ) and bipolar disorder (BP) are heritable but genome-wide association (GWA) studies have only partially uncovered possible involved genes. To further investigate linkage signals obtained in our sample of 48 extended pedigrees, we combined large scale RNA expression and single nucleotide variants (SNV) association analysis. **Methods.** We evaluated RNA expression in immortalized lymphocytes (IML) and SNVs in 453 subjects affected by SZ (N=59) or BP (N=84) and their unaffected relatives (N=313) from a subsample of 17 large kindreds of French-Canadian origin. Illumina microchips were used (HumanHT-12 Expression Beachip v4 and Infinium Human OmniExpress24; Génome Québec). We focused our analysis to the 12 linked chromosomal regions (LCR) previously identified in our large kindreds (Maziade Mol Psy 2005; Maziade EJHG 2009). After selection of the significantly detectable expressed probes using Illumina criteria, a statistical adjustement of RNA expression values was made for age, sex, housekeeping genes, and microchip. Residual expression data were then compared between affected and unaffected subjects, taking familial correlation into account. The probes showing some evidence of differential expression between unaffected subjects and SZ, BP or SZ+BP (p≤0. 01) were then evaluated for association (Zhao Biometrics 2014) with SNVs located at no more than 50 kb from the corresponding genes. **Results.** Of the 19195 probes showing significantly detectable expression in IMLs, 338 (SZ),187 (BP), and324 (SZ+BP) showed a differential expression between affected and unaffected family members. In LCRs, the strongest results for SZ+BP was observed for ITGB5 (p=8x10-5; p=0. 06 after Bonferroni) and, for SZ, was with the candidate gene ATXN1 (p=0. 002). RNA expressions of these two genes were confirmed by qPCR. Approximatly 197 SNVs were located within 50 kb of ATXN1 from which 92 were not in strong LD (r2 ≤ 0. 8) and showed a minor allele frequency (MAF) of at least 5%. Association to SZ+BP trough expression of ATXN1 was detected for nine of these SNVs (0. 0042≤p≤0. 045). This is consistent with our previously observed association with SZ and BP symptoms and the SCA1 repeat within ATXN1 (Chagnon Am Soc Hum Genet 2004). No significant SNV was observed for ITGB5. **Conclusion.** RNA expression maybe useful to identify potential candidate genes for SZ and BP within LCR (cis effect) while between LCRs, or trans effects, remain to be evaluated.

1058W

Next-generation profiling to identify the molecular etiology of Parkinson's dementia. *A. Henderson-Smith[1], J. Corneveaux[1], L. Cuyugan[2], M. Huentelman[1], W. Liang[2], T. Beach[3], T. Dunckley[1,3].* 1) Neurogenomics, TGen, Phoenix, AZ; 2) Collaborative Sequencing Center, TGen, Phoenix, AZ; 3) Banner Sun Health Research Institute, Sun City, AZ.

Parkinson's disease (PD) is characterized by degeneration of mid-brain dopaminergic neurons and a high prevalence of dementia associated with the spread of degenerative pathology to vulnerable cortical regions, including the posterior cingulate cortex. RNA sequencing (RNA-seq) is an attractive approach to uncovering the etiology of this and other complex neurodegenerative diseases. The present mRNA-seq study includes differential gene expression and alternative splicing analyses of the posterior cingulate cortex from neurologically normal control patients, patients with Parkinson's disease (PD), and patients with Parkinson's Dementia (PD-D). Genes overexpressed in both disease states were predictably involved with an immune response, while shared underexpressed genes function in signal transduction or a cytoskeleton component. Alternative splicing analysis produced a pattern of immune and RNA processing disturbances. Genes with the greatest degree of differential expression did not overlap with genes exhibiting significant alternative splicing activity. Resulting variation from the two types of expression analyses indicates the importance of broadening expression studies to include exon-level changes since there can be significant splicing activity with potential structural consequences, a subtlety that is not detected when examining differential gene expression alone, or is under-represented with probe-limited arrays.

1059T

Quantitative trait loci analysis of the serotonin-NAS-melatonin pathway in Autism Spectrum Disorders. *M. Benabou[1,2,3], A. Mathieu[1,2,4], J. Buratti[1,2,4], G. Huguet[1,2,4], R. Delorme[1,2,5,6], M. Leboyer[6,7,8,9], C. Pagan[10], J. Callebert[10], JM. Launay[10], T. Bourgeron[1,2,4,6].* 1) Human Genetics and Cognitive Functions Unit, Institut Pasteur, Paris, France; 2) CNRS UMR3571 Genes, Synapses and Cognition, Institut Pasteur, Paris, France; 3) University Paris Descartes, Sorbonne Paris Cité, Paris, France; 4) University Paris Diderot, Sorbonne Paris Cité, Paris, France; 5) AP-HP Department of Child and Adolescent Psychiatry, Robert Debré University Hospital, Paris, France; 6) Fondation Fondamental, Créteil, France; 7) AP-HP Department of Psychiatry, Henri Mondor-Albert Chenevier Hospital, Créteil, France; 8) INSERM U955, Institut Mondor de Recherche Biomédicale, Psychiatric Genetics, Créteil, France; 9) Faculty of Medicine, University Paris-Est, Créteil, France; 10) Service de Biochimie et Biologie Moléculaire, Department of Biochemistry, University Hospital Lariboisière, AP-HP, INSERM U942, Paris, France.

Increased blood serotonin and decreased melatonin have been repeatedly reported in patients with in Autism Spectrum Disorders (ASD). These observations have suggested that a disruption of the serotonin-NAS-melatonin pathway could be a risk factor for ASD. Several mutations affecting the main enzymes of this pathway, AANAT and ASMT were identified, but the underlying mechanisms of these biochemical anomalies remain largely unknown. It was previously shown that blood serotonin and melatonin levels are inheritable, but the genetic variants associated with these quantitative traits are unknown. Here, we measured serotonin blood level, the intermediate molecule N-acetylserotonin (NAS) and melatonin, as well as enzyme activity of AANAT and ASMT in a large cohort of 1196 individuals including 286 families with ASD and 308 controls. In addition, for a subset of these individuals (260 patients with ASD, 385 parents, 129 unaffected siblings, 30 affected siblings, 6 other relatives and 118 controls), we have genotyped more than 400 000 SNPs. We observed that serotonin and NAS levels were significantly increased in patients with ASD compared with the control group, while melatonin level, ASMT and AANAT activities were decreased ($p<0.0001$ for each trait). Each trait showed a significant parent-offspring correlation, which was particularly high for NAS (serotonin: 14%, $p=0.001$; AANAT: 27%, $p<0.0001$; NAS: 37%, $p<0.0001$; ASMT: 24%, $p<0.0001$; melatonin: 20%, $p=0.004$). The heritability of each trait was relatively high, particularly for NAS (serotonin: 30%, $p<0.0001$; AANAT: 38%, $p<0.0001$; NAS: 61%, $p<0.0001$; ASMT: 34%, $p=0.0002$; melatonin: 28%, $p=0.0003$). Our pilot genome wide quantitative trait loci (QTL) analysis has identified 5 genome wide significant SNPs. Three were identified for melatonin levels in patients and 2 of them (rs7993991/punadjusted=6.3x10-8 and rs11872926/punadjusted=1.2x10-7) were located in genes (*DNAJC3* and *DLGAP1* respectively). Two SNPs were also identified for NAS in the control group, but none of them was found in a known gene. In summary, our analyses provide the first estimates of the heritability for the whole serotonin-NAS-melatonin pathway in humans. The candidate SNPs associated with these traits should be confirmed in larger cohorts.

1060F

Function-based GWAS identifies novel candidate genes in autism spectrum disorder. *L. K. Davis[1], E. R. Gamazon[2], J. S Sutcliffe[3], E. H. Cook[4], N. J. Cox[2].* 1) Section Genetic Medicine, The University of Chicago, Chicago, IL; 2) Vanderbilt Genetics Institute, Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN; 3) Vanderbilt Brain Institute, Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN; 4) Department of Psychiatry, University of Illinois, Chicago, IL.

ASD is highly heritable and polygenic, with risk spread across thousands of DNA variants throughout the genome. Here we demonstrate a 'functional-unit' based GWAS, which, instead of utilizing all SNPs on the array (the vast majority of which will fall under the null expectation of no association), requires only a subset of SNPs that are coding (i. e. , missense, nonsense or frameshift) as well those that have been show by previous studies to be cis- or trans-eQTLs involved in regulating gene expression in brain. Coding variants and eQTLs are then assigned to the genes that they affect for a gene-based association analysis. The main objective of this study is to use biologically informative genomic annotations to effectively reduce the number of multiple tests conducted, thereby maximizing power to detect meaningful associations. We applied our method to both a discovery (N = 2,274) and replication sample (N = 2,976), and conducted a meta-analysis including both sets of results. We constructed gene-level annotations for each known gene in the human genome by annotating each gene with regulatory variation (i. e. , eQTLs) previously identified in the parietal cortex and the cerebellum as well as coding variants (nonsense, frameshift, and missense) within the gene. A gene-level p-value was obtained using a combined statistic that integrates SNP-level evidence, derived from the traditional single-variant test, for association with the trait. A total of 13,487 genes, represented by at least two variants, were annotated and included in each analysis. A Bonferroni corrected p-value of 3.7x10-6 was imposed to correct for the number of gene-based tests conducted. The most significant associations discovered were ZNF204P (2x10-5) within the AGRE/iControl analysis and ITFG2 (2x10-5) within the SSC/SAGE analysis, although neither result exceeded the threshold for genome-wide significance. However, upon meta-analysis, three genes were found to exceed genome-wide significance including ITFG2 (p=1.01x10-6), MTHFD2 (p=1.05x10-6), and ZNF2 (p=1.83x10-6). We present a novel gene-based approach to GWAS, which incorporates functional annotations including brain eQTLs and coding variants in a noise-reduction GWAS framework. We present meta-analysis data implicating three genes (ITGF2, MTHFD2, and ZNF2) and evaluate their potential as ASD candidate genes.

1061W

Linkage analysis revisited: three regions of potential risk found in extended multiplex Autism Spectrum Disorder (ASD) families. *s. Iuzi[1], J. Jaworski[1], M. L. Cuccaro[1,2], E. R. Martin[1,2], A. J. Griswold[1], H. N. Cukier[1], S. Slifer[1], I. Konidari[1], P. L. Whitehead[1], J. R. Gilbert[1,2], J. L. Haines[3], M. A. Pericak-Vance[1,2].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miller School of Medicine, miami, FL; 2) Dr. John T. Macdonald Department of Human Genetics, University of Miami, Miami, FL, USA; 3) Department of Epidemiology and Biostatistics, Institute for Computational Biology, Case Western Reserve University School of Medicine, Cleveland, Ohio, USA.

Autism spectrum disorder (ASD) is a highly prevalent neurodevelopmental disorder with a complex genetic architecture. It is characterized by severe deficits in three main areas: social interactions skills; verbal and non-verbal communication skills; behaviors and interests. The search for ASD genes has only identified ~25% of genetic components to date. Linkage studies are still considered a very useful tool to identify loci of interest in the presence of allelic heterogeneity, especially with large and well characterized pedigrees. We ascertained 86 extended of European ancestry with multiple cases of ASD and performed genome-wide genotyping followed by extensive parametric two-point and multipoint linkage analysis. For the parametric two-point models we assumed a dominant affecteds only model for analysis Two point results showed 31 autosomal regions with HLOD >3. 3. Of these, enhanced significance was noted at 3q24, 12q14 and 15q21 loci with HLOD>4. 4, HLOD>5. 4 and HLOD>3. 5, respectively. Multipoint parametric analysis confirmed our findings on the 3q24, 12q14 and 15q21 loci showing HLOD>2. 5, HLOD>2. 9 and HLOD>3, respectively, three regions previously implicated in ASD risk. These linkage regions will be compared with sequence data generated from the families in order to identify rare variants segregating with disease risk.

1062T

The burden of Multiple Sclerosis variants in continental Italians and Sardinians. *S. Sanna[1], N. Barizzone[2,3], I. Zara[4], M. Sorosina[5], S. Lupoli[6], E. Porcu[1,7], M. Pitzalis[1], M. Zoledziewska[1], F. Esposito[5], A. Mulas[1,7], E. Cocco[8], F. R. Guerini[9], P. Brambilla[5], G. Farina[10], R. Murru[8], F. Deidda[1], So. Sanna[1], A. Loi[1,7], C. Barlassina[6], D. Vecchio[11], A. Zauli[5], F. Clarelli[5], D. Braga[6], F. Poddie[7], J. Frau[8], L. Lorefice[8], D. Cusi[6], F. Cucca[1,7], F. Martinelli Boneschi[5,12], S. D'Alfonso[2,3].* 1) IRGB-CNR, Monserrato, CA, Italy; 2) Interdisciplinary Research Center of Autoimmune Diseases IRCAD, University of Eastern Piedmont, Novara, Italy; 3) Department of Health Sciences, University of Eastern Piedmont, Novara, Italy; 4) Centro di Ricerca, Sviluppo e Studi Superiori in Sardegna (CRS4), Pula (Cagliari); 5) Laboratory of Genetics of Complex Neurological Disorders, Institute of Experimental Neurology (INSPE), Division of Neuroscience, San Raffaele Scientific Institute, Milan, Italy; 6) Department of Health Sciences, University of Milano, Milan, Italy; 7) Department of Biomedic Science, University of Sassari, Sassari, Italy; 8) Centro Sclerosi Multipla, Department of Public Health, Clinical and Molecular Medicine, University of Cagliari, Cagliari, Italy; 9) Don C. Gnocchi Foundation IRCCS, Milan, Italy; 10) Department of Clinical and Experimental Medicine, University of Sassari, Italy; 11) Department of Transational Medicine, Section of Neurology, University of Eastern Piedmont, Novara, Italy; 12) Department of Neurology, Division of Neuroscience, Scientific Institute San Raffaele, Milan, Italy.

Recent studies identified >100 non-HLA Multiple Sclerosis (MS) susceptibility variants in Northern European populations, but their role in Southern Europeans is unclear. Here we investigated the cumulative impact of those variants in two Mediterranean populations: continental Italians and Sardinians. We calculated four weighted Genetic Risk Scores (wGRSs) using a log additive sum of risk alleles at up to 102 non-HLA MS risk variants and 5 HLA MS susceptibility markers in 1,691 patients and 2,194 controls from continental Italy, and 2,861 patients and 3,034 controls from Sardinia. We then assessed differences between populations using Nagelkerke's R2 and the area under the ROC curves. As expected, the genetic burden (mean wGRS value) was significantly higher in MS patients than in controls, in both populations (p<10-45 for all wGRSs). Of note, the burden was remarkably higher in Sardinians (p<10-20 for all wGRSs), and this difference was not attributable to 14 non-HLA variants showing different allele frequencies in continental Italians (Fst>0. 01). Conversely, the proportion of variability explained and the predictive power of the assessed variants were significantly higher in continental Italians (AUC 0. 725 versus 0. 681 in Sardinians, p=4x10-5). This genetic risk score is not meaningful in clinical settings, but the observed difference between populations suggests that either the SNPs identified through large general populations are capturing the causal variants less efficiently in Sardinians and/or additional risk variants remain to be detected in this population. Furthermore, Sardinians showed another peculiar feature, with a notably higher burden of non-HLA variants in patients who do not carry HLA risk alleles (p=7x10-4). No similar trend was seen in controls. This suggests a counter-balancing aggregation effect of non-HLA loci versus HLA markers, which could be explained by selective pressure and/or interaction with environmental factors. The observed differences in the MS genetic burden between these two Mediterranean populations highlight the need for more genetic studies in these and other South Europeans to further expand the knowledge of genetics of MS.

1063F

Identification of pleiotropic loci involved in the comorbidity of addiction with major depression and anxiety using randomly ascertained extended pedigrees. *K. Hodgson[1], L. Almasy[2], E. E. M. Knowles[1], J. W. Kent Jr. [3], S. R. Mathias[1], N. Yao[1], J. E. Curran[2], T. D. Dyer[2], H. H H Göring[2], R. L. Olvera[4], R. Duggirala[2], P. T. Fox[5,6], J. Blangero[2], D. C. Glahn[1,7].* 1) Department of Psychiatry, Yale University School of Medicine, New Haven, CT; 2) South Texas Diabetes and Obesity Institute, University of Texas Health Science Center at San Antonio & University of Texas of the Rio Grande Valley, Brownsville, TX; 3) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 4) Department of Psychiatry, University of Texas Health Science Center at San Antonio, San Antonio, TX; 5) Research Imaging Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX; 6) South Texas Veterans Health System, 7400 Merton Minter, San Antonio, TX; 7) Olin Neuropsychiatric Research Center, Institute of Living, Hartford Hospital, Hartford, CT.

Comorbidity is common in addiction, with patients frequently having additional diagnoses of major depression and anxiety disorders. The genetic architecture of these comorbidities remains unclear, but could provide an important insight into the underlying etiology of these disorders. To investigate these genetic relationships, we used a randomly ascertained sample of Mexican-American extended pedigrees. Variance decomposition methods were used to explore the role of genetics in addiction phenotypes, depression and anxiety disorders. Genome-wide univariate and bivariate linkage scans were conducted to localize the chromosomal regions influencing these traits and the high rates of comorbidity observed between them. In the univariate analyses, quantitative trait loci were identified for substance dependence (14q13. 2-q21. 2, LOD=3. 322), cocaine use (7q34, LOD=3. 211), anxiety (12q24. 32-q24. 33, LOD=2. 918) and depression (22q11. 1-q11. 21, LOD=3. 144). Considering the overlap of addiction with depression and anxiety, significant genetic correlations were observed and bivariate linkage analyses were used to identify loci with pleiotropic effects on these correlated traits. Bivariate quantitative trait loci were found for alcohol dependence-anxiety (9q33. 1-q33. 2, LOD=3. 054), substance dependence-anxiety (18p11. 23-p11. 22, LOD=3. 425) and cannabis use-major depression (11q23. 1-q23. 2, LOD=3. 229). This study confirms the shared genetic underpinnings of addiction with depression and anxiety and identifies a number of genomic loci involved in the etiology of these comorbid disorders. These quantitative trait loci show convergence with previous literature in the field, but also implicate novel regions of the genome in addiction genetics.

1064W

Genome-wide association study for facet-level traits of neuroticism in a health study cohort of Korean adults. *S. Kim[1], H. Kim[1], H. Shin[2], J. Cho[2], M. J. Kwon[2], Y. Chang[2], S. Ryu[2], H. L. Kim[1].* 1) Ewha Womens University, Seoul, South Korea; 2) School of Medicine, Sungkyunkwan University, Seoul, South Korea.

Neuroticism is a personality trait closely linked to psychological distress and a risk ofmood disorders. Converging evidence indicates that neuroticism shared genetic and biological components of psychiatric disorders. The revised NEO Personality Inventory (NEO-PI-R) measures neuroticism which consists of six facets of traits : anxiety (N1), angry hostility (N2), depression (N3), self-consciousness (N4), impulsivity (N5), vulnerability (N6). In the present study, we conducted the genome wide association study (GWAS)for neuroticism in the lower-order levels. The analysisincluded 2,064 individuals aged 19-69 years. Anxiety (N1) was associated with rs76982298 (1. 54x10-6) located in MAPK10 (mitogen-activated protein kinase 10) gene. MAPK signaling was previously known to affect anxiety and depressive-like behavior in mouse model. Depression (N3) was associated with rs8088304 (P=9. 96x10-6) located withMBP (myelin basic protein) gene. Reduced mRNA transcription of MBP was found in schizophrenia in the previous report. Self-consciousness (N4)was linked to rs4557199 (2. 53x10-7)located near ROBO2 (roundabout, axon guidance receptor homolog 2) gene. The current findings usingGWAS of neuroticism in facet level showed suggestive loci related with genetic susceptibility of mental disorders. In addition, the present study might produce insights into the biological pathways of psychological factors associated with various psychiatric disorders.

1065T

Genetic loci in periodic hypersomnia/Kleine-Levin syndrome type. *E. Mignot[1], R. Hillary[1], HM. Ollila[1], J. Faraco[1], L. Lin[1], I. Arnulf[2].* 1) Stanford University, Palo Alto, CA. USA; 2) Sleep Disorder Unit, Pitie-Salpetriere Hospital, Assistance Publique - Hôpitaux de Paris, 75013 Paris, France.

Background:Kleine-Levin syndrome (KLS) is a rare sleep disorder that affects ~ 1 person in a million and has been suggested to be more frequent in Ashkenazi Jewish. The disorder typically strikes adolescent males but improve with age, often resolving by age 30. In KLS, patients have recurrent episodes lasting several weeks when they sleep nearly 24 hours per day. When awake, patients during episodes experience apathy, cognitive disturbances and occasionally hyperphagia and hypersexuality. Between episodes, patients are totally normal. Aim: Our aim was to identify genetic variants that contribute to KLS predisposition. Methods: As a part of an international collaboration we performed GWAS in 650 KLS cases and 30, 000 controls. The sample comprised KLS cases and matched controls from United States, Europe and Asia and additional controls from the GERA consortium. Genotyping was done using Affy 6. 0 and Affymetrix Axiom World Array with ethnicity specific platforms that were imputed to 1000 genomes. Analyses were controlled for population stratification and ethnicity (Caucasian, Ashkenazi Jewish, Asians, other) . Results: Genome-wide significant loci were found near TRANK1 and ATXN1. Most interestingly, the leading TRANK1 variant is same as reported in other GWAS for schizophrenia and Bipolar disorder. Conclusions: The findings give the first biological evidence for disease mechanisms in KLS. Importantly, these results suggest a partially overlapping genetic composition for schizophrenia and KLS. The fact these patients are not primarily depressed or psychotic during episodes, completely reverse to normality between episodes could suggest important pathophysiological clues linking sleep and episodic psychiatric conditions.

1066F

An international, collaborative genome-wide association study of Tourette Syndrome in 13,000 individuals identifies a non-coding RNA expressed early in human brain development as a TS susceptibility gene. *J. M. Scharf[1,2] on behalf of the TSAICG, GGRI and TIC Genetics consortia.* 1) Mass General Hospital/Harvard Medical School, Boston, MA; 2) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA.

Tourette Syndrome (TS) is a childhood-onset developmental disorder with one of the highest heritabilities among neuropsychiatric diseases with non-Mendelian inheritance. TS is more common than previously recognized, and is now believed to be present in 0. 3-0. 8% of children, many of whom have additional, co-occurring neuropsychiatric disorders, such as OCD, ADHD, or autism-spectrum disorder (ASD). We previously demonstrated that TS is highly polygenic and that the majority of TS heritability can be captured by SNP arrays. However, to date the sample size of TS genetic studies has been too small to identify genome-wide significant loci. Here, we present results from the largest TS genome-wide association study (GWAS) to date and a mega-analysis of all existing TS GWAS data, totaling 4,232 European ancestry cases and 8,919 ancestry-matched controls. After standard QC procedures, the final analysis had a lambdaGC=1. 065 (lambda1000=1. 014) including SNPs with MAF>1% imputed to 1000 Genomes. One SNP, rs2708146, on chromosome 2p16. 1, achieved genome-wide significance (p=2. 8x10-8). This SNP lies within a 1Mb haplotype spanning an uncharacterized long, intergenic, non-coding RNA (lincRNA), *LINC01122,* that is expressed primarily in mid-fetal human brain development (15-16 pcw), with strongest expression in developing cerebral cortex. Pairwise correlational analysis of 22,731 genes from the Brainspan project expressed across 7 human pre-natal time points and 16 brain regions demonstrated the highest co-expression with *BCL11A*, a gene 1 Mb centromeric to *LINC01122* that has been found previously to harbor recurrent *de novo* loss-of-function mutations in ASD samples. Subsequent analyses have integrated multiple sources of epigenomic data from human brain tissue and *in vitro* human stem cell models of developing cortical neurons, including 1) H3K27Ac peaks to mark active gene enhancers, 2) developmental time-point-specific gene expression to identify regional genes that correlate with enhancer activity, and 3) biophysical models to predict changes in transcription factor binding site affinity caused by candidate causal GWAS SNPs lying within these active enhancers. These data, combined with functional *in vitro* expression studies and *in vivo* mouse experiments are in progress to characterize this regulatory locus further and its potential role in the development of cortico-striatal-thalamic-cortical circuits that have been hypothesized to be abnormal in TS patients.

1067W

ADHD risk loci identified by genome-wide association meta-analysis. *D. Demontis, the iPSYCH-SSI-Broad/MGH ADHD Workgroup and the Psychiatric Genomics Consortium: ADHD Subgroup.* Biomedicine, Aarhus University, Aarhus, Århus, Denmark.

Attention-deficit hyperactivity disorder (ADHD) is a highly heritable common childhood behavioural disorder affecting 3-6% of school-age children and ~4% of adults around the world. Several moderately sized genome-wide association studies (GWASs) have been performed, but until now no single markers have passed the threshold for genome-wide significance. The SNP heritability of ADHD has been estimated to 0. 28, indicating that common SNPs contribute substantially to ADHD susceptibility and that increasing GWAS sample sizes is likely to produce significant results. Here we present the meta-analysis of GWASs of ADHD, which is a large-scale collaboration, between the Danish *i*PSYCH (Lundbeck Foundation Initiative for Integrative Psychiatric Research) program, the Broad Institute and the Psychiatric Genomics Consortium (PGC). The study, which represents the latest ADHD freeze, includes ten ADHD PCG samples, in total consisting of ~6000 cases and ~14500 controls and the iPSYCH sample in total comprising ~8700 cases and ~11300 controls. Standard quality control procedures and genome-wide association analyses are performed on each dataset individually and subsequently a meta-analysis is conducted across datasets. A 23andMe self-reported ADHD dataset is used as replication sample in order to follow-up on the strongest independent signals from the GWAS meta-analysis. Results from the primary GWAS meta-analysis and the combined analysis of the PGC samples, the iPSYCH sample and the 23andMe replication sample will be presented. The results will reveal the first genome-wide significantly associated loci with ADHD. SNP heritability estimates for evaluating to what extent, common genetic variants explain the heritability will also be presented. Risk score profiles will be calculated and subsequently odds ratios will be estimated in order to evaluate the increased risk of ADHD for individuals belonging to the group with the highest burden of risk alleles. Additionally enrichment analyses will be presented elucidating e. g. potential enrichment of SNPs located in brain specific enhancer regions among the most associated markers. The results presented here include the identification of the first genome-wide significant loci in ADHD. Identification of specific risk loci as well as the enrichment analyses and heritability estimates presented represent an important step towards a better understanding of the genetic architecture of ADHD.

1068T

A novel oligogenetic framework identifies age of onset modifying mutations in Alzheimer's disease. *J. I. Vélez[1,2], C. A. Mastronardi[1], H. R Patel[1], A. S. Johar[1], F. Lopera[2], M. Arcos-Burgos[1,2].* 1) John Curtin School of Medical Research, The Australian National University, Canberra, ACT, Australia; 2) Neuroscience Research Group, University of Antioquia, Medellín, Colombia.

The identification of modifier genes (MGs) in complex traits, including Alzheimer's disease (AD), is of great interest in genetic research. Modifier genes can alter the expression of another gene, consequently affecting the natural history of complex traits. For example, MGs can affect penetrance, dominance, expressivity and pleiotropy, and are responsible for changes in the age of onset (AOO) and disease progression. Here we present a novel oligogenetic framework for the identification of MGs that combines the use of genetic isolates, high-throughput sequencing, extreme phenotype sampling, and advanced data mining tools. We applied this framework to patients from the Paisa genetic isolate in Antioquia, Colombia, who belong to the world's largest known pedigree segregating early-onset familial AD caused by the p. Glu280Ala (E280A) fully penetrant mutation in the *PSEN1* gene. Whole-exome genotypes from 71 mutation carriers with AD from this pedigree were analyzed using linear mixed-effects models. Four AOO protective variants harbored in the *APOE* (rs7412, *P*FDR = 2. 13x10-30), *GPR20* (rs36092215, *P*FDR = 6. 58x10-22), *FCRL5* (rs16838748, *P*FDR = 8. 61x10-10), *RC3H1* (rs10798302, *P*FDR = 1. 86x10-4) genes with AOO delaying effects of up to 17 years above the average AOO in this pedigree (of ~48 years), were identified. Similarly, five harmful variants in the *TRIM22* (rs12364019, *P*FDR = 1. 15x10-14), *AOAH* (rs12701506, *P*FDR = 5. 69x10-8), *PINLYP* (rs2682585, *P*FDR = 1. 67x10-6), *IFI16* (rs62621173, *P*FDR = 8. 63x10-6) and *DFNA5* (rs754554, *P*FDR = 3. 62x10-2) genes resulted in an AOO that is up to 12 years ealier than average. Genes harboring these variants play an important role in cell proliferation, protein degradation, apoptotic and immune dysregulation processes, and could be important as eventual therapeutical targets. Data mining techniques were used to construct a model with some of these genetic variants to predict the development of early- or late-onset AD in E280A mutation carriers. The results of this study provide the foundation for better management of these patients and enhanced assessment, highlighting the potential of this oligogenetic model in the clinical setting.

1069F

Genome wide association in families identifieds a genetic basis for two components of speech, phonological processing and articulation. *H. Voss-Hoynes[1], C. Stein[1], B. Truitt[1], J. Tag[2], L. Freebarin[2], B. Lewis[2], S. K. Iyengar[1].* 1) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH.

Speech sound disorders (SSD) are communication disorders occurring in 16% of 3 year olds. These disorders can affect phonological processes—sound patterning—or articulation—sound making. One phonological process, syllable sequencing, can be measured with multisyllabic word repetition tasks (MSW) while accuracy in articulation of consonant sounds is commonly measured with the Goldman-Fristoe Test of Articulation (GFTA). Using 1. 5 million markers on the Omni 2. 5 Exome chip, we performed a separate quantitative trait genome wide association analysis in PLINK on the Z-scores of MSW and GFTA scores. While no SNPs reached the multiple testing threshold of 3. 5x10-8, analysis of 545 Caucasian individuals (147 families) identified interesting loci. The 50 most significant SNPs (p=1x10-6 - 1. 1x10-5) for each phenotype did not overlap. The top SNPsfor MSW (rs73075822 and rs56260247, p=1x10-6) are intronic and located in CREB5 which is a transcription factor on chromosome 7. Another SNP of interest (rs116882098 p= 6x10-6) is located in GRIK2, a gene previously associated with intellectual disability, epilepsy, and behavioral disorder. While there does not seem to be an overt pattern for MSW genes, for GFTA, the top SNPs tend to be found in genes associated with growth and development. The top three SNPs are located in DAB1 (rs78728168 p=1x10-6), SH3B4 (rs74789325 p=1x10-6), and SLIT3 (rs35506927 p=1x10-6). DAB1 encodes a regulator of the reelin pathway, necessary for cell positioning in the brain; SH3B4 is thought to indirectly regulate cell growth and proliferation and has been associated with peripheral neuropathies; and SLIT3 encodes a slit protein which prevents developing neurons from crossing the midline. These findings indicate that despite phenotypic correlation, the genetic basis of phonological processing and articulation may have some distinct features. While the exact variants we identified are unique, there are similarities with GWAS of endophenotypes associated with language and reading. Like to the association between GRIK2 and behavioral disorder, ZNF385D has been associated with language impairment and ADHD. Additionally the neural-growth control function if SLIT3 is similar to a gene associated with dyslexia, NODAL, which maintains neural asymmetry. Future work will determine if the unique gene sets for each component of speech, language, and reading participate in the same networks. Acknowledgements: DC00528, DC012380, T32-HL007567.

1070W

Genome-wide linkage and family-based association in the Western Australian Family Study of Schizophrenia (WAFSS). *N. S. McCarthy[1], P. E. Melton[1], J. C. Badcock[2], M. Dragovic[2], B. Morar[2], V. A. Morgan[2], E. K. Moses[1], A. Jablensky[2].* 1) Centre for the Genetic Origins of Health and Disease, The University of Western Australia, Perth, Western Australia, Australia; 2) Centre for Clinical Research in Neuropsychiatry, University of Western Australia, Perth, Western Australia, Australia.

Background: Family studies provide some advantages over unrelated cohorts in the potential identification of the 'missing heritability' for SZ and other complex diseases. Pedigrees are potentially enriched for rare variants, and are less susceptible to the problem of population stratification. In addition, families are ideal for the detection of pleiotropy, an important consideration in a known heterogeneous disorder like SZ. In the present study, we conducted a genome-wide linkage and association analyses in the Western Australian Family Study of Schizophrenia (WAFSS) to identify genomic regions associated with increased SZ risk. **Methods:** Our WAFSS study sample included 336 individuals of European descent from 64 families with multiple cases of SZ or spectrum disorder. Clinical assessment included a standardised diagnostic interview and best-estimate diagnosis (ICD-10 and DSM-IV), both established by two senior clinicians. Clinical symptoms were comprehensively assessed with the Diagnostic Interview for Psychosis. Genotyping was performed on the *Illumina HumanCoreExome Beadchip*, which assays approximately 250,000 common tag single nucleotide polymorphisms and 250,000 moderately rare, exonic variants. Standard quality control filters were applied using *Plink*. For linkage analysis, variants were thinned based on linkage disequilibrium (r2 = 0. 8). Non-parametric linkage (NPL) and family-based association analyses were performed separately adjusting for age and sex using *Merlin*, and logarithm of the odds (LOD) scores calculated. **Results**: For the family-based association, the most significant genetic signal was located downstream of the glutamate receptor gene (*GRIN3A)* on chromosome 9 (rs1022749, LOD 4. 1, *P*=1. 4x10-5). In the NPL analyses the most significant genetic signal was located upstream of the RNA-binding protein *CUGBP2* on chromosome 10 (rs2398243, LOD 3. 1, *P*=7. 6x10-5). **Discussion:** Comparison of these genome-wide linkage and association results demonstrate the complementary nature of these two methods and implicate different genomic regions. The role of the NMDA-receptors is well described in SZ, and other variants in *GRIN3A* have previously been associated with SZ. In contrast, little is known about the potential role of *CUGBP2* in SZ. Additional work will include gene-centric burden testing using rare, exonic variants and prioritizing genomic regions implicated in this study for whole genome sequence data analysis in the WAFSS cohort.

1071T

UK BiLEVE, the first genetic study in UK Biobank, identifies 5 novel regions associated with smoking behaviour. *M. D. Tobin[1,13], L. V. Wain[1], N. Shrine[1], S. Miller[2], V. Jackson[1], I. Ntalla[1], M. Soler Artigas[1], J. P. Cook[1], A. P. Morris[3], E. Zeggini[4], J. Marchini[5,6], P. Deloukas[7], A. Hansell[8], R. Hubbard[9], I. Pavord[10], N. C. Thomson[11], D. P. Strachan[12], I. P. Hall[2], UK BiLEVE consortium.* 1) Department of Health Sciences, University of Leicester, Leicester, United Kingdom; 2) Division of Respiratory Medicine, University of Nottingham, Queen's Medical Centre, Nottingham, United Kingdom; 3) Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom; 4) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom; 5) Department of Statistics, University of Oxford, Oxford, United Kingdom; 6) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 7) William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University London, London, United Kingdom; 8) Faculty of Medicine, School of Public Health, Imperial College London, London, United Kingdom; 9) Faculty of Medicine and Health Sciences, School of Medicine, University of Nottingham, Nottingham, United Kingdom; 10) Respiratory Medicine, University of Oxford, Oxford, United Kingdom; 11) Institute of Infection, Immunity & Inflammation, University of Glasgow, United Kingdom; 12) Population Health Research Institute, St George's University of London, London, United Kingdom; 13) National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester, United Kingdom.

UK Biobank recruited 502,682 UK individuals aged 40-69 years (95% European ancestry), collecting detailed health and lifestyle information, DNA, and physical measures, including spirometry and smoking behaviour. In the UK Biobank Lung Exome Variant Evaluation (UK BiLEVE) we genotyped 50,000 samples from the extremes and middle of the lung function distribution amongst 25,000 heavy smokers (mean 35 pack years) and 25,000 never smokers in order to study the genetics of lung function and smoking behaviour. Genome-wide genotyping was undertaken using a custom designed Affymetrix array that included 130K rare missense and loss of function variants (selected to be polymorphic in UK populations based on currently available "exome chip" data), 642K variants selected for optimal imputation of common variation and improved imputation of low frequency variation (MAF 1-5%). A case-control analysis between the heavy and never smokers was undertaken across 28. 5 million genotyped and imputed variants. We discovered five new genome-wide significant (P<5x10-8) signals for smoking behaviour at 1p21. 3 (*LPPR5*), 2q22. 3, 6p21. 2 (*DNAH8*), 11q23. 2 (*NCAM1*) and 20q11. 21 (*NOL4L*), and confirmed two previously reported loci at 15q25 and 7p14. Two of our novel signals separately implicate *NCAM1* (neural cell adhesion molecule 1): one SNP is located within an intron of *NCAM1*; a second variant on a different chromosome (2q22. 3) and has a *trans* effect on expression of *NCAM1* in brain tissue. The 2q22. 3 variant is also a *trans* eQTL in substantia nigra (brain) tissue for *WDR61,* which is 300kb from the known 15q25 locus containing the nicotine-receptor genes *CHRNA3* and *CHRNA5*. Another novel signal in *NOL4L* is near to, but independent of, a SNP in *C20orf203* previously implicated in nicotine dependence. By conditioning on the primary signals, we identified secondary independent signals, which did not reach genome-wide significance, at 3 of the loci associated with smoking behaviour including a novel rare (MAF=0. 09%) intergenic SNP near *NCAM1*. A further 8 loci with suggestive evidence of association with smoking behaviour (5x10-8<P<5x10-7) include a variant in *CHRNA4* (20q13. 33), a gene previously implicated in smoking behaviour. The genotype data for the UK BiLEVE study has been made available to the scientific community via UK Biobank, as has subsequent data from a similar array in additional UK Biobank participants. This research has been conducted using the UK Biobank Resource.

1072F

Investigation into the Genetics Underlying Receptive Vocabulary Processing. *A. K. Adams[1], J. D. Eicher[1], N. R. Powers[1,2], N. T. Truong[2], J. R. Gruen[1,2,3].* 1) Department of Genetics, Yale University, New Haven, CT; 2) Department of Pediatrics, Yale University, New Haven, CT; 3) Department of Investigative Medicine, Yale University, New Haven, CT.

Receptive vocabulary processing is a complex neurobehavioral process requiring the integration of many different pieces of information for proper language comprehension. Receptive vocabulary processing is a genetically complex trait, which has only recently come under study from a genomic perspective. Developmental language impairment (despite normal intelligence) has been shown to be heritable, with twin studies revealing heritability estimates of greater than 50%. Studies have also shown a prevalence of 7. 4% in 5-year-old Midwestern children (Tomblin et al. , 1997). Using results from the Peabody Picture Vocabulary Test (PPVT) for receptive vocabulary, and a 550,000 marker genetic panel, an age-corrected genome wide association study (GWAS) was undertaken on 440 normally developing children of European ancestry from the Pediatric Imaging, Neurocognition, and Genetics (PING) cohort. This study revealed a single genome wide association with rs11918964 in the gene *MAGI-1* (p=3. 66 x 10 -8) as well as suggestive associations with other SNPs in both *MAGI-1* and the gene *BMP7*. *MAGI-1* is a scaffold protein that has previously been linked to neural disorders including schizophrenia, while *BMP7* is a gene affiliated with neural development. As a follow-up study to further investigate these results, both candidate SNP and full genome wide studies have been undertaken using the Hispanic American population (n=856) genotyped on a 2. 5 million marker array in the Genes, Reading, and Dyslexia (GRaD) cohort. These studies are ongoing and results are forthcoming.

1073W

Cross disorder analyses in attention deficit hyperactivity disorder and autism spectrum disorder - an update. *M. Mattheisen[1,2], The iPSYCH-SSI-Broad/MGH, the PGC ADHD, and the PGC ASD working groups.* 1) Department of Biomedicine, Aarhus University, Aarhus, Denmark; 2) Lundbeck Initiative for Integrative Psychiatric Research (iPSYCH).

The recent success in GWAS for schizophrenia (SCZ) has helped to create a momentum in psychiatric genetics that now also spreads into other mental disorders. One of the positive developments is a substantial increase in available samples for genetic studies. Especially in childhood onset disorders, such as ADHD and ASD, contributions from private foundations and philanthropy helped to achieve an at least 2-3 fold increase in sample size for these disorders when compared to their state 1 year ago. Here, we will present our cross-disorder analyses for the two largest ADHD and ASD GWAS data sets currently available worldwide (each comprising of approx. 15k cases) with the aim to study their shared genetic risk architecture at multiple levels. The datasets used in this study stem from the individual disorder working groups in the PGC as well as a large scale collaborative project between iPSYCH, the Statens Serum Institut (SSI), and the Broad/MGH. The latter dataset is a population-based sample comprised of individuals born in Denmark after 1991 that have been diagnosed with either ADHD or ASD (a total of approx. 15k cases). We will report results from analyses that use LD score regression (LDSC), GCTA, and polygenic risk scores to study SNP heritability and genetic correlation (rG) of three sub-groups comprised of individuals diagnosed with (1) ADHD only, (2) ASD only, and (3) ADHD and ASD. For these sub-groups we will also report rG with SCZ and other adulthood onset disorders, as well as related quantitative traits (e. g. brain imaging phenotypes and cognitive traits). Furthermore, results will be presented for a full meta-analysis of all cases diagnosed with either ADHD and ASD and the respective controls. Finally, we will employ a multivariate GWAS approach in a large subset of our data (iPSYCH-SSI-Broad/MGH data) for which we have multiple psychiatric and somatic phenotypes available. Preliminary analyses based on 85% of our final dataset suggest four genome-wide significant regions for the meta-analysis of ADHD and ASD. This includes a region on chromosome 5 that harbors miR-9 and showed the strongest evidence for association in our analyses. An analysis of the final dataset is needed to provide further evidence for the importance of our preliminary findings in ADHD and ASD etiology.

1074T

A polymorphism in *CRAT* is associated with *HLA-DQB1*06:02* negative essential hypersomnia. *T. Miyagawa[1], SS. Khor[1], H. Toyoda[1], H. Kojima[2], T. Futagami[2], M. Yamasaki[1], H. Saji[2], K. Mishima[3], Y. Honda[4], M. Honda[4, 5], K. Tokunaga[1].* 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) HLA Foundation Laboratory, Kyoto, Japan; 3) Department of Psychophysiology, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan; 4) Japan Somnology Center, Neuropsychiatric Research Institute, Tokyo, Japan; 5) Sleep Disorders Project, Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan.

Essential hypersomnia (EHS) is a lifelong sleep disorder characterized by excessive daytime sleepiness but no cataplexy, can be divided into two broad classes based on the presence or absence of *HLA* (human leukocyte antigen)-*DQB1*06:02* allele. *HLA-DQB1*06:02* positive EHS and narcolepsy are associated with the same susceptibility genes. However, there are fewer studies of *HLA-DQB1*06:02* negative EHS. We hypothesized that *HLA-DQB1*06:02* negative EHS has a different etiological pathway from that of narcolepsy and *DQB1*06:02* positive EHS. Therefore, in order to identify susceptibility genes associated with *HLA-DQB1*06:02* negative EHS, we performed a genome-wide association study of 476,446 SNPs in 119 Japanese patients with *HLA-DQB1*06:02* negative EHS and 1,582 Japanese healthy individuals. A replication study was also conducted on 191 Japanese patients with *HLA-DQB1*06:02* negative EHS and 433 Japanese healthy individuals. SNP rs10988217 located in *CRAT* (carnitine acetyltransferase) was found to be significantly associated with *HLA-DQB1*06:02* negative EHS ($P < 5\times10\text{-}8$, odds ratio = 2. 8). An eQTL analysis showed that SNP rs10988217 was significantly correlated with expression levels of *CRAT* in various tissues or cell types, including brain tissues ($P < 0. 05$). *CRAT* gene encodes the carnitine acetyltransferase protein, which is a key enzyme for metabolic pathways involved with the control of the acyl-CoA/CoA ratio in mitochondria, peroxisomes and endoplasmic reticulum. The Metabolomics GWAS Server (Shin SY et al. , Nat Genet. doi: 10. 1038/ng. 2982.) showed that rs10988217 affected succinylcarnitine levels in blood ($P < 10\text{-}17$). The results suggest that *HLA-DQB1*06:02* negative EHS may have an underlying dysfunction in fatty acid oxidation pathway.

1075F

Search for new susceptibility genetic factor associated with panic disorder-pathway analyses based on genome-wide association study-. *M. Shimada-Sugimoto[1], T. Otowa[2], T. Miyagawa[1], S. Khor[1], K. Kashiwase[3], N. Sugaya[4], Y. Kawamura[5], T. Umekage[6], H. Kojima[7], T. Futagami[7], H. Saji[7], A. Miyashita[8], R. Kuwano[8], H. Kaiya[9], K. Kasai[2], H. Tanii[10], Y. Okazaki[11], K. Tokunaga[1], T. Sasaki[12].* 1) Department of Human Genetics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 2) Department of Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 3) Japanese Red Cross Kanto-Koshinetsu Block Blood Center, Tokyo, Japan; 4) Department of Epidemiology and Public Health, Graduate School of Medicine, Yokohama City University, Kanagawa, Japan; 5) Department of Psychiatry, Sakae Seijinkai Hospital, Kanagawa, Japan; 6) Division for Environment, Health and Safety, The University of Tokyo, Tokyo, Japan; 7) HLA Foundation Laboratory, Kyoto, Japan; 8) Department of Molecular Genetics, Center for Bioresources, Brain Research Institute, Niigata University, Niigata, Japan; 9) Panic Disorder Research Center, Warakukai Med. Corp. , Tokyo, Japan; 10) Department of Psychiatry, Institute of Medical Life Science, Graduate School of Medicine, Mie University, Mie, Japan; 11) Tokyo Metropolitan Matsuzawa Hospital, Tokyo, Japan; 12) Department of Physical and Health Education, Graduate School of Education, The University of Tokyo, Tokyo, Japan.

Panic disorder (PD) is one of anxiety disorders characterized by panic attacks and anticipatory anxiety. Both genetic and environmental factors have been thought to trigger PD onset, but little is known about the pathogenesis of PD. To date, *TMEM132D*, which was found to be associated with PD in European genome-wide association study (GWAS), is the only genetic factor that was replicated in several studies. We previously performed GWAS with Japanese PD samples and found some genes with marginal associations with PD. However, there seems to be other genetic factors which remain to be elucidated by the analysis just focusing on SNPs with low P values. In this study, we performed pathway analyses and aimed to identify new genetic factors, considering the possibility that functional combinations of SNPs would affect PD pathogenesis. We performed three different types of pathway analyses. Each analysis indicated that immune-related pathways showed the strongest association with PD (DAVID, P = 2. 08 ×10-6, i-GSEA4GWAS, P < 10-3; ICSNPathway, P < 10-3) and that human leucocyte antigen (HLA) had the potential of contributing to the immune-related association. We then re-examined the previous GWAS findings, focusing on the *HLA* region, and found that several SNPs in this region showed suggestive associations with PD, even though these associations were not significant at the genome-wide level. Based on these results, we examined *HLA-DRB1* and *HLA-B* regions as susceptibility locus for PD in 744 subjects with PD and 1418 control subjects. Patients with PD were found to be significantly more likely to carry *HLA-DRB1*13:02* (P = 2. 50×10-4, odds ratio = 1. 54). We then performed subgroup analyses of GWAS, dividing subjects with and without *HLA-DRB1*1302*. Several SNPs in *TMEM132D* gene showed suggestive or marginal associations with PD in the subgroup without *HLA-DRB1*1302*. Our study provides initial evidence that *HLA-DRB1*13:02* and genes involved in immune-related pathways are associated with PD. Moreover, the result of subgroup analyses suggests that an immunopathogenesis is more likely to play a role in PD patients with *HLA-DRB1*13:02*. To the contrary, *TMEM132D* might affect disease states in subjects without *HLA-DRB1*13:02*. Further studies are needed to confirm these results and clarify the underlying mechanisms causing PD.

1076W

Genome-Wide Association Study Implicates Protein Kinase N2 as a Risk Factor in Persistent Drug Abuse in Intravenous Drug Users. *SC. Wang[1], G. Kirk[2], S. Mehta[2], B. Maher[1].* 1) Johns Hopkins Bloomberg School of Public Health, Department of Mental Health; 2) Johns Hopkins Bloomberg School of Public Health, Department of Epidemiology.

Background: We performed a genome-wide association study (GWAS) to identify risk variants for continuously defined intravenous drug use in a prospective cohort ascertained for drug injection. **Methods:**The sample comprised 1228 individuals from the AIDS Linked to the Intravenous Experience (ALIVE) cohort with available genome-wide SNP data (Affymetrix 6. 0). Using the semi-annual report from the ALIVE participants over thirteen years, we used latent class growth models to divide the population into two groups: persistent drug use and cessation. The median number of follow up visit is nineteen. Prior to GWAS, we performed several procedures for quality control. We removed individuals with low genotyping rate (less than 10%). We excluded SNPs with minor allele frequency (MAF) less than 0. 08 and with more than 10% missing genotyping rate or Hardy-Weinberg test p-value less than 0. 0001. Population stratification was accounted for by Multidimensional Scaling (MDS). After frequency and genotyping pruning, 712,405 SNPs remained. Latent class growth model was performed using Mplus. After filtering, there were 700 cases and 482 controls. Genome-Wide Association Study was performed using PLINK and Haploview. **Results:** Three associations were found with p-value < 1x10-6. Two are on chromosome 2 (rs6749634, Odds ratio [OR] =0. 53, *p*= 1. 755e-7; rs12711921, OR =0. 58, *p*= 4. 858e-7) and one is on chromosome 10 (rs11255655, OR =0. 51, *p*= 7. 367e-7). The above three loci are not located in any gene. In addition, significant associations were also found on chromosome 1 (rs7522512, OR =1. 629, *p*= 9. 737e-6; rs6684950, OR =1. 666, *p*=9. 858e-6) mapped to the protein kinase N2 (PKN2), chromosome 5 (rs7731348, OR =0. 60, *p*= 5. 264e-6) mapped to the G protein-coupled receptor 98 (GPR98), and chromosome 13 (rs2793763, OR =0. 52, *p*=86. 223e-6) mapped to the citrate lyase beta like (CLYBL). (Figure 1)**Conclusions:** According to past research, no phenotype associations with the above SNPs (rs6749634, rs12711921, and rs11255655) or the above genes (PKN2, GRP98, and CLYBL) have been reported to our knowledge; moreover, these genes' in vivo function is yet unknown. Further research is needed confirm associations between these SNPs and intravenous drug use.

1077T

Genetic risk variants converge in NMDAR interactome underlying schizophrenia. *Z. Wei[1], X. Chang[2], L. Lima[2], T. Wang[2], A. Cederquist[2], E. Ryan[2], J. Garifallou[2], M. Khan[2], Y. Liu[2], J. Li[2], P. Sleiman[2,3,4], H. Hakonarson[2,3,4].* 1) New Jersey Institute of Technology, Newark, New Jersey, 07102, USA; 2) The Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA; 3) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, 19104, USA; 4) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, 19104, USA.

Hundreds of genomic loci have been identified with the recent advances of schizophrenia researches in genome-wide association studies (GWAS) and sequencing studies. However, the functional interactions among those genes remain largely unknown. We developed a network-based approach to integrate multiple genetic risk factors, which lead to the discovery of new susceptibility genes and casual sub-networks or pathways in schizophrenia. We applied the method on two large GWA studies. One is the largest GWAS conducted by Schizophrenia Psychiatric Genome-Wide Association Study Consortium (PGC), which recruited 36,989 cases and 113,075 controls (PGC phase 2). The other one is our previous GWAS on schizophrenia and bipolar disease with 12,362 schizophrenia and 1,032 bipolar cases and 34,676 controls. We identified significantly and consistently over-represented pathways, which are highly relevant to synaptic plasticity, neural development and signaling transduction, such as long-term potentiation, neurotrophin signaling pathway and the ERBB signaling pathway. We also demonstrated that genes targeted by common SNPs are more likely to interact with genes harboring de novo mutations (DNMs) in the protein-protein interaction (PPI) network, suggesting a mutual interplay of both common and rare genetic variants in schizophrenia. We further developed an edge-based search algorithm to identify the top-ranked gene modules associated with schizophrenia risk. Besides the genome-wide significant gene GRIN2A, multiple genes involved in the NMDAR interactome exhibited strong association with schizophrenia risk, such as DLG2 (rs12294291, P = 4. 90 α10-7), GRIN2B (rs11757887, P = 8. 81 α10-7), ATP2B2 (rs9879311, P = 2. 77 α10-6) and NOS1 (rs2293052, P = 1. 24 α 10-6). Taken together, our results demonstrate that the N-methyl-D-aspartate receptor (NMDAR) is highly targeted by multiple types of genetic risk factors, suggesting that the NMDAR interactome plays a leading role in the pathology of schizophrenia.

1078F

Population-level identity-by-descent haplotype sharing provides insight into demographic history and autism spectrum disorder. *A. R. Martin[1,2], M. Bertalan[3], A. Rosengren[3], A. Buil[3], J. H. Thygesen[3], W. Thompson[3], T. H. Sparsø[3], T. Als[4], D. P. Howrigan[1,2], J. Grove[4], C. Pedersen[4], P. B. Mortensen[4], B. M. Neale[1,2], T. Werge[3], M. J. Daly[1,2], iPSYCH-SSI-Broad collaboration.* 1) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Medical and Population Genetics, Broad Institute, Cambridge, MA; 3) Mental Health Centre, Sct. Hans and Institute of Biological Psychiatry, Psychiatric Center Sct. Hans, Roskilde, Denmark; 4) National Centre for Register-Based Research, Åarhus University, Åarhus, Denmark.

Regions of the genome shared via identity-by-descent (IBD) provide insight into the fine-scale recent demographic history within and between populations, which can be inferred from genotype data at a fraction of the cost of whole-genome sequencing. Enrichments of IBD sharing among disease cases compared to controls therefore provide insight into regions of the genome containing rare risk-conferring mutations as well as the ages of their risk-conferring haplotypes. IBD mapping has been shown to be especially powerful compared to standard GWAS approaches for weakly selected regions of the genome with an accumulation of recently-arisen deleterious alleles, as is expected for example in fecundity-reducing psychiatric diseases. Here, as part of the iPSYCH program, we investigate high-resolution population structure across >34,000 genotyped Danish blood spot samples, including >7,900 ADHD and >8,900 autism cases. We leverage pairwise IBD sharing to assess effective population sizes through time on the scale of 10s-100s of generations across birth regions in Denmark. We also investigate runs of homozygosity, a special case of IBD, across subpopulations in Denmark and identify a marked increase in individuals of Middle Eastern and South Asian descent. On average, unrelated pairs of Danish individuals share 52. 4 ± 11. 8 cM of their genome in 32. 7 ± 6. 7 IBD segments ≥ 1 cM in size. We develop an IBD association framework that takes into account haplotype cluster size, haplotype length, and low complexity regions of the genome, and we apply this framework to discover rare autism risk haplotypes likely carrying recent, stronger acting variation than would be discovered via single-point GWAS analysis.

1079W

Cross-disorder analysis of 23 brain diseases and 200,000 patients using shared heritability methods reveals novel patterns in the genetic susceptibility to brain diseases. *V. Anttila[1,2,3], B. Bulik-Sullivan[1,3], H. Finucane[4,5], S. Ripke[1,2,3], R. Malik[6,7], P. Lee[8], A. Biffi[2,9,10,11], K. Kendler[12,13], J. Scharf[8], J. Smoller[8], A. Palotie[2,8,14], M. Daly[1,2], J. Rosand[9,10,11], A. Corvin[15], B. Neale[1,2,3], Brainstorm consortium.* 1) Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 4) Department of Mathematics, Massachusetts Institute of Technology, Massachusetts, USA; 5) Harvard School of Public Health, Boston, Massachusetts, USA; 6) Institute for Stroke and Dementia Research, Klinikum der Universitét München, Ludwig-Maximilians-University, Munich, Germany; 7) Munich Cluster for Systems Neurology (Synergy), Munich, Germany; 8) Psychiatric and Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston; 9) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA, USA; 10) Division of Neurocritical Care and Emergency Neurology, Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; 11) J. Philip Kistler Stroke Research Center, Massachusetts General Hospital, Boston, MA, USA; 12) Departments of Psychiatry and Human Genetics, Virginia Commonwealth University, Richmond, VA, USA; 13) Virginia Institute for Psychiatric and Behavior Genetics, Richmond, VA, USA; 14) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 15) Department of Psychiatry, Trinity College Dublin, Dublin, Ireland.

Many neurological and psychiatric diseases have considerable co-morbidity (e. g. Bipolar Disorder-Migraine, Autism-Epilepsy). In the Brainstorm project, we set out to quantify the extent of the shared genetic basis of these diseases by integrating the most recent available genome-wide association results from both neurological and psychiatric diseases, representing over 200,000 cases and 700,000 controls across 23 brain diseases. The broad and continuous nature of psychiatric phenotypic spectra has been clinically recognized for a long time. ,Recent work in psychiatric genetics has already demonstrated strong genetic correlations across mental illnesses, with the most well-known example being schizophrenia and bipolar disorder. We have recently developed an approach to estimate heritability and genetic correlation (LDSC) in order to evaluate the extent of shared genetic risk across neurological and psychiatric disease. We found that in general, psychiatric diseases as a set have considerable shared common variant genetic risk factors. Notably schizophrenia shows significant genetic correlations with most of the studied psychiatric phenotypes. In contrast, current shared genetic analyses of neurological diseases show greater specificity. Notably, a few diseases - Parkinson's, Alzheimer's and multiple sclerosis - show little or no overlap with other neurological phenotypes. Furthermore, neurological and psychiatric phenotypes tend to show little common variant genetic overlap. However, significant positive genetic correlation is observed between migraine and attention deficit hyperactivity disorder as well as depression. In this study, we set out to integrate the best available neurological and psychiatric genome-wide association results, and to use them to explore the patterns surrounding the common genetic risk factors for psychiatric diseases, to suggest directions for future cross-disorder studies. We report a number of novel, significant genetic overlaps both within and between classical disease categories suggesting new avenues for future studies in brain diseases.

1080T

Genome-wide parametric multipoint linkage analyses of extended Alzheimer's disease (AD) families identifies a Chromosome 1 risk locus for familial late-onset AD. *J. Jaworski[1], B. Kunkle[1], S. Barral[2, 3, 4], B. Vardarajan[2, 3, 5], G. Beecham[1], E. Martin[1], L. Cantwell[6], A. Partch[6], T. Bird[7, 8], W. Raskind[8, 9], A. Destafano[10], R. Carney[1, 11], M. Cuccaro[1, 12], J. Vance[1, 12], L. Farrer[10, 13], A. Goate[14], T. Foroud[15], R. Mayeux[2, 3, 4, 16], G. Schellenberg[7], J. Haines[17], M. Pericak-Vance[1, 13].* 1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 2) The Taub Institute of Research on Alzheimer's Disease, Columbia University, New York, NY, USA; 3) The Gertrude H. Sergievsky Center, Columbia University, New York, NY, USA; 4) The Department of Neurology, College of Physicians and Surgeons, Columbia University, New York, NY, USA; 5) Department of Psychiatry and Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, MO, USA; 6) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 7) Department of Neurology, University of Washington, Seattle, WA, USA; 8) Department of Medicine, University of Washington, Seattle, WA, USA; 9) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, USA; 10) Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA; 11) Department of Psychiatry and Behavioral Sciences, Miller School of Medicine, University of Miami, Miami, FL, USA; 12) Dr. John T. Macdonald Foundation Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA; 13) Departments of Medicine (Biomedical Genetics), Neurology, Ophthalmology, and Epidemiology, Boston University Schools of Medicine and Public Health, MA, USA; 14) Department of Psychiatry and Hope Center for Neurological Disorders, Washington University School of Medicine, St. Louis, MO, USA; 15) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA; 16) The Department of Epidemiology, School of Public Health, Columbia University, New York, NY, USA; 17) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH, USA.

ABSTRACT

Background. While more than two dozen loci that contribute to late-onset Alzheimer disease (LOAD) have been identified, few high penetrance mutations (e. g. , APP mutations) that explain risk in families heavily burdened with LOAD have been found. Linkage analyses in large multiplex pedigrees heavily affected with LOAD could point to genomic regions harboring these rare risk variants. Thus, we performed genome-wide parametric multipoint linkage analyses on 41 Non-Hispanic White (NHW) families exhibiting a pattern of dominantly inherited LOAD, free of known AD mutations and not clustering for the APOEe4 allele. Methods. Genome-wide genotyping of 385 individuals from these 41 families was used to perform a parametric multipoint affecteds-only analysis in MERLIN. The 41 NHW families selected met three criteria: 1) exhibiting dominant inheritance of LOAD; 2) no mutations at known familial AD loci; and 3) low occurrence of the APOEϵ4 allele. All were among families who had available genome-wide genotyping of common variants on either the Illumina HumanHap 550 or 1M arrays.

Results. A locus at 1p13. 2 produced a genome-wide significant linkage HLOD score of 3. 33 (significance defined by *LOD=>3. 3). This region encompasses approximately 1. 23 Mb and contains 23 genes or gene features (e. g. miRNA). Several of these are interesting functional AD candidates: WNT2B is differentially expressed in AD brains, MOV10 is a protein integral to synaptic strength and plasticity, RHOC's expression is down-regulated by beta-amyloid, leading to decreased migration of neural stem/progenitor cells, and LRIG-2 is a BACE1 substrate. Suggestive linkage (defined as *LOD>=2. 2) was observed at 2p22. 1 (HLOD=2. 42) and we confirmed our previously reported non-parametric multipoint result at 14q32. 2 (HLOD=4. 36). Regions previously identified in AD linkage scans (PSEN1, PSEN2, APP, MAPT, APOE) showed no evidence of linkage in these analyses. Conclusions. Parametric multipoint linkage analysis of 41 multiplex NHW families identified several loci of significant interest in an initial analysis of LOAD families undergoing whole-genome sequencing. Analyses of these prioritized regions using WGS data may help determine the causative variants for these loci.

1081F

Identification of chromosomal regions interacting with susceptibility loci for Alzheimer's disease. *M. Kikuchi[1], N. Hara[2], M. Hasegawa[3], A. Miyashita[2], R. Kuwano[4], T. Ikeuchi[2], A. Nakaya[1].* 1) Department of Genome Informatics, Graduate School of Medicine, Osaka University; 2) Department of Molecular Genetics, Brain Research Institute, Niigata University; 3) Division of Molecular Biology, Research Institute for Biomedical Sciences (RIBS), Tokyo University of Science; 4) Asahigawaso Research Institute.

The most common cause of dementia is late-onset Alzheimer's disease (AD), which occurs in individuals aged >65 years and leads to neuronal death. The pathogenesis of AD is strongly correlated with genetic factors. To identify single nucleotide polymorphisms (SNPs) associated with AD, genome-wide association studies (GWAS) have been performed across some cohorts and ethnic groups. At present, 334 susceptibility loci for AD are registered in the GWAS catalog database. However, 292 (87. 4%) of these loci are located within introns and intergenic regions. The contribution of AD-associated SNPs to the onset of AD remains unknown. In this study, we explored whether AD-associated SNPs are in functional non-coding regions such as enhancers and silencers. Given that enhancers and silencers participate in long-range interactions via chromatin loops, we determined these interactions for AD-associated SNPs in candidate regions of enhancers/silencers and searched for genes in the neighborhood of AD-associated SNPs. We collected 334 AD-associated SNPs from the GWAS catalog database. We explored AD-associated SNPs in regulatory elements using RegulomeDB, a database that annotates SNPs with regulatory elements (eQTL, transcription factor binding motifs, DNase-peaks, etc.). To identify long-range chromatin interactions for AD-associated SNPs, we performed tethered conformation capture (TCC) in the human neuroblastoma SK-N-SH and human astrocytoma U-251 MG cell lines. We found that 107 (36. 6%) out of 292 AD-associated SNPs lying in non-coding regions were related to one or more regulatory elements, suggesting that some AD-associated SNPs work as enhancers/silencers and influence the transcription of distant genes. For instance, eight SNPs in the *APOE* gene cluster, which includes *APOE*, a strong risk factor for AD, were localized in transcription factor binding regions. Furthermore, analysis of the chromatin structure by TCC revealed that the *APOE* gene cluster conformationally interacted with some genes involved in AD pathology. The chromatin interactions of the *APOE* gene cluster were observed only in the SK-N-SH cell line. Taken together, our study suggests that some AD-associated SNPs affect the transcriptional regulation of genes at distinct genomic regions through chromatin higher-order structure.

1082W

A variant in Sarcolemma Associated Protein (*SLMAP)* achieves genome-wide significance in African Americans with Alzheimer's disease using informed conditioning on clinical covariates. *J. Mez[1], J. Chung[2], J. Kriegel[1], K. L. Lunetta[4], J. Haines[12], R. P. Mayeux[9, 10, 11], M. A. Pericak-Vance[7, 8], G. D. Schellenberg[6], L. A. Farrer[1,2,3,4,5], Alzheimer's Disease Genetic Consortium.* 1) Neurology, Boston University School of Medicine, Boston, MA; 2) Medicine, Boston University School of Medicine, Boston, MA; 3) Ophthalmology, Boston University School of Medicine, Boston, MA; 4) Biostatistics, Boston University School of Public Health, Boston, MA; 5) Epidemiology, Boston University School of Public Health, Boston, MA; 6) Pathology and Laboratory Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 7) The John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 8) Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 9) Taub Institute on Alzheimer's Disease and the Aging Brain, Columbia University, New York, NY; 10) Gertrude H. Sergievsky Center, Columbia University, New York, NY; 11) Neurology, Columbia University, New York, NY; 12) Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.

Late onset Alzheimer's disease (LOAD [MIM104300]) risk is influenced by multiple known genetic, clinical and environmental factors. Including these factors as covariates in traditional genetic analyses of subjects ascertained in a case-control design may decrease power to detect an association. However, an informed conditioning approach, based on a liability threshold model with parameters informed by external prevalence data, can be used to extract risk information from known genetic and non-genetic factors without a reduction in power. A previous genome wide association study (GWAS) by the Alzheimer's Disease Genetic Consortium (ADGC) found that, in addition to the *APOEϵ4* allele, a variant in the *ABCA7* gene (rs115550680) was significantly associated with LOAD at the genome-wide level in African Americans (AAs). In the current study, we conducted a GWAS in AAs employing informed conditioning that included these genetic factors and clinical and environmental factors associated with LOAD. Subjects included 1,910 well-characterized African American AD cases and 3,829 cognitively normal AA controls from 9 ADGC datasets. First, we used LTSOFT software to derive a posterior liability conditioned on age, sex, diabetes status, smoking history, education level and AD status. Next, we evaluated association of the posterior liability with a genome-wide set of 6 million imputed markers using linear regression that included principal components of population substructure and *APOEϵ4* and rs115550680. We meta-analyzed results across datasets. We obtained genome-wide significant evidence of association (p=2. 75x10-8) between the posterior liability and rs115317194 (minor allele frequency=0. 07), a single nucleotide polymorphism (SNP) in the Sarcolemma Associated Protein (*SLMAP*) gene that is unique to African ancestry. *SLMAP* is expressed in brain and its protein product binds the protein phosphatase 2A-containing striatin-interacting phosphatase and kinase complex that regulates tau phosphorylation. In the ADGC's previous African American GWAS of AD that used a traditional case-control design, the same *SLMAP* SNP reached a p-value of 3. 39x10-5 which is three orders of magnitude less significant than the one yielded for the posterior liability model. Sensitivity analyses revealed that adding the clinical and environmental factors accounted for most of the change in the p-value. The current study suggests an informed conditioning approach can improve association signals in GWAS.

1083T

Novel risk variants and the genetic architecture of amyotrophic lateral sclerosis. *W. van Rheenen[1], A. Shatunov[2], A. Al-Chalabi[2], L. H. van den Berg[1], J. H. Veldink[1], Project MinE consortium.* 1) 1. Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands; 2) 2. Department of Basic and Clinical Neuroscience, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom.

Amyotrophic lateral sclerosis (ALS) is a fatal neurodegenerative disease affecting 1 in 400 people at a median age at onset of 65 years. Genome-wide association studies have identified several risk loci in ALS. The majority of ALS heritability, however, remains unexplained. Considering our incomplete understanding of ALS, we need to expand our knowledge of the genetic basis of this disease. Furthermore, elucidating the genetic architecture of ALS is essential to design effective future genetic studies in ALS. We genotyped new cases and controls and collected existing genotype data of all previously published GWAS in ALS. Variants obtained from whole genome sequencing 1,246 ALS cases and 615 controls merged with the 1000 Genomes Project reference panel were imputed (1000GP). SNP associations were tested in a meta-analysis and using a linear mixed model (LMM). Predictions were made using polygenic risk scores and heritability was estimated across the allele frequency spectrum. In total 12,577 cases and 23,475 cases passed GWAS quality control. Imputing variants from our custom reference panel improved the imputation accuracy (r2 = 0. 63 and 0. 83 for 1000GP and custom panel respectively). The meta-analysis confirmed genome-wide significant ($p < 5 \times 10\text{-}8$) associations for the three known ALS loci. Furthermore, a rare coding variant was found in a novel locus that was subsequently replicated (joint meta-analysis $p = 1.8 \times 10\text{-}12$), and we observed an enrichment of rare coding variants in this gene (OR 2. 33, p = 0. 013). Using a LMM we found three more loci reaching genome-wide significance. Polygenic risk score predictions indicated a polygenic architecture of ALS ($R2 = 0.7\%$, $p = 2.3 \times 10\text{-}16$). Finally, we estimated the SNP-based heritability at 8. 5% (SE 0. 5%) with an important role for less common variants (1-10%). This is supported by the observation that less common to rare variants exhibit an overall risk increasing effect on ALS suggesting purifying selection against ALS risk alleles while ALS patients are perfectly fit to produce offspring. Identifying 7 genome-wide significant loci including a new ALS gene this study substantially increases our knowledge of the genetic basis of ALS. Furthermore, it highlights a polygenic architecture where less common to rare variants confer to the overall risk of this disease. These results emphasize the need for large-scale screens for rare variants in ALS and searching the cause of selection against ALS risk alleles.

1084F

Variants in *CNTN5* Associated with ADHD Susceptibility: A Meta-Analysis of Two Pediatric Cohorts. *J. J. Connolly[1], B. Almoguera[1], F. Mentch[1], L. Vazquez[1], D. J. Abrams[1], P. Sleiman[1,2], T. Lingren[3], B. Namjou[4], T. Froehlich[5], J. B. Harley[4], H. Hakonarson[1,2], The Electronic Medical Records and Genomics (eMERGE) Network.* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; 3) Department of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA; 4) Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center and US Department of Veterans Affairs Medical Center, Cincinnati, OH, USA; 5) Department of Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA.

Attention deficit hyperactivity disorder (ADHD) is the most common neurobiological disorder in children, with a prevalence of ~6-7% that has remained stable for decades. We present results of a meta-analysis of two genome-wide association studies (GWASs) involving European American (EA) cohorts from The Children's Hospital of Philadelphia (CHOP – 762/4,907 cases/controls) and Cincinnati Children's Hospital Medical Center (CCHMC – 222/402 cases/controls). All cases and controls were defined by an electronic health record (EHR)-based algorithm, which mines health EHRs for pertinent diagnostic (i. e. ICD9 codes) and medication information. The algorithm was run and validated on the CHOP cohort of >80,000 patients, and also externally at the CCHMC, n=>6,000. Chart review of 50 cases and 50 controls at CHOP produced positive and negative predictive values of 96% and 96%. Respective values for CCHMC were 89% and 95%. All samples were genotyped on Illumina on HumanHap550 or Human610-Quad and imputation of non-observed genotypes was carried out using SHAPEIT and IMPUTE2. Association analysis was conducted using SNPtest and 10 principal components were generated using Eigenstrat 3. 0 to adjust for population stratification. Metal was used for meta-analysis of the two cohorts. Three single nucleotide polymorphisms (SNPs) at the contactin 5 (*CNTN5*) gene were nominally significant in each cohort, but statistically significant in the combined analysis (P=1. 34 x 10-8, P <2. 39 x 10-8, and 3. 88 x 10-8 respectively). *CNTN5* represents a solid ADHD candidate gene as it encodes a neuronal membrane protein, which functions as a cell adhesion molecule and mediates cell surface interactions during nervous system development. While a CNV study previously associated *CNTN5* with ADHD, this is the first time that a single nucleotide variant is observed to be associated with ADHD at a genome-wide significant level.

1085W

Identifying Genetic Modifiers of an Inherited Peripheral Neuropathy in a GWAS Design. *F. Tao[1], G. Beecham[1], N. Vasudeva[1], L. Abreu[1], R. Schüle[1,2], S. Blanton[1], D. Pareyson[3], M. Reilly[4], M. Shy[5], S. Züchner[1], Inherited Neuropathy Consortium.* 1) Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Department of Neurology, University of Tübingen, Tübingen, Germany; 3) Department of Clinical Neurosciences, C. Besta Neurological Institute, Milan, Italy; 4) MRC Centre for Neuromuscular Diseases, UCL Institute of Neurology, London, UK; 5) Department of Neurology, University of Iowa, Iowa City, IA, USA.

The most common subtype of inherited peripheral neuropathies is Charcot-Marie-Tooth disease type 1A (CMT1A), caused by a uniform 1. 5Mb duplication including PMP22 gene on chromosome 17. Although sharing the same genetic cause, CMT1A patients show great differences in their symptom onset and disease severity, suggesting that other factors such as genetic modifiers may play a role in modulating the phenotypic presentation. This study uses genome-wide association tests as the primary study design to identify genetic variants with modifying effects in CMT1A. We genotyped 715K SNPs in 919 CMT1A patients and performed a genome-wide association study (GWAS) to identify genetic variants that associate with the following phenotypic outcomes: difficulty in walking, difficulty in balance, decreased sensation and foot dorsiflexion. The genotypes were also utilized to confirm PMP22 duplication in each CMT1A patient. Two different association tests were performed: genome-wide association analyses with family data (GWAF) for SNP-based association test, and sequence kernel association test (SKAT) for gene-based association test. Both SNP-based GWAF and gene-based SKAT generated multiple genetic variants of interest associated with the four clinical outcomes. However, none of the candidate SNPs and genes reached genome-wide statistical significance after Bonferroni correction to adjust for multiple testing. The top SNPs identified by GWAF are: rs7196367 for difficulty in walking (raw p-value = 6. 7×10-7), rs2406176 for difficulty in balance (raw p-value = 1. 3×10-6), rs4346964 for decreased sensation (raw p-value = 3. 3×10-6), rs11212437 for foot dorsiflexion (raw p-value = 3. 7×10-6). The top genes identified by SKAT are: LOC441025 for difficulty in walking (raw p-value = 1. 4×10-4), NR5A1 for difficulty in balance (raw p-value = 3. 9×10-5), DIS3L2 for decreased sensation (raw p-value = 2. 8×10-4), PLCXD3 for foot dorsiflexion (raw p-value = 1. 9×10-5). This study, thus far, did not show a genome-wide statistically significant association of genetic modifiers with the phenotypic variability in CMT1A patients. We are currently conducting pathway-based analysis from the top 5% genes of interest, and are also complementing these studies with an extreme phenotype approach using whole genome sequencing (WGS). Meanwhile, the Inherited Neuropathy Consortium is working on expanding the sample size for GWAS to achieve higher statistical power in detecting modifier variants.

1086T

Investigating the Genetic Architecture of Major Depression and Body Mass Index in 10649 Han Chinese Women using Sparse Whole Genome Sequencing and Molecular Signatures of Stress. *R. Peterson[1], T. Bigdeli[1], N. Cai[2], A. Edwards[1], W. Kretzschmar[2], F. Yang[3], J. Marchini[2], J. Flint[2], K. Kendler[1], S. Bacanu[1], B. Webb[1], CONVERGE Consortium.* 1) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, PO Box 980126, Richmond VA, 23298, USA; 2) Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, Oxfordshire, UK; 3) Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine, No. 600 South Wanping Road, Shanghai, P. R. China.

Obesity is a major public health concern, and is often comorbid with psychiatric conditions including major depression (MD). Notably, several symptoms of MD are related to energy balance, including changes in appetite, weight, and sleep. It is conceivable that stressful life events (SLE) and dysregulation of biological systems of stress may underlie a shared liability between mental and physical health. However, despite numerous associations between obesity and various mental health outcomes there has been limited research on shared genetic and environmental liability. We sought to investigate the relationship between obesity and MD by (1) testing genetic variants previously identified as influencing body mass index (BMI) for association with BMI, MD, and depression subtypes, (2) constructing genetic risk scores (GRS) and test for cross-trait association, and (3) evaluating the moderating effects of SLEs and molecular signatures of stress. We used sparse whole-genome sequencing data from the CONVERGE (**C**hina, **O**xford a**n**d **V**CU **E**xperimental **R**esearch on **G**enetic **E**pidemiology) study of 6,000 Han Chinese women with recurrent MD and 6,000 screened controls. In aggregate, common SNPs accounted for 15. 5% of the variance in BMI (p=8. 0x10-7), 22. 6% in MD diagnosis (p<1. 0x10-16), and 16. 1% in atypical depression liability (p=0. 002). Associations with individual variants previously implicated in BMI were not robust (p*FTO*=0. 005). However in aggregate, a GRS constructed from 92 BMI-associated genetic variants was significantly associated with BMI, accounting for ~0. 75% of the variance (p<1. 0x10-16). The GRS was negatively associated with MD (effect=-0. 72, p=0. 014) but not atypical depression. Phenotypic analyses indicated significant environmental main effects of childhood sexual assault (OR=2. 9, p=2. 6x10-19), childhood physical abuse (OR=4. 6, p=4. 2x10-23), and serious childhood neglect (OR=4. 8, p=2. 5x10-51) on MD. We also observed a significant association between MD and both mtDNA levels (OR=1. 33, p=9. 0x10-42) and telomere length (OR=0. 85, p=2. 8x10-14). Analysis of SLEs and molecular signatures of stress will be applied to models of comorbidity. Initial results were suggestive of partially shared genetic risk between MD and BMI. Furthermore, results suggest genetic variation associated with BMI may only partially overlap between European and Han Chinese populations, highlighting the importance of studying genetic contributions to complex traits in diverse populations.

1087F

Sparse whole genome sequencing identifies two loci for major depressive disorder. *T. B. Bigdeli[1], N. Cai[2], W. Kretzschmar[2], Y. Li[2], B. Riley[1], Q. Li[3], X. Gan[4], B. T. Webb[1], S. A. Bacanu[1], R. E. Peterson[1], Y. Chen[5], G. Fu[1], F. Yang[6], E. Cong[6], J. Marchini[2], H. Yang[3], J. Wang[3], S. Shi[6,7], R. Mott[2], Q. Xu[8], J. Wang[9,10], K. S. Kendler[1], J. Flint[2], CONVERGE consortium[1-10].* 1) Virginia Institute for Psychiatric and Behavioral Genetics, Richmond, VA; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford, Oxfordshire, OX3 7BN, United Kingdom; 3) BGI-Shenzhen, Floor 9 Complex Building, Beishan Industrial Zone, YantianDistrict, Shenzhen, Guangdong, 518083, China; 4) Department of Comparative Developmental Genetics, Max Planck Institute for Plant Breeding Research, Carl-von-Linne-Weg 10, Cologne, 50829, Germany; 5) CTSU, Richard Doll Building, University of Oxford, Old Road Campus, Oxford, Oxfordshire, OX3 7LF, United Kingdom; 6) Shanghai Jiao Tong University School of Medicine, Shanghai Mental Health Centre, No. 600 Wan Ping Nan Road, Shanghai, 200030, China; 7) Fudan University affiliated Huashan Hospital, No. 12 Wulumuqi Zhong Road, Shanghai, 200040, China; 8) National Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences & Neuroscience Center, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, 10005, China; 9) Department of Biology, University of Copenhagen, Ole Maal Oes Vej 5, Copenhagen, 2200, Denmark; 10) Macau University of Science and Technology, Avenida Wai long, Taipa, Macau 999078, China, Taipa, Macau, 999078, China.

Background: Major depressive disorder (MDD) is one of the commonest psychiatric illnesses and a leading cause of morbidity worldwide. Although modestly heritable (30-40%), MDD is a genetically complex and etiologically heterogeneous trait, complicating efforts to identify variants conferring risk or protection. For these reasons, we investigated the genetic basis of MDD in subjects for whom known sources of phenotypic and genetic heterogeneity were minimized and known risk factors documented. **Approach**: The CONVERGE (China, Oxford and Virginia Commonwealth University Experimental Research on Genetic Epidemiology) consortium recruited 11,670 Han Chinese women through a collaboration involving 58 hospitals in 45 cities and 23 provinces of China. Whole genome sequence (WGS) was acquired to a mean depth of 1. 7X per individual from which 32,781,340 SNP sites were identified. Comparison of genotypes from the low coverage sequencing to genotypes called from 10X sequencing data indicated 98. 1% concordance (*N*=9). Following stringent quality controls, we performed a genome-wide association study (GWAS) of 5,303 cases with recurrent MDD and 5,337 controls screened to exclude MDD. **Results**: We identified and replicated in an independent Han Chinese sample (*N*=6,417) two genome-wide significant loci contributing to risk of MDD on chromosome 10: one near *SIRT1* (*P* = 2. 53x10-10) and the other in an intron of *LHPP* (*P* = 6. 45x10-12). Comparison with results from the Psychiatric Genomics Consortium (PGC) mega-analysis of European studies failed to provide robust replication for our top SNPs. However, the proportion of associations in the same direction in the two studies exceeded chance expectations (*P*< 0. 001). **Conclusions**: We have identified the first robustly associated, replicated genetic associations with MDD. We attribute our success to the recruitment of relatively homogeneous cases with severe illness, and to our having conducted the largest WGS study of a single cohort to date. Our results clearly demonstrate that obtaining low sequence coverage of a large number of individuals can be an effective way to screen the genome for association signals. We discuss relevant implications for future genetic studies, as well as emergent findings from a joint PGC-CONVERGE trans-ethnic meta-analysis of MDD.

1088W

Cross-Disorder Genome-Wide Analyses and Genetic Relationship between ASD and OCD. *W. Guo[1], JF. Samuels[2], Y. Wang[2], MA. Grados[2], MA. Riddle[2], OJ. Bienvenu[2], FS. Goes[2], B. Maher[2], AE. Pulver[2], D. Valle[3], G. Nestadt[2], YY. Shugart[1].* 1) Unit of Statistical Genetics, NIMH, Bethesda, MD; 2) Johns Hopkins University School of Medicine, Department of Psychiatry and Behavioral Sciences, Baltimore, MD 21205, USA; 3) John Hopkins University School of Medicine, Institute of Human Genetics, Baltimore, MD 21205, USA.

Obsessive-compulsive disorder (OCD) and Autism spectrum disorder (ASD) are highly heritable neurodevelopmental disorders that are thought to share genetic risk factors. However, the underlying genetic determinants are largely unknown. The authors report a combined genome-wide association study (GWAS) of ASD and OCD. The OCD dataset includes 3087 individuals in 1065 families (comprising 1406 patients with OCD and 2895 individuals in total). The ASD dataset includes 6782 individuals in 2227 trios. GWAS summary statistics were examined for enrichment of functional variants associated with gene expression levels in brain regions. Polygenic score analyses were conducted to investigate the genetic relationship within and across the two disorders. The heritability on both data were examined using genome-wide common variant data by the program Genome-wide Complex Trait Analysis (GCTA). We examined genomic architectures of ASD and OCD by chromosome and MAF bin.

1089T

Identification of novel genetic variants of DSM-5 alcohol use disorder: Exome Array analyses in NESARC-III. *J. Jung, H. Zhang, RB. Goldstein, BF. Grant.* Laboratory of Epidemiology and Biometry, National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Rockville, MD.

DSM-5 Alcohol use disorder (AUD) is a complex psychiatric disorder that affects about 32% of European Americans (EA) and 22% of African Americans (AA). Moderate to severe DSM-5 AUD and DSM-4 alcohol dependence (AD) are similarly genetically influenced with heritability of 50-65% and genetic correlation between them of 95%. We analyzed data from the National Epidemiologic Survey on Alcohol and Related Conditions-III (NESARC-III) to perform a Exome Array study of alcohol use disorder with moderate to severe symptoms (AUD >=4). EA and AA in NESARC-III samples were genotyped by Affymetrix Axiom Exome Chip 319K and 100K additional customized SNPs. EA samples of 2114 cases and 2697 supernormal controls with no psychiatric disorders, and AA samples of 436 cases and 1448 supernormal controls were analyzed. A single SNP based analysis with common variants (minor allele frequency (MAF)>= 0. 01) was performed to test association with AUD using an additive model and a gene based analysis with rare variants (MAF<0. 05) was conducted using Sequence Kernel Association Test after controlling for sex, age, family income, marital status, education, and two population stratification scores. Based on a single SNP analysis, we identified three novel SNPs in *ST5* (rs200029552), *GRIN3A* (rs144427058), and *DEPDC5* (rs199749859), with p-value < 2e-7 in EA. In addition, we found that a SNP rs117916638 located in exon region of gene *EGLN2* (p-value=3. 74e-6) is significantly associated with AUD in EA samples, and we replicated it in AA samples (p-value=6. 2e-3). In a gene based analysis, we identified 3 novel genes (*FER1L6-AS1, FOLR2, H2BFM*) with p-value <2. 5e-6 of a gene level threshold in EA and *EGLN2* was detected in both EA (4. 2e-5) and AA (6. 2e-3). Both gene based approach and a SNP based approach confirmed that *ADH1B* (p=1. 3e-5) is associated in only EA. In addition to confirming previous findings on associations of *ADH1B, GRIN3A* and *FOLR2* with alcohol-related phenotypes, our preliminary data identify novel associations that open up promising path to improved understanding to the etiology of AUD.

1090F

Genome-wide Linkage on Chromosome 10q26 for a Dimensional Scale of Major Depression. *E. E. M. Knowles[1], J. W. Kent[2], D. R. McKay[1], E. Sprooten[3], S. R. Mathias[1], J. E. Curran[4], M. A. Carless[2], M. A. A. de Almeida[4], H. H. H. Goring[2], T. D. Dyer[4], R. L. Olvera[5], P. T. Fox[5], R. Duggirala[2], L. Almasy[5], J. Blangero[4], D. C. Glahn[1].* 1) Yale University, New Haven, CT; 2) Texas Biomedical Research Institute, San Antonio, Texas, USA; 3) Icahn School of Medicine at Mount Sinai, New York, NY; 4) The University of Texas, Rio Grande Valley, Edinburg, TX; 5) University of Texas Health Science Center, San Antonio, Texas.

Major depressive disorder (MDD) is a common and potentially life-threatening mood disorder. Identifying genetic markers for depression might provide reliable indicators of depression risk, which would, in turn, substantially improve detection, enabling earlier and more effective treatment. The aim of this study was to identify rare variants for depression, modeled as a continuous trait, using linkage and post-hoc association analysis. The sample comprised 1221 Mexican-American individuals from extended pedigrees. A single dimensional scale of MDD was derived using confirmatory factor analysis applied to all items from the Past Major Depressive Episode section of the Mini-International Neuropsychiatric Interview. Scores on this scale of depression were subjected to linkage analysis followed by QTL region-specific association analysis. Linkage analysis revealed a single genome-wide significant QTL (LOD = 3. 43) on 10q26. 13; the majority of the linkage signal originated from a large multiplex pedigree within the larger sample (LOD = 1. 84). QTL-specific association analysis conducted in the entire sample revealed a suggestive variant within an intron of the gene *LHPP* (rs11245316, p = 7. 8x10-04; LD-adjusted Bonferroni-corrected p = 8. 6x10-05). Family-specific association analysis within the multiplex MDD pedigree (from which the majority of the linkage signal originated) revealed a variant within an intron of CPXM2 (rs7913161, p = 2. 4x10-05) that met the peak-adjusted alpha. The signal for this variant in the entire sample was not significant (p = 0. 19), suggesting that the variant rs7913161 is likely marking a functional and rare variant present only in the multiplex pedigree. This region of the genome has previously been implicated in the etiology of MDD; the present study extends our understanding of the involvement of this region by highlighting a putative gene of interest (*LHPP*), and implicating the possible role of rare variation within the gene *CPXM2* in a newly identified multiplex MDD pedigree.

1091W

Bivariate linkage and association for recurrent major depression and right hippocampal volume. *S. R. Mathias[1], E. E. M. Knowles[1], J. W. Kent[2], D. R. Mckay[1], M. A. Carless[2], J. E. Curran[3], M. A. A de Almeida[3], H. H. H. Goring[2], T. D. Dyer[3], R. L. Olvera[4], P. T. Fox[4], L. Almasy[4], J. Blangero[3], D. Glahn[1].* 1) Yale University, New Haven, CT; 2) Texas Biomedical Research Institute, San Antonio, TX; 3) The University of Texas Rio Grande Valley, Edinburg, TX; 4) University of Texas Health Science Center, San Antonio, TX.

Previous work has shown that the hippocampus is smaller in the brains of individuals suffering from major depressive disorder (MDD) than in those of healthy controls. Moreover, right hippocampal volume specifically has been found to predict the probability of subsequent depressive episodes. The present study explored the utility of right hippocampal volume as an endophenotype of recurrent MDD (rMDD). We observed a significant genetic correlation between the two traits in a large sample of Mexican American individuals from extended pedigrees (ρ_g = -0. 34, s. e. = 0. 14,p = 0. 013). A bivariate linkage scan revealed a significant pleiotropic quantitative trait locus on chromosome 18p11. 31-32 (LOD score = 3. 61). Bivariate association analysis conducted under the linkage peak revealed a variant within an intron of the gene *SMCHD1* (rs12455524) met the level of LD-adjusted peak-wide significance (p = 7. 4 x 10-5). Univariate association analyses of each phenotype separately revealed that the same variant was significant for right hippocampal volume alone (p= 2. 6 x 10-5), and also revealed a suggestively significant variant within the gene *DLGAP1* for rMDD alone (p = 0. 0001). The results implicate right-hemisphere hippocampal volume as a possible endophenotype of rMDD, and in so doing highlight a potential gene of interest for rMDD risk.

1092T

Genome-Wide Association Study of Smoking Quantity in American Populations: Identification of Novel Risk Loci in Both African- and European-Americans. *J. L. Montalvo-Ortiz[1], H. R. Kranzler[3], R. Sherva[4], L. Almasy[5], A. I. Herman[1,2], R. Koesterer[4], H. Zhao[6], L. A. Farrer[4,7], J. Gelernter[1,2].* 1) Department of Psychiatry, Yale University School of Medicine, New Haven, CT; 2) Departments of Genetics and Neurobiology, Yale University School of Medicine, New Haven, CT; 3) Department of Psychiatry, University of Pennsylvania, Philadelphia, PA; 4) Department of Medicine (Biomedical Genetics), Boston University, Boston, MA; 5) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 6) Departments Biostatistics and Genetics, Yale School of Public Health, New Haven, CT; 7) Departments of Neurology, Ophthalmology, Genetics & Genomics, Epidemiology and Biostatistics, Boston University School of Medicine, Boston, MA.

BACKGROUND: We report a genome-wide association study (GWAS) of the trait of smoking quantity, defined as the number of cigarettes smoked per day in the lifetime in European-American (EA) and African-American (AA) populations. METHODS: Our GWAS sample included the Yale-UPenn sample (2114 EA and 2602 AA) used in our previous GWAS studies, augmented by additional subjects ascertained identically. We conducted the GWAS analysis by considering smoking quantity as an ordinal trait, assessed by the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA), which is a polydiagnostic assessment for psychiatric traits. RESULTS: In AAs, *IRX3* SNP rs16953233 located in chromosome 16 was genome-wide significant (GWS) (p = 2. 84 × 10-8). SNP rs61040077 mapped to chromosome 6 region near *RGS17* was also nominally associated with smoking quantity (p = 3. 76 × 10-7). Although it is not genomewide significant it may be interesting since we have previously shown *RGS17* to be associated with risk for substance dependence diagnoses including alcohol, cocaine, opioid and marijuana dependence and has been also shown to be associated with smoking behaviors. In EAs, three SNPs mapping to *IQCF3* were the most significant results obtained (minimal p = 4. 65 × 10-7). Pathway analysis identified association with calmodulin and calcium signaling in both populations, which has been previously found to be involved in the effects of nicotine dependence and other substance use disorders. CONCLUSIONS: These risk variants for smoking behavior provide novel insights into potential therapeutic and prevention interventions.

1093F

Genetics of depression across adolescence. *H. M. Sallis[1,2], L. Paternoster[1], J. Evans[2], G. Davey Smith[1].* 1) MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, United Kingdom; 2) Centre for Academic Mental Health, School of Social and Community Medicine, University of Bristol, Bristol, UK.

Despite many attempts to understand the genetic architecture of depression, little progress has been made. The heritability of depression in adults has previously been estimated at around 0. 37 (Wray et al, 2013), while the Psychiatric Genomics Consortium used SNP-based methods to estimate a heritability (h2SNP) of 0. 21 (Lee et al, 2013). The majority of these studies have been carried out in adults and have ignored the potential influence of development. However a recent publication by Nivard et al (2014) looked at heritability of depression across the life course. This study showed a substantial genetic component to depression at earlier ages, with a distinct drop in heritability around the age of puberty. However, this finding coincided with a switch from parent to self-report measures. The Avon Longitudinal Study of Parents and Children (ALSPAC) collected consistent self-report measures of depressive symptoms between ages 10 and 18 and is uniquely placed to investigate the heritability of depression across adolescence, an important developmental period involving major biological changes and a rising incidence of depression. We used genome-wide complex trait analysis (GCTA) to investigate h2SNP of depression across adolescence, in addition to bivariate GCTA to look at the genetic correlation between depressive symptoms. We investigated the role of puberty by stratifying estimates of h2SNP according to onset at ages 10 and 13. A genome-wide association study of depressive symptoms during early and late adolescence was performed to identify genetic variants associated with depression at different stages of development. We found fluctuations in h2SNP across adolescence, with a peak at age 13 (h2SNP=0. 17, se=0. 06). A genetic correlation between depressive symptoms at ages 13 and 18 of 0. 40 was found, although the statistical evidence for this was weak. h2SNP estimates differed according to timing of pubertal onset. No genetic variants achieved genome-wide significance at ages 10 or 18, however, we found a signal on chromosome 7 at age 13. We are currently attempting to replicate this finding. In conclusion we found evidence of the heritability of depression across adolescence but no evidence to support an abrupt drop in heritability with the onset of puberty. There is evidence of partial overlap of the genetic architecture of depression between mid and late adolescence. A genetic variant on chromosome 7 may contribute to the heritability of depression at 13.

1094W

Schizophrenia Polygenic Risk Score is Strongly Associated with Bipolar 1 Disorder in a Multi-ethnic Sample. *C. Schaefer[1], L. Shen[1], K. Thai[1], E. Jorgenson[1], T. Hoffmann[2], Y. Banda[2], M. Kvale[2], N. Risch[1,2].* 1) Division of Research, Kaiser Permanente Northern California Division of Research, Oakland, CA; 2) Institute for Human Genetics, UCSF, San Francisco, CA.

Growing evidence supports a large, shared polygenic contribution of common variants to bipolar disorder (BD) and schizophrenia (SCZ). To better understand shared genetic variation of SCZ with bipolar phenotypes and with BP1 in minority race/ethnicity groups, we calculated a polygenic risk score for schizophrenia (SCZ-GRS) based on 108 genetic variants significantly associated with SCZ in a recent GWAS (PGC, Nature, 2014), using a multi-ethnic sample of 5,056 BD1 cases and 65,114 controls. BD1 cases and controls are participants in the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH); cases were identified by multiple treatment episodes of BD1 and/or hospitalization with BD1 from electronic medical records (EMR). Persons with diagnoses of SCZ, schizoaffective disorder, or non-affective psychoses were excluded. The resulting sample was 68% female; average age at study entry was 47 years, and 25% were minority or mixed race-ethnicity. Controls were screened using EMR data to exclude those with Axis 1 disorders. Cases and controls were genotyped on the Affymetrix Axiom system imputed to 1000 Genome Project (March 2012 release). The SCZ-GRS was calculated as the weighted sum of the number of risk alleles; one-sided p-values were used due to the strong hypothesis of association of SCZ-GRS with BD. In analyses of non-Hispanic whites (3726 cases; 54620 controls), SCZ-GRS was significantly higher in cases (mean = 10. 44; s. d. = 0. 54) than controls (mean = 10. 36; s. d. = 0. 55; b=0. 270, p = 4. 31E-17). BD1 cases with psychotic symptoms (N=2418) had higher SCZ-GRS than those without (N=1307)(b=0. 204, p = 0. 0013). The SCZ-GRS was significantly associated with BD1 in each stratum of age at onset (<= 20, 21-39, >=40 years). The SCZ-GRS was associated with BD1 in East Asians (cases 298; controls 5777;b=0. 255, p = 0. 053) and Latinos (cases =732; controls = 4770;b=0. 141, p = 0. 077), but not in African Americans (cases= 300; controls=2167)(b=0. 121, p = 0. 373). These results show that genetic variation is broadly shared by BD and SCZ, including among Latinos and Asians, and specific BD1 phenotypes, such as psychotic symptoms, show increased sharing. Further examination of BD1 phenotypes, as well as gene pathways or networks may point to additional specificity in the shared variation of BD and SCZ.

1095T

Genetics of Alcohol Drinking Behaviors for Koreans: a Population with one third of Slow Metabolizers. *S. Son[1], S. Yang[1], M. Kim[2], J. Sung[1].* 1) Department of Epidemiology, Graduate School of Public Health, Seoul National University, Seoul, South Korea; 2) Shattuck-St. Mary's School, Faribault, MN.

It is well documented that individuals with Lys mutation (Glu504Lys) (=rs671 G>A) of *Aldehyde Dehydrogenase 2 (ALDH2)* gene are so-called "slow metabolizers" and less prone to alcohol abuse. In Korea and East Asian countries, about one third of population has one or two copies of the "slow" allele. Many of "slow metabolizers", however, drink as much and as frequently in Korea, and it remains unclear whether other differences in genetic constitutions explain drinking behaviors between slow and fast metabolizers. To identify genetic architecture of alcohol drinking behaviors, we performed genome wide association studies for a range of alcohol-related traits in general Korean population. We performed GWAs for 8,852 subjects from two independent cohort (3,479 from family-based, 5,373 from population-based cohort). 3,479 subjects from family-based cohort were measured for details of drinking habits and genetic markers with 3. 7 million SNP markers. After excluding 735 abstinent participants for religious reasons, we performed GWAs for self-reported facial flushing reaction with alcohol use, three domains (hazardous drinking, harmful use, and alcohol dependence) of the Alcohol Use Disorders Identification Test (AUDIT), and frequency and amount of drinking as phenotypes. All the analyses were repeated by stratifying the rs671 alleles, and SNPs interacting with the rs671 alleles on alcohol consumption behaviors were examined genome-widely. Only *ALDH2* rs671 explained facial flushing (P=3. 1×10-68). For various aspects of alcohol drinking habits, some other SNPs within the LD block of rs671 showed stronger associations, but adjusting for rs671 attenuated the significance (from P=10-68 to P=10-6). When we performed stratified analysis by fast/slow metabolizers, novel loci on the 3p14. 1 (P=6. 4×10-9) and 10p14 region(P=1. 4×10-7) showed associations with hazardous and harmful drinking domains for "fast metabolizers"; also a novel locus on the 20p12. 1 region (P=2. 2×10-7) was associated with dependence domain of AUDIT for "slow metabolizers". The same 20p12. 1 region also showed probable interaction with rs671 (P=2. 0×10-6) for dependence domain. *ALDH2* gene has a large LD block (81 kb), and it is not clear whether some other loci near rs671 also participate in the susceptibility of certain drinking habits, particularly binge drinking habits. Our findings suggest genetic susceptibility for problematic drinking behaviors might differ between fast and slow metabolizers.

1096F

Genome-wide meta-analyses identify copy number variation associated with alcohol dependence in African and European populations. *A. Sulovari[1,2], D. Li[1,3,4].* 1) Microbiology and Molecular Genetics, University of Vermont, Burlington, VT; 2) Cell, Molecular and Biomedical Sciences Graduate Program, University of Vermont, Burlington, VT; 3) Department of Computer Science, University of Vermont, Burlington, VT; 4) Neuroscience, Behavior and Health Initiative, University of Vermont, Burlington, VT.

Background: Alcohol dependence is one of the most common causes of morbidity, mortality and disability among individuals with psychiatric disorders. We report genome-wide meta-analyses between copy-number variation (CNV) and alcohol dependence (AD) in samples of European and African descent. **Methods**: We analyzed CNVs in a total of 7,467 unrelated samples from three different studies, including OZALC (Genome-wide association study of alcohol use and alcohol use disorder in Australian Twin-Families), CIDR (Collaborative Study on the Genetics of Alcoholism Case Control Study) and SAGE (Study of Addiction: Genetics and Environment). CNV detection was carried out using PennCNV and QuantiSNP. Of all samples 5,778 were of European ancestry and 1,585 of African descent. For each of the two populations we conducted a meta-analysis under the model of fixed effects (Cochran-Mantel-Haenszel, CMH) for association with AD. **Results:**In samples of European descent, CNVs in region 7p22. 3 were significantly associated with AD diagnosis (CMH OR = 3. 7,*P* =6. 82 × 10-5). In samples of African descent, CNVs in regions 14q32. 33 and 11p11. 2 were associated with AD (CMH OR = 2. 24,*P* =7. 51 × 10-6and CMH OR = 2. 36,*P* =6. 33 × 10-5, respectively). **Conclusions:** Our genome-wide meta-analyses identified three CNV risk loci that may contribute to the development of alcohol dependence in the European and African populations.

1097W

Cooperativeness Character Domain of Temperament and Character Inventory (TCI) is Associated with 15q26. 3 region in General Korean Population. *S. Yang[1], J. H. Kim[3], J. Sung[1,2].* 1) Complex Disease & Genome Epidemiology Branch, Department of Epidemiology, School of Public Health, Seoul National University, Korea; 2) Institute of Environment and Health, Seoul National University, Korea; 3) Department of Psychiatry, Samsung Medical Center, Sungkyunkwan University School of Medicine, Korea.

An individual's personality affects health by predisposing the person to a higher chance of specific health behaviors. Despite moderate heritability of 0. 3-0. 6 for most personality domains, specific genetic variants that explain personality have been elusive in studies with Caucasian samples. We conducted a genome-wide search for both TCI domains and multivariable TCI-Five-Factor Model (FFM, for sub-samples) in a population cohort of Korea, where cultural environments are different from Western populations. This study includes a total of 3342 individuals (1419 men, 691 families, 557 monozygotic twins) from the Healthy twin study of Korea with detail epidemiologic, clinical information and TCI measures; 1169 (476 men) also fulfilled the FFM. Two Platforms (Affymetrix Genome-Wide Human array 6. 0, Illumina Infinium Humancore Beadchips) were used for genotyping and the markers were imputed using 1kG Asians (n=280). For statistical analysis, we conducted a family-based univariate association test using mixed-effect variance component approach for TCI scales (MERLIN), and multiple family-based quasi-likelihood score test (MFQLS) of linear mixed model, which incorporates best linear unbiased predictor as an offset, for multivariable analysis. Univariate TCI domain analysis showed cooperativeness (CO) domain were associated with ADAMTS17 gene (on 15q26. 3, p= 9. 0e-8), but other domains did not reach significance level. When we performed a principal component analysis for both TCI and FFM domains, several latent layers of personality traits were evident. In multivariate analysis, persistence (PE-TCI) and openness (OP-FFM) (r=0. 32, RP5-897D18. 1, 20q13. 31, p= 1. 9e-7), PE-TCI and extraversion-FFM (r=0. 37, GALNT18, 11p15. 3, p= 2. 0e-7), self-transcendence-TCI and OP-FFM (r=0. 57, ADGRG7, 3q12. 2, p= 2. 6e-7) showed meaningful genetic associations. Our findings suggest that 1) some personality domains, particularly character traits might have different biological background in Koreans, 2) not a single measure but a multiple measures of personality traits might better capture the genetic architecture of personality traits. Because genetic studies on personality in Asian populations and studies which measured both TCI and FFM are scarce, we currently await collaborations to replicate our findings. This work was supported by the Post-Genome Technology Development Program (10050164, Developing Korean Reference Genome) funded by the Ministry of Trade, Industry & Energy (Korea).

1098T

Exome-wide rare variant analyses reveal NT5DC1 gene as a new candidate gene for adult ADHD. *T. Zayats, K. Jacobsen, IMPACT Consortium, J. Haavik, S. Johansson.* K. G. Jebsen Centre for Research on Neuropsychiatric Disorders, Department of Biomedicine, University of Bergen, Bergen, Norway.

Objectives: Attention-deficit/hyperactivity disorder (ADHD) is a highly heritable childhood-onset neuropsychiatric condition that often persists into adulthood. The genetic architecture of ADHD, and particularly ADHD in adults, is largely unknown. We performed an exome-wide scan of adult ADHD using Illumina HumanExome BeadChip, which allows examination of over 250,000 common and rare variants. Methods: DNA samples were collected by the International Multicenter persistent ADHD CollaboraTion (IMpACT). Statistical analyses were divided into 3 steps: (1) gene-level analysis of rare-variants (MAF<5%), (2) single marker association tests of common variants (MAF≥1%) and (3) pathway analyses. To account for multiple testing, Bonferroni correction was applied. Results: In total, 9. 365 individuals (1. 846 cases and 7. 519 controls) were analyzed. Analysis of rare variants revealed a new, genome-wide significant candidate gene for adult ADHD: NT5DC1 gene (p=3. 17E-08). The strongest signal from common variants was observed at rs9325032 in *PPP2R2B* (p=1. 61E-05). In pathway analyses, the strongest association was observed for the termination of RNA polymerase II transcription (GO:0006369 term, p=2. 48E-05), surviving Bonferroni correction for multiple testing of GO terms. Conclusion: In this largest molecular genetic study of adult ADHD performed so far, the results point to the involvement of haloacid dehalogenase superfamily of hydrolases in the etiology of adult ADHD, providing new insights into possible mechanisms underlying this condition.

1099F

RNA-sequencing identifies differentially expressed novel gene isoforms in bipolar disorder. *N. Akula, FJ. McMahon.* Human Genetics Branch, 35 Convent Drive, National Institute of Mental Health, National Institutes of Health, Bethesda, MD.

Bipolar disorder (BD) is a complex brain disorder that affects ~2% of the population. Genome-wide association studies (GWAS) have consistently identified several common variants associated with BD, but their impact on gene expression remains largely unexplored. Recent RNA sequencing (RNA-seq) studies have revealed that many of the BD GWAS-implicated genes encode thousands of alternative transcripts in the fly brain (Brown et al. 2014), suggesting that alternative splicing may be an important mechanism underlying risk for BD. To test this hypothesis, we performed deep RNA-seq (over 276 million paired-end reads/sample) in dorsolateral prefrontal cortex tissue obtained from 6 BD cases and 6 age- and sex- matched controls (Akula et al. 2014). We also identified 29 genes from the NIH GWAS Catalogue that mapped near SNPs associated with BD or related disorders at p<10-7. We analyzed alternative isoforms in these BD GWAS-implicated genes using STRINGTIE. This method identified 515 isoforms, of which 190 were apparently novel, since they have not been called by ENSEMBL, GENCODE, or other public sources. Of the 80 apparently novel isoforms that were abundant enough for testing, 22 isoforms were differentially expressed in BD at nominal p<0. 05. These apparently novel isoforms of genes implicated in BD GWAS were more likely to be differentially-expressed in BD than novel isoforms of other genes (hypergeometric p-value < 0. 0001). At the more stringent FDR<0. 05, 9 isoforms were differentially expressed, mapping to 7 of the genes implicated by BD GWAS: ADCY2, ANK3, CACNA1C, LINC00473, LMAN2L, PBRM1, and SYNE1. Most of the differentially expressed isoforms showed intron retention; the rest represented alternative 5' or 3' splicing sites or exon skipping events. While the sample size is small and experimental validation of these apparently novel isoforms is needed, these results suggest that alternative splicing events in the brain may play an important role in etiology of BD.

1100W

A search for novel causative genes for autosomal recessive hereditary spastic paraplegia based on exome sequencing of singletons from individual families. *H. Ishiura[1], J. Mitsui[1], H. Shimazaki[2], K. Koh[3], Y. Ichinose[3], Y. Takahashi[4], J. Goto[1,5], J. Yoshimura[6], K. Doi[6], S. Morishita[6], H. Sasaki[7], Y. Takiyama[3], S. Tsuji[1], JASPAC (Japan Spastic Paraplegia Research Consortium).* 1) Department of Neurology, The University of Tokyo, Tokyo, Japan; 2) Department of Neurology, Jichi Medical University, Tochigi, Japan; 3) Department of Neurology, Yamanashi University, Yamanashi, Japan; 4) Department of Neurology, National Center of Neurology and Psychiatry, Tokyo, Japan; 5) Department of Neurology, International University of Health and Welfare Mita Hospital; 6) Department of Computational Science, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan; 7) Department of Neurology, Hokkaido University Graduate School of Medicine, Hokkaido, Japan.

[Background]Hereditary spastic paraplegia (HSP) is a neurodegenerative disorder characterized by spasticity and pyramidal weakness of the lower limbs. The mode of inheritance of HSP can be autosomal dominant (ADHSP), autosomal recessive (ARHSP), X-linked, and mitochondrial. The causative genes for ARHSP are more heterogeneous than those of ADHSP and about 60% of ARHSP patients remain to be elucidated. In the majority of the families, affected individuals are singletons, making it difficult to efficiently narrow the candidate region with linkage analyses. We tried to identify novel causative gene for ARHSP based on exome sequencing of singletons from individual families. [Methods]One hundred and eight patients of undiagnosed HSP whose family histories were compatible with autosomal recessive mode of inheritance were enrolled. Nonsynonymous and splicing variants whose minor allele frequencies were <0. 2% were selected from exome data. We then searched for patients carrying biallelic variants, and examined if the genes with biallelic variants were shared by multiple families. [Results]We identified four families, where affected individuals shared biallelic variants in the same gene. We found two homozygous nonsense variants and one homozygous missense variants in two families. Homozygous nonsense variants strongly support loss-of-function mutations relevant to the disease. The missense variant is located in a conserved catalytic domain and the haplotypes surrounding the missense mutation were identical in the two families. The clinical presentation of these four patients are characterized by ataxia or dysarthria in addition to spastic paraplegia. Thus, the gene is considered to be causative gene for ARHSP. [Discussion]Utilizing exome data, we attempted to identify causative genes for ARHSP, focusing on genes with biallelic variants shared by multiple families. We identified a novel putative gene for ARHSP, based on the findings that 4 families have homozygous variants including two independent nonsense variants. The present study revealed that our study paradigm is efficient for identifying causative genes even though only singletons from individual families are available for the analysis. On the other hand, causative genes for ARHSP seem to be heterogeneous with each accounting for less than 1-2% of families individually. To further identify causative genes based on this strategy, it will be essential to further increase the number of families.

1101T

Characterization of the transcriptome of vestibular dysfunction associated with *SLC26A4* mutations. *Y. H. Chan[1], Y. C. Lu[1], C. W. Yu[1], I. S. Yu[2], Y. H. Lin[1,3], P. L. Chen[4,5,6,7], C. C. Wu[1,7], C. J. Hsu[1].* 1) Department of Otolaryngology, National Taiwan University Hospital,Taipei, Taiwan; 2) Transgenic Mouse Models Core, Division of Genomic Medicine, Research Center for Medical Excellence, National Taiwan University, Taipei, Taiwan; 3) Graduate Institute of Molecular Medicine, National Taiwan University College of Medicine, Taipei, Taiwan; 4) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 5) Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan; 6) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 7) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan.

Recessive mutations in *SLC26A4* are responsible for non-syndromic enlarged vestibular aqueduct (EVA) and Pendred syndrome (PS), constituting a common cause of hearing impairment in children. To date, several mouse models with *Slc26a4* mutations have been established. Corresponding to the human counterpart, mice with defected *Slc26a4* revealed both auditory and vestibular dysfunction. In contrast to the phenotype of hearing loss which has been exhaustively investigated in these *Slc26a4*-defected mice, the pathogenetic mechanisms underlying the vestibular dysfunction associated with *Slc26a4* mutations remain largely unexplored. In the present study, we performed RNA-seq analyses on the inner ear extracts obtained from three groups of mice at 8 weeks: *Slc26a4*-defected mice with circling behavior (C; n=6), *Slc26a4*-defected mice without circling behavior (NC; n=6), and wild-type mice (WT; n=6). The resulting expression data were then subjected to Ingenuity Pathways Analysis (IPA), and the differentially expressed genes were validated in vivo via immunohistochemistry. We identified 90 genes differentially expressed between the C and WT groups, 8 genes differentially expressed between the NC and WT groups, and 11 genes differentially expressed between the C and NC groups. Genes considered highly correlated to vestibular dysfunction included: *Dynlt1f*, *Entpd4*, *Fam25c*, *Gm3776*, *Hbq1a*, *Hist1h2ai*, *Hist1h4m*, and *Lst1*. Among these genes, we have validated the expression of 6 genes in the inner ear using immunostaining. Canonical pathways identified by IPA included glutathione-mediated detoxification and NRF2-mediate oxidative stress response, indicating that reactive oxygen species might play a pivotal role in the development of vestibular phenotypes. Our results provide insights into the molecular pathology of vestibular dysfunction related to *SLC26A4* mutations, and may inform future studies on the potential therapeutic implications of modulating the associated genes or pathways.

1102F

The genetic architecture of Autism Spectrum Disorders in the Faroe Islands. *C. Carton[1,2,3], G. Huguet[1,2,3], J. Buratti[1,2,3], A. Mathieu[1,2,3], A. Boland[4], J. F. Deleuze[4], J. Halling[5], E. Billstedt[5], C. Gillberg[5], T. Bourgeron[1,2,3,5,6].* 1) Institut Pasteur, Paris, Paris, France; 2) CNRS UMR3571 Genes, Synapses and Cognition, Institut Pasteur, Paris, France; 3) University Paris Diderot, Sorbonne Paris Cite, Human Genetics and Cognitive Functions, Paris, France; 4) Centre for National Genotyping, Evry, France; 5) Gillberg Neuropsychiatry Centre, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 6) Fonda-Mental Foundation, Creteil, France.

Autism spectrum disorders (ASD) are a group of neuropsychiatric disorders characterized by deficits in social communication, as well as presence of restricted interests, stereotyped and repetitive behaviors. The genes associated with ASD remain largely unknown mostly because of the clinical and genetic heterogeneity of this complex syndrome. Here, we studied the genetic architecture of 349 individuals from the Faroe Islands including 33 patients with ASD, 105 relatives and 210 controls using genome-wide genotyping of >5 millions SNPs and whole-exome sequencing. The Faroe Islands are located between the Norwegian Sea and the North Atlantic Ocean, approximately at half distance from Norway and Iceland. As expected, we first observed a higher number of runs of homozygosity (ROH) compared to other populations (P<0. 0001). Interestingly, within the Faroe population, patients with ASD had inbreeding coefficient compared with controls (F(ASD)=0. 007 vs. F(controls)=-0. 0045; P=0. 0002), suggesting that, in a subset of patients, recessive mutations could contribute to the risk of ASD. We then analyzed the contribution of *de novo* mutations (CNVs, SNVs and indels) in families with both parents DNA available (12 for SNVs and 17 for CNVs). For 3 patients, we identified *de novo* mutations in loci previously associated with ASD: 1 deletion of 22q11, 1 deletion of *NRXN1*, and 1 damaging missense mutation of *MECP2*. We also identified inherited rare CNVs altering exons of genes previously associated with ASD (*ADNP*, *BCL9*, *IMMP2L*, *TBL1X*, *TBL1XR1* and *ACACA*). We then observed a significant enrichment of rare inherited loss of function (LoF) mutations in the Ephrin receptor *EPHA1* (P=0. 01) and in the sodium channel *SCN3A* (P=0. 04), which could represent risk factors for ASD. Finally, our analysis also revealed inherited duplications of *IQSEC3* and a *de novo* stop mutation of *RIMS4*, two genes playing a role in the establishment of neuronal connectivity. In summary, based on the results from this pilot study, it seems that the genetic architecture of ASD in the Faroe Islands is not dramatically different from other populations. We were able to identify a genetic cause of ASD in at least 10% the individuals (20% of those actually examined). Further analyses are currently in progress, with a view to better understanding the interplay between common and rare variants in the susceptibility and severity of ASD in these patients.

1103W

Whole exome sequencing of parent-proband trios to identify genes implicated in hyperserotonemia and autism spectrum disorder. *R. Chen[1], J. Sutcliffe[1], S. Guter[2], L. Davis[3], N. Cox[1], B. Li[1], E. Cook[2].* 1) Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN; 2) Department of Psychiatry, University of Illinois at Chicago, Chicago, IL; 3) Department of Medicine, University of Chicago, Chicago, IL.

Autism spectrum disorder (ASD) is a highly heritable neuropsychiatric disorder with heterogeneous genetic etiology. Numerous lines of evidence have linked serotonergic dysfunction to this disorder, including that (1) significant elevation of platelet serotonin (5-HT) was observed in ~35% of ASD cases, termed hyperserotonemia; (2) selective serotonin reuptake inhibitors can efficiently relieve some symptoms including ameliorating anxiety and irritability related to insistence on sameness and other rigid-compulsive behaviors. This link suggests the possibility of a serotonergic pathway-related mechanism contributing to the complex etiology of ASD. Serotonin itself is highly heritable, but the genetic basis of 5-HT and its relationship to ASD are largely unknown. Given that we observe elevation of serotonin in ASD probands relative to their parents, we hypothesize that *de novo* mutations (DNMs) and rare variants with recessive inheritance may play an important role in the predisposition to hyperserotonemia in ASD probands, and propose to use these observations to identify genes and pathways contributing to hyperserotonemia and ASD. In this study, we carried out whole exome sequencing of 118 ASD parent-proband trios with 5-HT measurements. We followed best-practice procedures for data processing and variant calling. In particular, we experimentally validated all functional DNM candidates and obtained 125 functional DNMs in 124 genes. We used the phase-by-transmission algorithm to construct haplotypes and achieved almost certain phasing for infrequent-to-rare variants (allele frequency < 0. 05), enabling accurate identification of compound heterozygotes (CH; 2-hit) to investigate the recessive model. We dichotomized the probands into high and low groups and identified 149 (AF<0. 05) and 49 (AF<0. 01) CH genes enriched in the high group. Gene-set enrichment analysis showed that these CH genes overlap significantly with the TADA gene lists reported previously as ASD-associated genes; the significance was obtained empirically to control a gene-length bias by sampling random gene sets matching the gene length distribution. Analysis of DNMs showed genes harboring functional DNMs significantly overlap with genes involved in ASD such as FMRP targets, chromosome remodeling genes and synapse-related genes. We are applying network analysis of DNMs and CH genes in the high vs. the low group to further elucidate the pathways encoding serotonergic function dysregulated in ASD.

1104T

Assessing the impact of inherited and *de novo* CNVs in Autism Spectrum Disorder in high-risk infant siblings using high-resolution microarrays. *L. D'Abate[1,2], S. Walker[1], R. K. C. Yuen[1], K. Tammimies[1,3], M. J. Gazzellone[1], B. Thiruvahindrapuram[1], J. Howe[1], J. Brian[4], S. Bryson[5], K. Dobkins[6], R. Landa[7], D. Messinger[8], S. Ozonoff[9], I. Smith[5], W. Stone[10], Z. Warren[11], G. Young[9], L. Zwaigenbaum[12], S. W. Scherer[1,2].* 1) The Centre for Applied Genomics, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada; 3) Center of Neurodevelopmental Disorders (KIND), Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden; 4) Autism Research Centre, Bloorview Research Institute, Toronto, Ontario, Canada; 5) IWK Health Centre, Dalhousie University, Halifax, Nova Scotia, Canada; 6) Department of Psychology, UC San Diego, La Jolla, California, USA; 7) Hugo Moser Research Institute, Kennedy Krieger Institute, Baltimore, Maryland, USA; 8) Department of Psychology, University of Miami, Coral Gables, Florida, USA; 9) MIND Institute, UC Davis, Davis, California, USA; 10) Autism Centre, University of Washington, Seattle, Washington, USA; 11) Vanderbilt Kennedy Center Treatment and Research Institute for Autism Spectrum Disorders, Vanderbilt Kennedy Centre, Nashville, Tennessee, USA; 12) Autism Research Centre, University of Alberta, Edmonton, Alberta, Canada.

Background: Studies conducted on trios and quartet families have implicated rare *de novo* and inherited copy number variants (CNVs) in the complex etiology of Autism Spectrum Disorder (ASD). The Baby Siblings Research Consortium (BSRC), spanning 21 sites world-wide, is a unique sibling cohort with comprehensive data on the development and therapeutic outcomes of infant siblings of children with ASD. As clinical microarrays are commonly used to identify the genetic contributors of ASD, we are investigating whether genetic testing in young children can inform of ASD status prior to the typical age of onset of symptoms. Here, we correlated our genetic findings with the clinical data available on infant siblings, allowing us to assess the predictive value of early genetic testing. **Methods:** A total of 473 individuals (239 probands, 78 siblings with ASD, 156 unaffected siblings) were genotyped on the high-resolution Affymetrix CytoScan HD platform. All infant siblings were clinically assessed periodically and given an ASD diagnosis if appropriate at 3 yrs. of age. High-confidence CNVs were detected using a combination of four algorithms, with rare CNVs being identified by comparing with 873 platform-matched controls from the Ontario Population Genomics Platform. **Results:** From the first 200 families analyzed, we observed a *de novo* CNV rate of 5. 5% in affected individuals. This included a 610 kb deletion in the 16p11. 2 locus, corresponding to the region affected by the microdeletion syndrome. A *de novo* rate of 1. 5% was observed in unaffected siblings, which is comparable to the rate observed in population controls. Likely pathogenic rare CNVs were found in 7. 2% of affected individuals and 2. 9% of unaffected siblings. Examples include a 57 kb deletion in *MBD5* in two affected male siblings and a 200 kb deletion in *NRXN1* in a male proband and his female unaffected sibling. A 90 kb deletion affecting the third exon of *PTCHD1-AS1* was observed in a male proband and his male sibling of unknown status. Some unaffected siblings harbouring potentially pathogenic variants displayed ASD-like symptoms later in development and are currently being clinically reassessed. **Conclusions:** Although analysis of the cohort is ongoing, initial findings show that identification of potentially pathogenic variants can inform on an ASD phenotype in early childhood.

1105F

Identification of *de novo* and rare inherited mutations in autism spectrum disorder by whole genome sequencing. *H. El-Shanti[1], M. Kambouris[1], Y. Al-Sarraj[1], X. Jin[2], M. Wang[2], R. K. C Yuen[3], V. Chini[1], Y. Begaoui[1], K. Errafii[1], R. Z. Taha[1], H. Boulos[4], N. Khattab[1], H. Shaath[1], X. Wu[2], J. Ju[2], M. Al-Mutawa[1], W. Habbab[1], I. R. Thompson[1], H. Abu Al-Khair[5], M. Abdel Gani[5], F. Alshaaban[1], K. Shalaby[1], G. Wang[2], Z. Wang[2], N. Chen[2], A. Shih[6], G. Dawson[6], Y. Jiang[7], S. Scherer[3], A. Farooq[8].* 1) QBRI Medical Genetics Center, Doha, Qatar; 2) BGI-Shenzhen, Bei Shan Road, Yantian, Shenzhen, China; 3) The Centre for Applied Genomics, Hospital for Sick Children and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada; 4) Human Genetics Department, University of Chicago, Chicago, IL, USA; 5) Shafallah Center for Children with Special Needs, Doha, Qatar; 6) Autism Speaks, New York, NY, USA; 7) Duke University, Durham, NC, USA; 8) Department of Biochemistry and Molecular Biology, Leonard Miller School of Medicine, University of Miami, Miami, FL, USA.

Autism Spectrum Disorder (ASD) is a lifelong neurodevelopmental disorder characterized by deficits in social communication and interaction, repetitive and restrictive behavior, extensive clinical and etiologic heterogeneity, as well as, a remarkably rising global prevalence rate. It is currently believed that ASD has an underlying genetic basis, ranging from effects of single genes to that of multiple genes and chromosomal regions, hallmarking the multifactorial and complex etiology. While a genetic etiology is identifiable in about one quarter of the cases, in the remaining three quarters it remains elusive. Recently, whole genome sequencing (WGS) emerged as a powerful tool for variant discovery due to its comprehensive and uniform coverage. We used WGS to examine a cohort of 20 trios with a child diagnosed with ASD by, at least, an ADI-R (Autism Diagnostic Interview-revised), for *de novo* or rare inherited genetic variants. Due to the nature of the ascertainment source, all probands had associated intellectual impairment. In all probands known causes of ASD were excluded, fragile X syndrome, Rett syndrome (for females) and Copy Number Variations (CNV) by SNP genotyping on an Illumina 1M-Duo SNP array. Autosomal recessive ASD alleles were inspected by WGS coupled with homozygosity mapping. The cohort is comprised of 16 male and 4 female probands and their parents. Twelve of the parent pairs were consanguineous (60%) with cumulative homozygosity intervals ranging from 37 Mb to 583 Mb. One *de novo* mutation in a known gene (*ADNP*) was identified in one family (5%) , but 5 other potentially pathogenic de novo mutations in putative genes were identified in 5 families (25%). Rare inherited X-linked mutations in two known ASD risk genes (*UPF3B*, *KDM5C*) were identified in two families (10%), but potentially pathogenic mutations in putative genes were identified in 3 other families (15%). We did not identify any rare inherited autosomal dominant mutations in the known ASD genes. One of the values of studying inbred populations is the evaluation of identical by descent variants in genes that can be further examined for compound heterozygosity in disease-specific databases. We identified rare putative inherited autosomal recessive homozygous variants in 3 families (15%). We conclude that WGS is a powerful tool for studying the genetic causes of ASD. We also conclude that consanguineous families can provide clues to autosmal recessive ASD risk genes.

1106W

No Evidence That Differences In Cortical DNA Methylation Contribute to Autism. *S. Ellis[1], S. Gupta[1], A. Moes[1], D. Absher[2], A. B. West[3], D. E. Arking[1].* 1) Human Genetics, Johns Hopkins University School of Medicine, Baltimore, MD; 2) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 3) Neurology, University of Alabama School of Medicine, Birmingham, Alabama.

Autism spectrum disorder (ASD) develops in early childhood and continues as a lifelong neurodevelopmental disorder that affects approximately one in 100 individuals worldwide. To date, GWAS and gene expression studies have implicated more than 100 loci contributing to the disorder; however, the majority of risk remains unexplained. Given prenatal brain development's critical role in the progression of ASD and a period sensitive to alterations in epigenetic pathways, researchers have begun to investigate the role of DNA methylation (DNAm) in ASD to uncover more of this unexplained risk. Microarray studies have looked for DNAm differences between ASD cases and controls but have been limited by small sample sizes, the digital nature of methylation arrays, and the use of material other than the brain – the primary affected tissue in ASD. To further test the hypothesis that DNAm plays a role in ASD, we measured CpG DNAm in a set of cortical brain samples (BA19) by two independent technologies, using both next-generation methylation sequencing (reduced representation bisulfite sequencing (RRBS), N=63) and methylation microarrays (Illumina Infinium HumanMethylation27 BeadChip, N=79). Samples comprised of controls and individuals with autism – a subset of ASD – all of whom were diagnosed using ADOS and ADI-R (among other exclusions). To assess the validity of the data generated, detection of previously reported brain meQTLs (26% at $p<0.05$) and age prediction based on DNAm ($R2=0.70$) were carried out in the RRBS and microarray data, respectively. Additionally, in a direct comparison of DNAm at overlapping sites between the two technologies, mean methylation values were highly correlated ($R2=0.92$). Taken together, these results suggest that DNAm was accurately captured by both technologies allowing for meaningful downstream analyses. Given our confidence in these DNAm measurements, we tested for global, regional, and individual CpG site DNAm differences between cases and controls, while accounting for both known and unknown covariates. No significant differences in DNAm were found in either the RRBS or the microarray data after bootstrapping. Further, when we focused on six previously reported regions of differential methylation, we were unable to replicate previous findings, despite our increased sample size and use of the primary affected tissue in ASD. Together, these data offer no evidence that cortical CpG DNAm contributes to differences in autism brains.

1107T

Distinct biomolecular modules group and distinguish autism genes. *F. Hormozdiari[1], O. Penn[1], E. Borenstein[1,2], EE. Eichler[1,3].* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Department of Computer Science and Engineering, University of Washington, Seattle, WA; 3) Howard Hughes Medical Institute, Seattle, WA.

There is compelling evidence that the large number of genes recently associated with *de novo* mutation in autism spectrum disorder (ASD) map to a smaller set of biological pathways and modules. Using co-expression and protein interaction data, we have enhanced our computational method named MAGI (http://eichlerlab. gs. washington. edu/MAGI/) to now incorporate private inherited mutations, in addition to *de novo* mutations for improved module discovery with better objective function. We denoted this extended method as MAGI+. We applied MAGI+ to ~10,000 exomes from 2500 simplex autism families from the SSC (Simons Simplex Collection). We group patient mutations into five biological networks with distinct biological properties. These modules in total include less than 2% of all the genes in protein interaction network yet cover over 22% of truncating mutations observed in probands. These include the WNT pathway (15. 4x, p<1. 4E-8), which is highly enriched in genes that regulate transcription (8. 8x, p<4. 6e-9), chromatin modification (especially the SWI/SNF complex) significantly enriched in cell cycle and proliferation (14. 2x, p<1. 5E-4) and long-term potentiation (33x, p<6. 1e-8) with the highest expression shown in postnatal brain. Two new modules emerge from these expanded data. This includes a module related to the neurotrophin pathway (14. 9x, 2. 4e-2) that promotes neuron survival and growth with the highest expression in fetal brain cortex. A fifth module is significantly enriched with the RIG-I-like receptor pathway (22. 4x, p<1. 5e-3) associated with the innate immune system and shows the highest expression in epithelium cells. We are currently incorporating data from an additional 3000 patients from publically accessible ASC (Autism Sequencing Consortium) and DDD (Deciphering Developmental Disorders) datasets and will present this expanded set as well as the relationship of these modules with associated phenotypes (e. g. , macrocephaly, microcephaly, intellectual disability, seizures, etc.). In addition to discovering and prioritizing novel autism genes, this network strategy provides a molecular approach to subtype autism patients for future and more effective therapeutic intervention.

1108F

Utilizing whole-exome sequencing in a multiplex family with autism spectrum disorder. *E. Koparir[1], M. C. Tarakcioglu[2], Z. Gormez[3], A. Sarac[4], M. S. Sagiroglu[3], B. Yuksel[4], A. Yuksel[5], M. Ozen[5].* 1) Medical Genetics, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul, Turkey; 2) Child and Adolescent Psychiatry, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul, Turkey; 3) Advanced Genomics and Bioinformatics Research Center, The Scientific and Technological Research Council of Turkey (TUBITAK-BILGEM), Kocaeli, Turkey; 4) Genetic Engineering and Biotechnology Institute, The Scientific and Technological Research Council of Turkey (TUBITAK-MAM), Kocaeli, Turkey; 5) Medical Genetics, Biruni University, Istanbul, Turkey.

Autism spectrum disorders (ASD) are childhood neurodevelopmental disorders characterized by impairments in social interaction, communication, and by restricted, repetitive, and stereotyped patterns of behavior. The prevalence of ASD is estimated to be around 0. 6%, making it one of the most prevalent disorders in childhood; however, the prevalence of multiplex families with autism spectrum disorder is very rare. Family and twin studies have provided strong evidence that genetic factors have a major role in the etiology of ASD. Although in recent years several high-throughput genetic studies have identified novel ASD genes in idiopathic autism, the vast majority of the underlying genetic factors remain still unknown. Several studies have been published to examine the role of whole-exome sequencing (WES) to identify genetic risk factors for ASD. In the current study, we utilized WES in a multiplex ASD Turkish family with three affected siblings to screen all disease related and candidate genes to find out the molecular etiology of the ASD. The mother and father were first-degree cousins. Two of the siblings were diagnosed with ASD; however one was diagnosed with non-syndromic mental retardation. In conclusion, we were able to identify two candidate genes in our informative family. Confirmation of the variants and segregation analysis are in progress.

1109W

One in three *de novo* variants seen in Autism Spectrum Disorder probands are present as standing variation in a cohort of more than 60,000 non-ASD individuals. *J. A. Kosmicki[1,2,3], K. E. Samocha[1,2,3], M. Lek[2,3], D. G. MacArthur[2,3], D. P. Wall[4], E. B. Robinson[2,3], M. J. Daly[1,2,3].* 1) Harvard Medical School, Boston, MA; 2) Massachusetts General Hospital, Boston, MA; 3) Broad Institute of Harvard and MIT, Cambridge, MA; 4) Stanford University, Palo Alto, CA.

One of the fundamental challenges of human disease genetics is differentiating risk-conferring variants from the overwhelming amount of neutral variation in the genome. Given limited sample sizes, clear interpretations of variant impact are often restricted to those variants with large effect sizes. By leveraging the power of 60,706 reference exomes in the Exome Aggregation Consortium (ExAC), we highlight classes of risk-conferring *de novo* and inherited variation in Autism Spectrum Disorder (ASD) data sets. Previous studies of *de novo* variation commonly assumed that the vast majority of *de novo* variants are novel and heavily selected against. We examined 6192 (4,256 ASD, 1,936 control) coding *de novo* variants from 6,103 trios, and found that 32. 3% are present as standing variation in non-ASD exomes from ExAC (herein referred to as recurrent *de novo* variants). [rR1]The recurrent *de novo* variants, independent of their functional impact, are enriched for CpG sites (p<10-137) and many are present at allele frequencies above 1% in non-European populations. The non-recurrent *de novo* variants are more strongly associated with neurodevelopmental risk than those present in standing variation. This is evidenced through stronger average associations between those variants and ASD (synonymous: p=0. 6, OR=1; missense: p<10-4, OR=1. 7; nonsense: p<10-5, OR=2. 8) and low IQ in ASD cases (p<10-4) as compared to previous work (Robinson et al. , 2014; de Rubeis et al. , 2014). In contrast, recurrent *de novo* variants do not exhibit the expected patterns of enrichment in ASD (synonymous: p=0. 92, OR=1; missense: p=0. 05, OR=1. 1; nonsense: p=0. 13, OR=0. 7), low-IQ (p=0. 89) or female (p=0. 69) individuals. We also examined inherited rare variants for presence in ExAC. As with the *de novo* variants, ASD individuals are enriched for inherited, singleton, loss-of-function (LoF) variants not present in ExAC (p<10-5; OR=1. 8). Additionally, several *de novo* variants not observed in ExAC were transmitted to ASD case individuals in other trio cohorts, notably a nonsense variant in *ANK2* and a probably-damaging missense variant in *RGL1*. In a separate case control cohort (1445 cases and 4424 controls), we found that ASD cases are also enriched for rare LoF variants not seen in ExAC (p<10-25; OR=2. 35). [rR1].

1110T

Investigating Somatic Mosaicism in Simplex Autism Spectrum Disorder. *D. Krupp[1], Y. Duffourd[2], S. Evans[1], R. Bernier[3], E. Fombonne[4], S. J. Webb[3], J. B. Rivière[2], B. J. O'Roak[1].* 1) Molecular and Medical Genetics, OHSU, Portland, OR; 2) Génétique des Anomalies du Developpement, Université de Bourgogne, Dijon, France; 3) Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 4) Psychiatry, OHSU, Portland, OR.

Autism spectrum disorder (ASD) is strongly genetic with complex architecture. Somatic mosaicism has been implicated in several neurodevelopmental disorders including epilepsy, cortical malformations, and overgrowth syndromes (Poduri et al. 2013). Pathways underlying these syndromes, i. e. mTOR/AKT/PI3K, are also implicated in syndromic and nonsyndromic ASD. Moreover, in a previous ASD exome study we found 5% of *de novo* events were consistent with mosaicism (O'Roak et al. 2012). We are systematically evaluating somatic mosaicism in exomes of ~2,500 families from the Simons Simplex Collection (SSC), including parents, proband, and an unaffected sibling. Our goals are to determine rates of mutations in affected and unaffected children, overlap with ASD genes/pathways, transgenerational risk, and to infer when in development these events might occur. We began by examining presumed *de novo* germline calls from the SSC (Iossifov et al. 2014, Krumm et al. 2015). We replicated previous observations, finding 4% of sites consistent with somatic events (binomial p<0. 001). We found 25 variants in genes recurrently mutated in ASD, including *SYNGAP1*. To identify variants with reduced allele fractions (that would have been missed previously), we developed an integrated calling and validation strategy. We are applying three somatic callers, VarScan 2. 3. 2, LoFreq 2. 1. 1, and our own caller. We are then validating these sites using molecular inversion probes (MIPs), enabling massive multiplexing of target sites and samples. Each MIP incorporates a unique molecular tag, allowing us to achieve statistical accuracy and correct sequencing errors. With this approach, we have been able to detect somatic mosaicism with 2-27% allele fractions in exome and resequencing data. For further analysis development and evaluation, we are leveraging 24 deeply sequenced (median ~200x) families. We have identified 9 high-confidence putative mosaic sites in probands and 8 in siblings, and are also examining parental mosaic variants transmitted to children. Using these data, we are refining the analysis process and applying it to 400 families. This systematic evaluation of mosaicism in ASD families will help elucidate how frequently somatic events occur during development and their contributions to pathogenesis of neurodevelopmental disorders. Moreover, the methods and tools developed here will allow for detailed analyses of specific tissues, i. e. brain, in future studies.

1111F

CNV Burden in ASD and its Association with Developmental and Behavioral Impairment. *J. M. Lee[1], A. J. Griswold[1, 2], J. R. Gilbert[1,2], M. A. Pericak-Vance[1,2], M. L. Cuccaro[1,2].* 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL.

Autism Spectrum Disorder (ASD) is a highly prevalent neurodevelopmental disorder. Copy number variations (CNVs) contribute to genetic risk for ASD and are more likely in individuals with ASD who have developmental impairments (DI). A common feature among individuals with DI is behavioral impairment (BI). For this study, we examined CNV burden in individuals with DI and BI. We hypothesized that CNV burden will be greatest in individuals with both developmental and behavioral impairments vs a single impairment. To test this hypothesis we investigated CNV burden in an ASD dataset (N=527 European ancestry) which was divided into four groups based on DI and BI values. DI scores were based on an adaptive behavior scale while BI scores were based on a behavior problem checklist. Using these scores we constructed four ASD groups: DI only (N=166), BI only (N=56), BI+DI (N=116), and no BI or DI (N=189). A genome-wide SNP array was utilized for CNV detection using QuantiSNP and PennCNV algorithms. To test for group differences in CNV burden we compared the groups on mean number of deletions and duplications using standard univariate tests. For our ASD dataset, 46. 95 (sd 22. 6) deletions and 20. 38 duplications (sd=11. 5) were detected per individual. When analyzed by DI (+/-), neither mean number of deletions (p=. 08) or duplications (. 38) differed between the DI and no DI groups. When analyzed by BI (+/-), neither mean number of deletions (p=. 61) or duplications (. 63) differed between the BI and no DI groups. Analyzing DI and BI jointly, the mean number of deletions was highest in the BI+DI group (49. 52, sd=26. 7) while the BI only group had the lowest number of deletions per individual (43. 30, sd=17. 7). The mean number of deletions across the four groups did not significantly differ (p=0. 29). Mean number of duplications was highest in the BI only group (22. 37, sd=13. 8) while the BI+DI had the lowest mean number of duplications (19. 92, sd=12. 2). The mean number of duplications across the four groups did not significantly differ (p=0. 56). Comparisons of mean deletion and duplication values by group and sex revealed no differences. In our dataset, the co-occurrence of developmental and behavioral impairments is not associated with increased CNV burden as we did not see excessive CNVs in any group. This is surprising given prior studies. Future studies should further examine the contributions of severity in associated phenotypes in dissecting genetic risk for ASD.

1112W

Genetic analysis for circadian rhythm abnormality in autism spectrum disorder. *A. Matsumoto[1], Y. Inaguma[2], Z. Yang[1], Y. Nakano[1], M. Goto[1], K. Nakayama[3], E. F. Jimbo[1], H. Osaka[1], S. Iwamoto[3], K. Nagata[2], T. Yamagata[1].* 1) Pediatrics, Jichi Medical University, Shimotsuke, Japan; 2) Department of Molecular Neurobiology, Institute for Developmental Research, Aichi Human Service Center, Kasugai, Japan; 3) Division of Human Genetics, Center for Molecular Medicine, Jichi Medical University, Shimotsuke, Japan.

(Background) About 44-83% of children with autism spectrum disorder (ASD) are suffering from sleep problems. Some circadian-relevant genes, *MTNR1A*, *MTNR1B*, and *PER1* etc. have been reported to associate with ASD. Therefore we screened for mutation of circadian-relevant genes in ASD patients, to elucidate the relation between ASD and circadian rhythm. (Methods) We screened 18 circadian-related genes, *ARNTL, CLOCK*, *PER1,2,3*, *CRY1,2*, *NR1D1*, and *TIMELESS* etc. , in 14 ASD with sleep disturbance, in 14 ASD without sleep disturbance, as well as in 23 controls. ASD was diagnosed according to the Diagnostic and Statistical Manual of Mental Disorders, 5 (DSM-5). Their DNA was extracted from lymphocytes after obtaining informed consent from their parents. We captured the all exons of circadian-relevant genes for subjects DNA and sequenced by next-generation sequencer (454 GS Junior, Roche). The detected mutations were verified by direct sequencing analysis. Some of the interesting genes were analyzed for their function. This study was approved by the Bioethics Committee for Human Gene Analysis of Jichi Medical University. (Result) Non-synonymous mutations detected only in each group were 6 out of 14 ASD with sleep disturbance group, 7 out of 14 ASD without sleep disturbance group, and 1 out of 23 controls. In all of ASD (P=0. 001), non-synonymous mutations were significantly frequently detected than control group. We detected p. F498S of Timeless in a patient with ASD and circadian rhythm disorder. He was short stature (-3. 4SD). He presented mild disturbance of sociability, repeated patterns of behavior, and mild ID. He fell into sleep in the morning around 5-7 o'clock since his infancy. The peak of the Melatonin of this patient and mother was 8-11 o'clock, which was 5 hours difference from normal pattern. Light therapy didn't change the circadian rhythm, backward nor forward. Timeless p. F498S mutation affected the neuronal function in utero electroporation analysis. (Discussion) Mutations in circadian-relevant genes were more frequent in patients with ASD than in controls. Circadian-relevant genes may be involved in the psychopathology of ASD. *TIMELESS* has a role of DNA replication and checkpoint. It was reported that *Timeless* knockdown mouse arrested the development. Melatonin profile and abnormal neuronal migration indicated that *TIMELESS* may contribute to ASD and circadian rhythm disorder. We will further analyze the function using TIMELESS knockout mouse.

1113T

Rare loss-of-function mutations in ASD individuals. *D. P. Moreira, E. S. Moreira, N. C. V Lourenço, I. M. W. Silva, T. F. Almeida, V. L. R. Tavares, M. Zatz, M. R. Passos-Bueno.* Dept of Genetics and Evolutionary Biology, Institute of Biosciences - University of Sao Paulo, São Paulo, Brazil.

Autism Spectrum Disorder (ASD), a neurocognitive condition, has a strong genetic component with heritability estimated between 50-60%. Most recently, exome/genome sequencing has shown that rare truncating variants, *de novo* or inherited, strongly contribute to ASD etiology. These alterations are enriched in ASD affected probands as compared to unaffected siblings. However, it is still a challenge to define which of these variants are indeed causative of the phenotype, especially because some of them are inherited, being associated with non-penetrance of the phenotype. Thus, we have been investigating rare truncating variants shared among ASD related individuals with the aim of identifying those that contribute to the phenotype. We performed exome sequencing of 26 ASD-affected individuals belonging to 12 families. Rare variants (MAF < 0. 01 in international databases and in a Brazilian genomic database) were initially selected for further analysis. Of these, we selected as potentially pathogenic variants those that are truncating alterations (premature stop codon variants or frameshift variants) located in genes that are expressed in central nervous system. We identified 13 truncating variants in 7 out of 12 families, which were shared between the affected individuals within a family. *In silico* functional analysis of these genes revealed that six out of the 12 genes are related with post-transcriptional and post-translational modifications, which are biological functions strongly associated with ASD. In most of these families we could not define if there is a major mutation associated with the phenotype. However, in one family, two affected sibs have a loss-of-function mutation in *TBCK*, inherited from a non-affected father, and an intragenic deletion involving just one exon of this same gene, maternally inherited. Our results suggest that homozygous loss-of-function mutations in *TBCK* leads to ASD, confirming the relevance of mutations in this gene for neurodevelopmental disorders. Although our sample is small, our preliminary data suggest that a proportion of ASD cases might be explained by Mendelian autosomal recessive model of inheritance. FAPESP/CEPID, CNPq.

1114F

Mitochondrial Variants in Autism Spectrum Disorder through off target exome read analysis. *A. Patowary, R. Nesbitt, M. Archer, Z. Brkanac.* University of Washington, Seattle, WA.

Autism spectrum disorder (ASD) is a severe neuropsychiatric disorder with complex genetics. Recent large-scale exome sequencing studies have identified a large number of genes where it is likely that *de-novo* variants are contributory, indicating substantial heterogeneity. Thus far, large-scale exome sequencing studies have concentrated on the nuclear genome, leaving the putative role of mitochondrial DNA (mtDNA) mainly unexplored. MtDNA mutations are causal for multiple early onset disorders that also have autistic features as part of the phenotype and thus mtDNA is intriguing as a candidate for mutation responsible for ASD. The mitochondrial genome (NC_012920) is circular and double stranded, and contains16,569 nucleotides coding for 37 genes. Recently, analytical tools have been developed to extract mitochondrial DNA reads from unmapped reads in whole-exome capture sequence data. Off-target sequence reads can be used to derive the mtDNA sequence with a high degree of accuracy. We have analyzed the mitochondrial DNA sequence derived from off-target exome reads from eight unrelated ASD probands from families with multiple affected cases. The mitochondrial genomes for all the samples were sequenced to an average depth of >25X. With a minimum heteroplasmy limit of 0. 8, we have identified 81 variants in eight mitochondrial genomes, including 16 non-synonymous variants. We will be extending our analysis of mtDNA to additional affected family members in the original eight families as well as cases from 15 additional families. Detailed findings will be presented.

1115W

A 1. 5Mb terminal deletion of 12p associated with autism spectrum disorder. *S. Raskin[1,2], I. M. Silva[1], J. A Rosenfeld[3,4], S. A. Antoniuk[5], V. S. Sotomaior[1].* 1) Group for Advanced Molecular Investigation (NIMA), School of Health and Biosciences, Pontifícia Universidade Católica do Paraná (PUCPR), Curitiba, Paraná, Brazil; 2) GENETIKA - Centro de Aconselhamento e Laboratorio de Genetica, Curitiba, Paraná, Brazil; 3) Signature Genomics, PerkinElmer, Inc. , Spokane, WA, USA; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX. USA; 5) Pediatrics Department, Universidade Federal do Paraná, Curitiba, Paraná, Brazil.

Autism spectrum disorder ischaracterized by delayed and/or unusual language, problems with social interactions, repetitive and stereotyped patterns of behavior and restricted interests and activities. The causes of autism have not all been clearly defined. However, there is a genetic basis, demonstrated by the high concordance between monozygotic twins, which can be as high as 90%. Although hundreds of potential autism susceptibility loci were identified in the last years, genes on 12p13. 33 have not been previously associated with Autism. We report here a 7-year-old boy with a “de novo” 1. 5-Mb terminal deletion within 12p13. 33 associated with autism. This deletion, detected by microarray-based comparative genomic hybridization, encompasses 13 genes (*B4GALNT3, CCDC77*, *ERC1*, *FBXL14*, *IQSEC3, KDM5A, LOC100292680, LOC574538, NINJ2*, *RAD52*, *SLC6A12, SLC6A13* and *WNK1)*. Furthermore, a maternaly inherited ~350kb Xp22. 31 duplication wasalso present, but wepresumed that this small, inherited aberration from the healthy mother do not contribute to the patient phenotype. There have been few previous reports about 12p13. 33 deletions, with all cases showing variable phenotypesbut none showing a clear diagnosis of Autism. The smallest region of overlap among our patient and the previously reported individuals is within the *ERC1 gene. W*e suggest that *ERC1* can be considered as a new possible candidate contributing to the autism phenotype as well as neurodevelopmental delay present in all patients. The wide range of phenotypic severity, from learning difficulties and speech delay in early childhood to autism, can be better explained by variable expressivity, as all showed, at least, learning disability during childhood. *ERC1* encodes two different protein isoforms, ERC1a that is expressed as a cytosolic protein outside of brain and does not bind to RIM proteins, wich are present in the presynaptic active zone, and ERC1b that is a brain-specific isoform and binds RIM proteins. RIMs are putative effectors for Rab3, a synaptic vesicle protein, which regulates neurotransmitter release. In conclusion we describe the first patient with an autism diagnosis associated with a 12p13. 33 deletion. While there are no others new reports of partial monosomy of distal 12p13. 33 and new information about the genes within this region, we suggest that *ERC1* is the best candidate for the neurodevelopmental delay present with variable severity in all reported patients.

1116T

Autism gene discovery in rare female-enriched multiplex families (FEMFs). *J. M. Tilghman[1,4], C. R. Stevens[2], M. J. Daly[2,3], A. Chakravarti[4].* 1) Predoctoral Training Program in Human Genetics, Johns Hopkins School of Medicine, Baltimore, MD; 2) Broad Institute of MIT and Harvard, Cambridge, MA; 3) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 4) Center for Complex Disease Genomics, Johns Hopkins University School of Medicine, Baltimore, MD.

Autism has a profound male sex bias and a higher recurrence risk for siblings of affected females compared to siblings of affected males. These observations are explained by multifactorial inheritance with a higher affection threshold for females than males (“Carter effect”). We hypothesize that affected females meet their higher threshold through a higher frequency, severity, or different biological function of pathogenic alleles. To test these hypotheses and enhance autism gene discovery, we are examining coding variation in female enriched multiplex families (FEMFs) containing 2 or more severely affected (ADIR+ and ADOS+) girls. Enrichment for more severe variants in these families should allow for the detection of autism genes within even a modest number of families. A pilot study using only 13 such families led to the identification of delta-catenin (*CTNND2*) as a definitive autism gene, and the suggestion of *CYFIP1*, showing the utility of this approach. To extend this analysis, we performed whole exome capture (Agilent 44Mb) and Illumina sequencing (>80% coverage at 20x depth) on 420 individuals from 101 FEMFs. These analyses are underway. As a start, we used data from a set of 70 probands to test for enrichment in 38 high-confidence autism genes identified primarily through recurrent loss-of-function (LoF) mutations in autism. We searched for rare (<0. 1% MAF) pathogenic variants (LoF, ±2 bp consensus splice variants and missense variants with PolyPhen2 score >0. 8). These individuals harbored a total of 5 LoF variants in 5 genes (*CHD2, CHD8, DYRK1A, SCN2A, WDFY3*) and 19 damaging variants in 17 genes (*ANK2, DIP2A, CHD2, CHD8, GRIN2B, RIMS1, DLG1, PTEN, WDFY3, KDM5B, SCN2A, SHANK3, MECP2, DSCAM, DYRK1A, SYNGAP1, TNRC6B*). Both sets of LoF and damaging variants were significantly enriched over expectations (P=0. 0031 and 0. 021 for variant counts; P<0. 0001 and 0. 0033 for gene counts), determined by sampling 10,000 replicates of 70 samples from 475 European Ancestry controls and counting rare pathogenic variants (as defined above). The observed enrichment of these variants confirms that known autism genes contribute to autism pathogenesis in FEMFs as well as low-risk families.

1117F

Whole exome sequencing in autism spectrum disorder: confirmation of ASD-risk genes and identification of new candidates. *J. VanGils[1,2,3], S. Calderari[1,2,3], J. Buratti[1,2,3], A. Mathieu[1,2,3], M. Benabou[1,2,3], A. Boland[4], J-F. Deleuze[4], B. Régnault[5], A. Maruani[6], F. Amsellem[6], A. Beggiato[6], M. Leboyer[7,8,9,10], R. Delorme[1,2,3,6], T. Bourgeron[1,2,3,7].* 1) Institut Pasteur, Human Genetics and Cognitive Functions Unit, Paris, France; 2) CNRS UMR 3571: Genes, Synapses and Cognition, Institut Pasteur, Paris, France; 3) Université Paris Diderot, Sorbonne Paris Cité, Human Genetics and Cognitive Functions, Paris, France; 4) Centre National de Génotypage, Evry, France; 5) Eukaryote Genotyping Platform, Genopole, Institut Pasteur, Paris, France; 6) Assistance Publique-Hôpitaux de Paris, Robert Debré Hospital, Department of Child and Adolescent Psychiatry, Paris, France; 7) INSERM U955, Psychiatrie Génétique, Créteil, France; 8) Université Paris Est, Faculté de Médecine, Créteil, France; 9) Assistance Publique-Hôpitaux de Paris, Hôpital H. Mondor and A. Chenevier, Département de Psychiatrie, Créteil, France; 10) FondaMental Foundation, Créteil, France.

Autism Spectrum Disorders (ASD) are heterogeneous neurodevelopmental disorders diagnosed in more than 1% of the population and characterized by impaired social communication and restricted or stereotyped interests. Thanks to whole exome sequencing technology, several ASD-risk genes were identified. Here, we analysed 99 whole exome sequencing data from 27 families (23 ASD families and 4 control families) to identify inherited and *de novo* variants associated with ASD. We found 28 damaging *de novo* mutations in 15 of 27 (55. 6%) families for an average of 0. 62 *de novo* damaging mutations per individual. Among these genes, 6 were already reported as ASD-risk genes. In three families, the identified mutations in *SCN2A* [MIM182390], *GRIN2B* [MIM138252] and *SHANK3* [MIM606230]) most likely contribute to the diagnosis of ASD. In two families with more than one affected child, the *de novo* ASD risk variants —*MECP2* [MIM300005] and *CNTN6* [MIM607220] —were not shared by the affected siblings. We also identified a *de novo* stopgain mutation in the *POGZ* gene [MIM614787], known as one of the major gene for ASD, in a control child. Finally, we identified *de novo* damaging mutations in 2 new candidate genes for ASD: *ACTL6B* (also known as *BAF53b*) [MIM612458] and *KIF1A* [MIM601255]. *ACTL6B* is a subunit of nBAF complex involved in the neuron-specific chromatin remodelling and has been reported in neurodevelopmental and intellectual disability (ID) disorders. *KIF1A* is involved in the transport of synaptic vesicles along axons and has already been reported as being involved in individual cases of nonsyndromic ID. Our results confirm the contribution of *de novo* mutations in ASD, but also highlight the crucial need of including unaffected siblings and healthy controls to better interpret the data. Our pilot study has only analysed the submerged part of the genetic iceberg of ASD. The integration of polygenic dimension, including the burden of inherited rare mutations and the accumulation of a high load of low-risk alleles, will allow us to better understand the complex genetic landscape of ASD.

1118W

A case of Autism Spectrum Disorder with a novel *SRCAP* mutation. *S. Walker[1], K. Tammimies[2], R. K. C. Yuen[1], B. Thiruvahindrapuram[1], G. Kaur[1], M. T. Carter[3], N. Hoang[3], R. Weksberg[3], B. A. Fernandez[4,5], P. Szatmari[6], S. W. Scherer[1,7].* 1) Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 2) The Center of Neurodevelopmental Disorders at Karolinska Institutet (KIND), Pediatric Neuropsychiatry Unit, Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden; 3) Department of Pediatrics, Division of Clinical and Metabolic Genetics, Hospital for Sick Children, Toronto, Ontario, Canada; 4) Disciplines of Genetics and Medicine, Memorial University of Newfoundland, St Johns Newfoundland Canada; 5) Provincial Medical Genetic Program, Eastern Health, St. Johns Newfoundland, Canada; 6) Centre for Addiction and Mental Health, University of Toronto, Toronto, Ontario, Canada; 7) McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada.

Autism Spectrum Disorder (ASD) is highly heterogeneous neurological condition characterised by limited communication skills, impaired social interaction and repetitive behaviours. There is a strong but complex genetic etiology with many contributing risk loci. Exome sequencing has been conducted on a cohort of 1400 Canadian individuals, including 825 affected cases. Sequencing was carried out using Life Technologies SOLiD5500xl and IonProton platforms. This cohort presents a useful resource to the field of ASD genetics. From these data, a single case has been discovered with a novel loss of function mutation in *SRCAP,* a gene implicated in ASD risk by rare copy number variant analysis. The mutation is a four base pair deletion in exon 27, leading to a premature stop codon which arose *de novo* (not present in the parents or three unaffected siblings). *SRCAP* functions in chromatin remodeling and transcriptional regulation. Loss of function mutations within exons 33 and 34 are known to cause Floating Harbor Syndrome (FHS) which is characterized by language delay, short stature and distinctive facial features; however, the phenotypic impact of mutations outside of this critical region is not known. The proband is male and presented with high functioning ASD. He had café au lait spots but did not meet criteria for a diagnosis of neurofibromatosis. At 8 years, his height and weight were at the 25th percentile and head circumference at the 98% percentile. He also had large ears and small eyes. Further investigation using additional genomic data and detailed phenotypic examination is required to determine whether this case is an atypical form of FHS or if other ASD cases with mutations at this locus may have similar features. Given the high levels of genetic and phenotypic variability in ASD, coupling genomic analyses with deep phenotyping will facilitate better understanding of highly penetrant mutations underlying ASD found in very rare cases.

1119T

Identification of rare disruptive variants in voltage-gated channel genes (*CACNA1C, CACNA1D, CACNA1S, CACNA1I*) in Japanese samples of schizophrenia and autism spectrum disorder using Ion Torrent PGM platform. *C. Wang, H. Kimura, J. Xing, K. Ishizuka, I. Kushima, Y. Arioka, A. Yoshimi, Y. Nakamura, Y. Shiino, Y. Oya, Y. Takasaki, Y. Uno, T. Okada, T. Iidaka, B. Aleksic, D. Mori, N. Ozaki.* Department of Psychiatry, Nagoya University, Nagoya, Aichi, Japan.

Several large-scale whole exome sequencing studies in schizophrenia (SCZ) and autism spectrum disorder (ASD) identified rare variants with modest or strong effect size as genetic risk factors. Dysregulation of cellular calcium homeostasis might be involved in SCZ and ASD pathogenesis, and genes coding for L-type voltage-gated calcium channel (VDCC) subunits Cav1. 1 (CACNA1S), Cav1. 2 (CACNA1C), Cav1. 3 (CACNA1D) and T-type VDCC subunit Cav3. 3 (CACNA1I) were recently identified as risk loci for psychiatric disorders. We investigated rare mutations with possibly damaging effects in those genes in Japanese sample of SCZ and ASD. We prioritized four candidate genes (CACNA1C, CACNA1D, CACNA1S, CACNA1I) for psychiatric disorders in subset of VDCC genes based on genome-wide association studies, exome sequencing and functional genomic studies. Then, mutation screening of exon regions of those 4 genes using Ion Torrent Personal Genome Machine (PGM) was performed in a Japanese sample of 370 SCZ patients and 192 ASD patients. Our AmpliSeq custom panel allowed us to cover 96. 84% of the targeted sequences. Average coverage of depth in the target region was ~200, and ~80% of sequenced region was covered over 100x coverage. Variant call and annotation were performed with Torrent Suite 4. 4 and Ingenuity Variant Analysis. Variant filtering were applied to identify those that were not registered in dbSNP database or have a minor allele frequency of less than 1% in East-Asian samples from the exome sequencing project (ESP) and 1000 Genomes, and are damaging non-synonymous, splicing site single nucleotide variants (SNVs) or small insertion and deletion predicted by in-silico analyses. All of those filtered mutations were confirmed by Sanger method. If parental samples were available, segregation analysis were employed for measuring the inheritance pattern. Under our filter, we discovered 1 nonsense SNV (p. C1471* in CACNA1D), 1 in-frame deletion (p. E1675del in CACNA1D), 2 de novo SNVs (p. A36V in CACNA1C, p. V947I in CACNA1S), 22 missense SNVs (list in poster) that are categorized as damaging by at least two different in-silico tools. Our analysis investigated several rare and possibly damaging variants on the risk for SCZ and ASD in VDCC genes, especially within L-type VDCC genes. Ongoing work includes (1) genotyping of selected rare variants in additional cohorts and rare-variant association analysis (2) further functional assessment of possible disease-causing variants.

1120F

Rare bi-allelic mutations in autism spectrum disorder illuminate gender differences and highlight novel genes. *T. W. Yu[1], R. N. Doan[1], E. T. Lim[1], S. De Rubeis[2], C. Betancur[3], D. J. Cutler[4], C. S. Poultney[2], A. G. Chiocchetti[5], C. M. Freitag[5], S. Goetze[1], M. H. Chahrour[6], E. H. Cook[7], J. S. Sutcliffe[8], M. E. Zwick[4], A. P. Goldberg[2], B. Devlin[9], K. Roeder[10], M. J. Daly[11], C. A. Walsh[1], J. D. Buxbaum[2], The Autism Sequencing Consortium.* 1) Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA; 2) Department of Psychiatry and Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 3) INSERM U1130, CNRS UMR 8246, Université Pierre et Marie Curie, Paris, France; 4) Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia, USA; 5) Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, JW Goethe University Frankfurt, Frankfurt, Germany; 6) Eugene McDermott Center for Human Growth & Development, UT Southwestern, Dallas, Texas, USA; 7) Institute for Juvenile Research, Department of Psychiatry, University of Illinois at Chicago, Chicago, Illinois, USA; 8) Vanderbilt Genetics Institute, Vanderbilt University Medical Centre, Nashville, Tennessee, USA; 9) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA; 10) Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; 11) The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

Systematic large scale sequencing studies of patients with autism spectrum disorder (ASD) are beginning to reveal insights into the pathogenesis of this complex disorder, with evidence of clustering of genes implicated by rare inherited and *de novo* CNVs and SNVs in pathways for synapse development and chromatin remodeling. Human gene knockouts, caused by biallelic mutations disrupting both copies of a gene, provide another potential avenue to illuminating disease-relevant biological processes. Here we report the largest analysis to date of rare, inherited, biallelic mutations in whole exome data from over 3,000 cases and 6,000 controls assembled by the Autism Sequencing Consortium. Rare (AF<1%) compound heterozygous or homozygous loss of function mutations are observed in 1 in ~32 unaffected individuals in the ASC overall, but are found significantly more often in affected individuals (1 in ~23, a 37% elevation in rate vs. unaffecteds, p=0. 008). The observed biallelic excess is not explained by differences in underlying heterozygous mutation rates. Affected females exhibit the most pronounced excess (67% elevation over baseline, p=0. 02), consistent with a female protective effect similar to that previously observed for *de novo* SNVs and CNVs. Genes potentially implicated by biallelic inactivation in this analysis include those previously known to cause recessive intellectual disability/ASD as well as interesting novel candidates with diverse neurobiological roles.

1121W

Genome-wide *de novo* mutation landscape in Autism Spectrum Disorder. *R. K. C. Yuen[1], D. Merico[1], H. Cao[2], B. Alipanahi[3], B. Thiruvahindrapuram[1], G. Pellecchia[1], X. Tong[2], D. Cao[2], Y. Sun[2], M. Li[2], W. Chen[2], X. Jin[2], T. Nalpathamkalam[1], M. Bookman[4], J. Bingham[4], S. Gross[4], D. Loy[4], S. Walker[1], J. L. Howe[1], M. Pletcher[5], C. R. Marshall[1,6], P. Szatmari[7,8], D. Glazer[4], B. J. Frey[3,9], R. H. Ring[5], S. W. Scherer[1,10,11].* 1) The Centre for Applied Genomics, Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; 2) BGI-Shenzhen, Yantian, Shenzhen, China; 3) Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada; 4) Google, Mountain View, California, USA; 5) Autism Speaks, Princeton, New Jersey, USA; 6) Department of Molecular Genetics, Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 7) Child Youth and Family Services, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 8) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 9) Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada; 10) Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada; 11) McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada.

Autism spectrum disorder (ASD) is a collection of neurodevelopmental conditions characterized by deficits of social interaction and communication, and presence of restricted and repetitive behaviors. *De novo* mutations (DNMs) play an important role in the etiology of ASD, but analyses so far have focused mainly on the protein coding regions of DNA, which account for only ~1% of the genome. To characterize genome-wide DNMs, we performed whole genome sequencing (WGS) on 200 ASD simplex trios with a depth of ~30x per genome using the Illumina HiSeq technology (part of the MSSNG WGS project). Using our improved variant detection pipeline, we identified 55. 4 *de novo* single nucleotide variants (SNVs), 4. 2 *de novo* insertion/deletions (indels) and 0. 12 *de novo* copy number variation per genome. Consistent with previous reports, we found that the father contributes the majority of the DNMs (72% of *de novo* SNVs and 70% of *de novo* indels), and that they are positively correlated with the paternal age ($p<0.0001$). Despite the different sensitivity on *de novo* mutation detection, we found that sequence context of DNMs is similar between ASD cases and population controls (based on 258 Dutch genomes). For the genic DNMs, we found a significant enrichment of predicted damaging DNMs in the ASD cases compare to the controls (OR: 1. 98; $p<0.0001$), of which 6. 5% are intronic splicing variants (half of them located over 10bp away from the exons). We found particularly high burden in genes (i) with medium to high expression in brain (OR: 2. 56; $p=0.0011$), (ii) targets of FMRP (OR: 5. 08; $p=0.0033$), (iii) with a functional role in human nervous system development (OR: 2. 82; $p=0.02$), (iv) implicated in mouse nervous system abnormal phenotypes (OR: 2. 55; $p=0.017$). Beyond the genic region, we also found a higher DNM rate of sites predicted to reduce transcription factor binding affinity (OR 1. 25; $p=0.018$). Our results revealed a substantial contribution of non-coding variants to the etiology of ASD and emphasized the importance of using WGS for comprehensive genetic analyses.

1122T

The Alzheimer's Disease Sequencing Project (ADSP) Discovery Phase: Data Production, Management, and Availability. *A. B. Partch[1], A. L. Destefano[2], L. B. Cantwell[1], K. M. Faber[3], M. Feolo[4], A. Stine[4], N. Gupta[5], W. J. Salerno[6], D. C. Koboldt[7], O. Valladares[1], J. C. Bis[8], D. M. Childress[1], S. H. Choi[2], R. L. Cweibel[1], S. P. Dugan-Rocha[6], P. Gangadharan[1], S. D. Lewis[7], D. Reyes[9], E. A. Boerwinkle[6], G. D. Schellenberg[1], S. Seshadri[10], T. M. Foroud[3], L. S. Wang[1], Alzheimer's Disease Sequencing Project Data Flow Work Group.* 1) Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 2) School of Public Health, Boston University, Boston, MA; 3) Department of Medical and Molecular Genetics, Indiana University, Indianapolis, Indiana; 4) National Center for Biotechnology Information, National Institutes of health, Bethesda MD; 5) Broad Institute of Harvard and MIT, Cambridge, MA; 6) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 7) The Genome Institute, Washington Univ. , St. Louis, MO; 8) Cardiovascular Health Research Unit, University of Washington, Seattle, WA; 9) Taub Institute for Research on Alzheimer's Disease, Columbia University, New York; 10) Department of Neurology, Boston University School of Medicine, Boston, MA.

Statement of purpose. The Alzheimer's Disease Sequencing Project (ADSP) was established in February 2012 as a Presidential Initiative to fight Alzheimer's Disease (AD). Developed jointly by National Institute on Aging (NIA) and the National Human Genome Research Institute (NHGRI), the specific aims of the ADSP are to: 1) identify protective genomic variants in older adults at risk for AD, 2) identify new risk variants among AD cases, and 3) examine these factors in multi-ethnic populations to identify therapeutic targets for disease prevention. **Methods used.** The ADSP Data Flow Work Group and the NIA Genetics of Alzheimer's Disease Data Storage Site (NIAGADS) were charged with supporting ADSP data production, sharing and management, and facilitating data access by the general research community. Samples and phenotypes were contributed by the Alzheimer's Disease Genetics Consortium (ADGC) and Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. Sample plating and shipping was coordinated by the National Cell Repository for Alzheimer's Disease (NCRAD). Whole-exome and whole-genome sequencing data were generated by the three NHGRI Large Scale Sequencing Analysis Centers (LSACs). The Data Flow Work Group maintains the ADSP web site to provide study design, cohort information, news releases, and instructions to apply for access to ADSP data. The ADSP Data Portal was designed through collaboration with the Database of Genotypes and Phenotypes (dbGaP), Sequence Read Archive (SRA), and NIAGADS. The portal allows users to explore the ADSP project data archived at dbGaP with NIH iTrust user authentication. Approved investigators can then identify samples and download files using a customizable filtering system and check-out cart function. **Summary.** The ADSP has completed sequencing the exomes of 10,939 unrelated cases and controls and the whole genomes of 578 members of 111 AD families. Accompanying each sample is GWAS data and a set of phenotypes that were harmonized according to protocols generated by the ADSP. With the completion of sequencing for the Discovery phase, this work group provides support to all work groups for data related issues, coordinates with dbGaP for data transfers, and reviews, posts, and notifies ADSP members of new results generated by other work groups. More information can be found at the ADSP website (www. niagads. org/adsp) or at dbGaP (study ID: phs000572).

1123F

Identification of candidate genes in familial early-onset essential tremor: whole exome sequencing study. *L. Clark[1], X. Liu[1], N. Hernandez[2], S. Kisselev[1], A. Floratos[3,4], A. Sawle[3,4], I. Ionita-Laza[5], R. Ottman[6,7,8,9], E. Louis[2].* 1) Department of Pathology and Cell Biology, Columbia University, New York, NY; 2) Department of Neurology, Yale School of Medicine, Yale University, New Haven, CT; 3) Department of Biomedical Informatics, Columbia University, New York, NY; 4) Center for Computational Biology and Bioinformatics, Columbia University, New York, NY; 5) Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY; 6) G. H. Sergievsky Center, Columbia University, New York, NY; 7) Department of Neurology, College of Phyisicans and Surgeons, New York, NY; 8) Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY; 9) Division of Epidemiology, New York State Psychiatric Institute, New York, NY.

Background: Essential tremor (ET) is one of the most common causes of tremor in humans. Despite its high heritability and prevalence, few susceptibility genes for ET have been identified. **Methods:** To identify genes for ET, we enrolled probands (affected with ET) and relatives in a family study of ET at Columbia University, New York (2011 – 2014). The study was approved by the Institutional Review Board at Columbia University and written informed consent was obtained from all enrollees. Whole exome sequencing was performed on at least two most distantly related affected (definite, probable or possible ET diagnosis) individuals from each of 37 total families. Sequence alignment to the human reference genome (UCSC hg19) was performed using the Burrows-Wheeler Aligner (BWA) algorithm, and variant calling was performed using the genome analysis toolkit (GATK; Broad Institute). The software tool pVAAST (Pedigree Variant Annotation, Analysis and Search Tool) was used to analyze the whole exome data from ET families using a dominant model and the maximum number of permutations of 1,000,000). **Results:** We identified candidate genes in independent ET families that were ranked by pVAAST and that reached genome wide significance. In two independent families, we identified likely pathogenic mutations in the same gene that co-segregated with disease. Both mutations result in amino acid substitutions of highly conserved amino acid residues that are predicted to be deleterious and damaging by *in silico* analysis. Likely pathogenic mutations in three other genes were also identified in three independent families. These genes are highly expressed in the cerebellum and Purkinje cells, and in several different ways, influence function of the GABA-ergic system. **Conclusions:** Candidate genes identified in this study are in concordance with recent evidence that the pathophysiological process in ET is associated with cerebellar degeneration, a reduction in Purkinje cells, and a decrease in GABAergic tone. Future functional studies are needed to evaluate the pathogenicity of the variants identified in these candidate genes.

1124W

Rare susceptibility variants for familial bipolar disorder: a role for G protein-coupled receptors. *C. Cruceanu[1, 2], J. F. Schmouth[2], J. P. Lopez[1], D. Spiegelman[2], D. Rochefort[2], P. Hince[2], P. A. Dion[2], M. Alda[3], G. Turecki[1], G. A. Rouleau[2].* 1) McGill University, Montreal, Quebec, Canada; 2) Montreal Neurological Institute, Montreal, Quebec, Canada; 3) Department of Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada.

Bipolar disorder (BD) is a complex psychiatric condition characterized by both manic and depressive episodes. Previous studies strongly support the role of genetics in BD, but likely due to genetic and phenotypic heterogeneity, few loci have been replicated across different studies and cohorts. We hypothesize that BD is caused by highly penetrant rare variants in many different genes across the population, which can be unveiled by through massively-parallel sequencing in well-characterized multiplex families. Thus, we re-sequenced the exomes of all affected individuals from 40 multi-generational families (3-8 individuals per family across 2-4 generations). To identify relevant BD susceptibility genes we prioritized rare variants segregating with affected status. In each family we prioritized potentially highly penetrant (e. g. protein-truncating, missense, or frameshift) or functionally relevant (e. g. 3'UTR, 5'UTR, splicing) variants. The variants that emerged from this analysis are in genes involved in brain development and neurogenesis, inflammation, and epigenetic regulation – all processes that have been suggested to be dysregulated in mood disorders including BD. The most interesting finding that emerged was an enrichment of putatively causal variants in genes belonging to the G protein-coupled receptor family, which are important drug targets and have previously been connected to psychiatric pathology. Furthermore, we followed up on the functional implications of some of the most deleterious mutations and showed targeted downstream GPCR dysregulation that could explain pathology. By focusing on rare variants in a familial cohort we have explained part of the missing BD heritability. In addition, we have narrowed in on some of the key biochemical pathways that are implicated in this complex condition.

1125T

Exome sequencing in multiplex synaesthesia families identifies rare coding variants in genes within candidate linkage regions. *K. S. Kucera[1], S. A. Graham[1], A. Vino[1], J. E. Asher[2], S. Baron-Cohen[2], S. E. Fisher[1,3].* 1) Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; 2) Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, UK; 3) Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, The Netherlands.

Synaesthesia is a rare neurodevelopmental non-disease condition estimated to affect <2% of the human population. People with synaesthesia experience cross-modal perceptions, such as visual images triggered by sounds, words that elicit tastes, or involuntary and stable associations of letters and/or numbers with colours. It has been hypothesized that synaesthesia is a consequence of relative hyperconnectivity in the brain. The well-documented familial clustering of synaesthesia suggests that genes play a significant role. However, results of previous linkage studies indicate that synaesthesia is likely to be genetically heterogeneous; thus, a similar phenotype may result from different combinations of contributing genes and their alleles in different cases. In the current study, we used next-generation sequencing to investigate families in which multiple relatives have synaesthesia. We sequenced exomes of people with and without synaesthesia in three such families (13 synaesthetes and 4 non-synaesthetes in total), to detect rare genetic variants that segregate with the trait. We thereby identified 77 rare (MAF<0. 05) nonsynonymous protein-coding variants that segregate with the phenotype in those families. Of the 77 variants, 10 were located within previously reported suggestive linkage regions on chromosomes 2q13 and 4q28. 3. Future investigations will use data from gene expression and protein function in cells and tissues to determine which of these candidate variants are most likely to have roles in aetiology. Furthermore, common variation in the same candidate genes can be tested for association in independent case-control datasets, to complement the family-based studies described here. Identification of genes implicated in synaesthesia will not only shed light on how this intriguing trait arises, but also promises novel insight into the molecular bases of neural connectivity in the human developing brain.

1126F

***PCDH19*-related Epileptic Encephalopathy: An unusual case of an affected male with mosaic *de novo* truncating variant in *PCDH19* gene.** *I. Thiffault[1,2,4], L. Smith[1,4,5], J. Lowry[3,4,5], L. Zellmer[1], E. Farrow[1], B. Black[5], C. Baldwin[5], N. Miller[1], S. Soden[1,4,5], C. Saunders[1,2,4].* 1) Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO; 2) Department of Pathology and Laboratory Medicine, Children's Mercy Hospitals, Kansas City, MO; 3) Department of Pediatrics, Children's Mercy Hospitals, Kansas City, MO; 4) University of Missouri-Kansas City School of Medicine, Kansas City, MO; 5) Individualized Pediatric Therapeutic Medicine Clinic, Children's Mercy Hospital, Kansas City, MO.

PCDH19 encodes protocadherin 19 and is located on the X chromosome. Heterozygous variants in *PCDH19* are associated with early infantile epileptic encephalopathy-9 (EIEE), also known as epilepsy and mental retardation restricted to females. Although autosomal dominant inheritance was tentatively suggested in several familial reports, *de novo* variants in *PCDH19*, were shown to be a frequent cause of sporadic EIEE in female. The syndrome is characterized by early normal development followed by febrile and temperature-induced seizures that tend to occur in clusters and may resemble Dravet syndrome. Hemizygous males are usually unaffected except for psychiatric/behavioral abnormalities. **Case presentation:** Our patient, CMH980, is a 6 year old male with focal epilepsy of unknown etiology that began at 9 months of age. In addition to seizures, by 3 years of age he developed behavioral challenges including irritability, aggression, rigidity, and poor sleep. Response to psychotropic medications has been inconsistent. His seizure frequency has decreased with age, and he is on stable doses of zonisamide and pyridoxine. Dysmorphic features noted included plagiocephaly in occipital/parietal area, agouti hair coloring, long posterior hair swirl, normal set ears but cupping of bilateral ears, intra digital webbing of phalanges, stereotypic movements of bilateral hands. Previous negative workup included *SCN1A*, karyotype, CGH-array, brain MRI and metabolic studies. Exome sequencing was performed on CMH980 and his healthy mother. CMH980 appeared to be heterozygous in *PCDH19* for a novel pathogenic nonsense variant, c. 605C>A (p. Ser202*), inconsistent with expectations for a male. This observation is confined to this single variant, ruling out an X chromosome dosage abnormality. This child's mother is negative for the variant; father was unavailable for testing. **Conclusion:** This is the second male reported with somatic mosaicism for *PCDH19*-deficency, the first being a large deletion detected by CGH-array. In mosaic males, as in affected female, some cells express PCDH19 and others do not. The loss of function at the cellular level results in a gain of function at the tissue level because of the abnormal interactions between cells expressing PCDH19 and cells lacking PCDH19. These observations further support cellular interference as the pathogenic mechanism for this condition, which leads to this unusual mode of inheritance in which females are more frequently affected than males.

1127W

Young Northern Finnish founder population reveals enrichment of rare recessive and dominant gene variants in neurodevelopmental disorders. *M. I. Kurki[1,2], O. Pietiläinen[3], E. Saarentaus[4], E. Hämäläinen[4], J. S. Moilanen[5], O. Kuismin[5], M. Daly[6], A. Palotie[1,2,4], Sequencing Initiative Suomi consortium.* 1) Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Stanley Center, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 3) Department of Stem Cell and Regenerative Biology, University of Harvard, Cambridge, Massachusetts, USA; 4) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 5) Department of Clinical Genetics, Oulu University Hospital, Medical Research Center Oulu and PEDEGO Research Group, University of Oulu, Oulu, Finland; 6) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA.

Genetic variants with strong reproductive disadvantages are evolutionary constrained and remain generally rare in population. However, these variants can still exist at higher frequencies in young populations, such as Finns, when the negative selection hasn't had time to counteract the effect of genetic drift on rare alleles. Thus, population isolates provide a valuable study design to explore the role of rare genetic variants in complex traits. In Finland, the youngest settlement is in the north and east parts of the country dating back to a small number of founder families only few centuries ago. In addition this region has higher prevalence of schizophrenia and intellectual disability (ID). We exploited this hypothesis by producing whole exome sequence (WES) and GWAS data from 352 patients from Northern Finland with ICD-10 diagnosis of ID of unknown etiology, and their 293 family members (97 trios, 109 duos and 146 index cases). The Northern Finland Intellectual Disability Project (NFID) exomes were combined with 8000 Finnish exomes sequenced in the Sequencing Intitative Suomi project (SISu, http://sisuproject. fi/). As expected, we observed comparable amount of large CNVs and *de novo* mutations as reported in similar patient collections, both of these categories being enriched in the NFID patients. Given the genetic origin of NFID, we expected to observe variants enriched in Finland that are 1) strong acting recessive variants that seem Mendelian but account for ~1% of a 1% phenotype rather than all of a 1/10000 phenotype and 2) dominant alleles with odds ratios in the range of 2-5. As per our hypothesis we discovered a Finnish-specific recessive cause of ID in 4 cases, homozygosity of a variant in CRADD (p=4e-8). The variant is not observed in homozygous state in 61 000 individuals worldwide (http://exac. broadinstitute. org/) or in 8000 Finnish individuals. We also observed Finnish enriched dominant missense variants in multiple genes (OR range 3-6) including a gene encoding for *TUBA1A1* (OR 5. 2, p:4e-8). Significant and promising variants are replicated by sequencing additional Northern Finnish ID cases and their family members (n=315; 150 cases; 51 trios). In conclusion, we demonstrate young founder populations as a powerful resource to study rare variants. Specifically, we show that an enrichment of deleterious alleles increases power to detect causal and disease associated variants that would require very large sample sizes in more diverse populations.

1128T

Exome and genome sequencing in Irish multiplex schizophrenia families. *B. P. Riley[1,2], B. T. Webb[1], B. Verrelli[3], S. A. Bacanu[1], M. Fromer[4], P. F. Sullivan[5,6], A. P. Corvin[7], K. S. Kendler[1,2], Irish Schizophrenia Genomics Consortium.* 1) Department of Psychiatry, Virginia Commonwealth University, Richmond, VA; 2) Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond, VA; 3) Department of Biology , Virginia Commonwealth University, Richmond, VA; 4) Department of Psychiatry, Icahn School of Medicine at Mt Sinai, New York, NY; 5) Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden; 6) Departments of Genetics and Psychiatry, University of North Carolina, Chapel Hill, NC; 7) Department of Psychiatry, Trinity College Dublin, Dublin, Ireland.

Affected members of multiplex schizophrenia (SCH) pedigrees have elevated recurrence risk compared to singleton cases, but mean polygenic risk scores between familial and singleton cases in our Irish samples do not differ. This suggess that higher familial recurrence risk is due in part to rare, higher impact variation. We are undertaking sequencing studies to identify such variation based on 3 hypotheses: 1) The study of multiplex pedigrees enriches the data for SCH risk variants; variants identified IBD in multiple families and those substantially more frequent in the families compared to population controls are of special interest. 2) Because of the population specificity of rare alleles, power is improved by studying familial and singleton SCH cases and controls from the same population; we have assembled 270 multiplex pedigrees, 3000 singleton cases and 3000 controls from Ireland, and 3781 additional UK controls. 3) We can include genomic information in analysis to improve signal detection, particularly important outside the exome; we annotate the genome with empirically-defined weights from GCTA based on data from GENCODE, ENCODE, base conservation and other sources to account for the differential prior probability that variation at a genomic position has functional consequence. In 76 pilot exomes, we identified a total of 106 variants in 65 unique genes with weights>1. Calcium channel (35-fold), cholinergic receptor (14-fold) and autism de novo variant (3. 4-fold) genes were all enriched for weighted variants. Most notably, the D112G missense change (rs137861662) in the N(alpha)-acetyltransferase 16 (*NAA16*) gene (site of de novo CNV in SCH) is predicted to be deleterious and is 19 times as frequent in familial cases as in controls. Further evolutionary analysis shows that conservation at this site is significant across >400 My of divergence (phyloP scores, 100-vertebrate: 8. 45; 46-vertebrate: 4. 49; mammalian: 2. 06; primate: 0. 52). A single ~20Kbp window centered on rs137861662 also exhibits significant phylogenetic conservation (p<0. 01). Haplotype-based modeling of this region in HapMap3 populations shows low haplotype diversity and an enrichment of rare alleles consistent with strong purifying selection seen in our deep time analyses. We are now applying this design to genome sequencing in order to identify additional variants for direct assessment in Irish singleton cases, Irish controls and UK controls.

1129F

The Diverse Landscape of Structural Variation Contributing to Autism. *R. L. Collins[1], H. Brand[1,2], C. Hanscom[1], J. T. Glessner[1,2], A. Stortchevoi[1], M. R. Stone[1], C. E. Redin[1,2], S. Eggert[1], V. Pillalamarri[1], S. Erdin[1], A. Ragavendran[1], C. W. Seabra[1,3], F. Kelley[4], T. Mason[4], L. Margolin[4], M. E. Talkowski[1,2,4].* 1) Center for Human Genetic Research, Massachussets General Hospital, Boston, MA; 2) Department of Neurology, Harvard Medical School, Boston, MA; 3) GABBA Program, University of Porto, Porto, Portugal; 4) Program in Medical and Population Genetics and Genomics Platform, Broad Institute, Cambridge, MA.

The genetic etiology of autism spectrum disorder (ASD) is complex. Recent studies suggest that alterations in gene dosage from *de novo* loss-of-function (dnLoF) mutations are among the most penetrant mutations, however these studies have only considered single nucleotide coding changes, indels, and large copy number variants. To date, the contribution of balanced and complex structural variation (SV) to ASD is largely unknown as these variants are cryptic to whole-exome sequencing (WES) and chromosomal microarray (CMA). To explore the landscape of cryptic SVs in ASD, We performed large-insert whole-genome sequencing (liWGS) of SV to deep physical coverage of mapped inserts (median = 97x) on 400 idiopathic ASD cases from the Simons Simplex Collection. Based on completed analyses (235 probands), we observed an average of 8. 26 Mb of the genome to be rearranged, representing at least 32 predicted LoF mutations per proband at the resolution of liWGS (~5 kb). We uncovered a diverse spectrum of SV, including complex variants such as germline chromothripsis involving 11 breakpoints and six chromosomes, as well as a new class of surprisingly common SV (8. 1% frequency in cohort) involving an inversion flanked by 'paired-duplications,' which we dub "dupINVdup. " Based on analyses to date, we estimate that at least 10. 6% of probands harbor a *de novo* SV that perturbs a locus of potential relevance to ASD, or approximately double the yield of CMA analyses alone in this cohort. Notably, we discovered dnLoF mutations that were cryptic to CMA and disrupted known ASD loci such as *NRXN1* and *IL1RAPL1*, as well as genes from which WES identified a single dnLoF in this cohort, which we now confirm via SV studies. We also identified putative novel ASD loci, such as a *de novo* 23 Mb inversion that disrupted *RPTOR*, a gene under strong evolutionary constraint that is associated with the mTOR pathway. Conversely, we observe probands with LoF SV inherited from a healthy parent that disrupt genes previously annotated as ASD associated loci based on multiple dnLoF from WES, suggesting that LoF mutations in these genes are likely to be incompletely penetrant. These results suggest that the yield of significant SVs in the SSC is likely to be about two-fold higher than is predictable by CMA alone, suggesting that cryptic and complex SVs represent an important and currently uncharacterized component of ASD etiology, with likely implications for other psychiatric disorders.

1130W

Structural variation in 111 families at risk for Alzheimer's disease from heterogenous whole-genome sequencing data. *N. Zhang[1], W. Salerno[2], J. Farrell[3], J. Malamon[1], L. Xia[4], A. Destefano[3], A. Naj[1], B. Vardarajan[5], A. Partch[1], D. Kobolt[6], D. Larson[6], W. Wang[1], L. Wang[1], K. Worley[2], L. Farrer[3], G. Schellenberg[1], Alzheimer's Disease Sequencing Project Structural Variant Work Group.* 1) University of Pennsylvania; 2) Baylor College of Medicine; 3) Boston University; 4) Stanford University; 5) Columbia University; 6) Washington University at St. Louis.

Determining the genetic foundations of common complex disease will increasingly depend on our ability to provide comprehensive analysis of genomic data pooled from multiple sources. As part of the Alzheimer's Disease Sequencing Project, we have characterized the structural variation in 111 families (67 Hispanic, 41 Caucasian, 1 African American and 2 Dutch Isolate) comprising 578 individuals diagnosed with or at risk for Alzheimer's Disease. Deep whole-genome sequencing (~30x) of these individuals was performed at three U. S. Sequencing and Analysis Centers and then analyzed via a unified structural variant (SV) workflow for quality control, discovery, genotyping, and annotation. Our Phase I SVs comprise a high-confidence set of prioritized deletions, insertions, and complex variants larger than 20 base pairs genotyped across all 578 individuals. These variants are prioritized based on overlap with linkage regions calculated from extended pedigrees, and compared with known structural events associated with Alzheimer's disease and linked to other common complex diseases. SV detection across the three sequencing centers requires rigorous statistical procedures for comparing and aggregating call sets across heterogeneous insert libraries and mapping pipelines. We developed a statistical framework that leverages pedigree information to assess and optimize call accuracy. While typically genotypes are used to check pedigree relationships with the kinship coefficient, we reversed the process and used the kinship coefficient to check SV genotyping. For SV programs that generate discovery sites and not genotypes, we developed the D-score which is an outgroup-based measure of sib-sharing. These two metrics take advantage of pedigree information to permit the accurate assessment of the strength of SV callers to detect structural variants of varying size ranges and types. We also developed an in silico procedure to spike-in SV of varying purity into real sequenced libraries, accommodating multi-library designs, and applied it to 3 samples that were sequenced at all three centers. This comprehensive QC process allowed us to benchmark mainstream SV callers across widely differing libraries, and to combine these callers in a library-specific way for improved specificity and sensitivity. Finally, we refine the candidate SV regions detected by an ensemble of SV callers with local assembly to provide precise breakpoints for follow-up genotyping.

1131T

Exome-based analysis of cardiac arrhythmia, respiratory control and epilepsy genes in sudden unexpected death in epilepsy. *R. D. Bagnall[1, 2], D. E. Crompton[3, 4], S. Petrovski[4, 5], L. Lam[1], C. Cutmore[1], S. I. Garry[4], L. G. Sadlier[6], L. M. Dibbens[7], A. Cairns[8], Z. Afawi[9], B. M. Regan[4], J. Duflou[2, 10], S. F. Berkovic[4], I. E. Scheffer[4, 11, 12, 13], C. Semsarian[1, 2, 14].* 1) Agnes Ginges Centre for Molecular Cardiology, Centenary Institute, Sydney, NSW, Australia; 2) Sydney Medical School, University of Sydney, Sydney, Australia; 3) Neurology Department, Northern Health, Melbourne, Australia; 4) Epilepsy Research Centre, Department of Medicine, University of Melbourne, Austin Health, Melbourne, Australia; 5) Institute for Genomic Medicine, Columbia University, New York, United States of America; 6) Department of Paediatrics and Child Health, School of Medicine and Health Sciences, University of Otago, Wellington, New Zealand; 7) Epilepsy Research Program, School of Pharmacy and Medical Sciences, University of South Australia, Adelaide, SA, Australia; 8) Neurosciences Department, Lady Cilento Children's Hospital, Brisbane, Australia; 9) Tel-Aviv Sourasky Medical Center, Tel Aviv, Israel; 10) Department of Forensic Medicine, Sydney, Australia; 11) Department of Neurology, The Royal Children's Hospital, Parkville, Melbourne, Victoria, Australia; 12) Florey Institute of Neurosciences and Mental Health, Melbourne, Australia; 13) Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Melbourne, Australia; 14) Department of Cardiology, Royal Prince Alfred Hospital, Sydney, Australia.

The leading cause of epilepsy-related premature mortality is sudden unexpected death in epilepsy (SUDEP). The cause of SUDEP remains unknown. We performed an exome-based analysis of rare variants to search for genetic risk factors in SUDEP cases, including a focused analysis of cardiac arrhythmia, respiratory control, and epilepsy genes. Demographic and clinical information of 62 SUDEP cases were collected. Exome sequencing and rare variant collapsing analysis with 2,936 control exomes were performed to test for genes enriched with damaging variants. Cardiac arrhythmia, respiratory control and epilepsy genes were screened for variants with a general population frequency of <0. 1% and predicted to be pathogenic with multiple *in silico* tools. The 62 SUDEP cases were categorized as 'definite SUDEP' (n=55), 'probable SUDEP' (n=5), and 'definite SUDEP plus' (n=2). The mean age at epilepsy onset was 10. 5 ± 8. 2 years (range 0-34 years) and the mean age at SUDEP was 28 ± 11. 9 years (range 1-53 years). SUDEP occurred in bed in 41 (66%) cases, with the position noted as prone in 27 and supine in two. We identified *de novo* mutations, previously reported pathogenic mutations, or candidate pathogenic variants in 29/62 (47%) cases. Six SUDEP cases (10%) had mutations in common genes responsible for the inherited cardiac arrhythmia disease, long QT syndrome (LQTS). Eight cases (13%) had candidate pathogenic variants in dominant cardiac arrhythmia genes. Fifteen cases (24%) had six mutations and ten candidate pathogenic variants in epilepsy genes associated with dominant disorders. No gene reached genome-wide significance with rare variant collapsing analysis; however, among the top 30 genes, genome-wide, were *DEPDC5* (p=0. 00016), which causes the focal epilepsy with variable foci syndrome, *KCNH2* (p=0. 0039), which causes LQTS type 2, and *NOS1AP* (p=0. 00023), which has common variants that are associated with sudden cardiac death. A sizeable proportion of SUDEP cases have clinically relevant mutations in cardiac arrhythmia and epilepsy genes. SUDEP may occur due to a predictable and preventable cause in cases with a LQTS gene mutation. These findings have relevance to understanding the causes of SUDEP, and raise the question whether specific cardiac arrhythmia or epilepsy genes confer an increased SUDEP risk. Understanding the genetic basis of SUDEP may inform cascade testing of at-risk family members.

1132F

***LGI2* heterozygous variant identified in a Japanese family with autosomal dominant cryptogenic West syndrome.** *N. Ishihara[1,2], Y. Azuma[2,3], T. Nakata[2,3], T. Takeuchi[2,3], T. Okuno[3], K. Ohno[3], J. Natsume[2].* 1) Dept Pediatrics, Fujita Health University School of Medicine, Toyoake, Japan; 2) Dept Pediatrics, Nagoya University Graduate School of Medicine, Nagoya, Japan; 3) Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan.

The purpose of this study was to identify the causal gene of familial West syndrome. West syndrome is an epileptic encephalopathy characterized by infantile onset epileptic spasms in cluster, developmental regression, and hypsarhythmia in interictal EEG. Some of the patients have no neurological deficit before onset of the epileptic seizure. In this study, we performed genetic analysis of a familial case of West syndrome which has no neurological deficit before the onset of the seizures, and no sequela after the treatment. Subjects: Proband is a 4-year-old girl who had epileptic spasms in cluster at the age of 4 months. She did not have any neurological complication before the onset of the seizure. Her brain MRI showed no abnormal findings, and blood test had no evidence of congenital metabolic disorder. Her interictal EEG showed hypsarhythmia, therefore she was diagnosed as cryptogenic West syndrome. She received ACTH therapy, and epileptic spasms were resolved in 3 days after starting the therapy. Her mother was also suffered from West syndrome during her infantile period, which resolved with ACTH therapy. Both the proband and her mother had no major neurological sequela, except cognitive function of the proband revealed subnormal. Methods: Whole exome sequence of the proband and her family was performed. We analysed mutations in common between the proband and her mother. Results: Analysis led to the identification of five candidate genes, and we found one of these genes had reported as causal for familial epilepsy in childhood with dogs. Leucine-rich gene glioma-inactivated 2 (*LGI2*) consisted of 545-amino acid protein contains an N-terminal signal peptide, a putative transmembrane region, and 1 partial and 4 complete leucine-rich repeats. Conclusions: There is no previous report of *LGI2* in human epilepsy syndromes, however we suggest that it may be one of the causal genes of familial West syndrome.

1133W

Evaluation of variants from epilepsy patients referred for genetic screening. *K. A. McCarty, C. da Silva, J. J. Alexander, M. Hegde, A. Escayg.* Dept of Human Genetics, Emory University, Atlanta, GA.

Epilepsy is a neurological disorder characterized by recurrent, unprovoked seizures caused by neuronal synchrony and hyperexcitability. More than a hundred genes causative of epilepsy have been identified to date. Despite this progress, a significant fraction of the genetics underlying epilepsy remains uncharacterized, and the mechanisms by which identified mutations lead to seizure generation remain elusive. Additionally, the time and financial costs associated with patient recruitment and genetic testing still create a barrier to epilepsy gene discovery. The Emory Genetics Laboratory (EGL) has developed a targeted sequencing library comprising 109 known and candidate epilepsy genes referred to as the 'epilepsy and seizure disorder panel' (ESDP). The ESDP is pulled from a larger mendeliome library of evidence-based disease genes, making this a valuable resource for identifying putative disease-causing alleles as well as new disease associations. Individuals are clinically referred to the EGL for genetic testing and are typically affected with severe childhood epilepsy. When testing is ordered, the coding exons of the 109 genes are sequenced and examined for the presence of variants; however, clinical laboratories identify variants which they are unable to classify as pathogenic without functional studies. We are utilizing this rich source of available sequence data to identify new disease-causing variants. This information will provide insight into the importance of specific residues and domains across the encoded proteins and enable pathogenic variants to be more reliably identified in the future. From the examination of 96 cases, we identified three novel variants in the voltage-gated sodium channel *SCN8A*, which is associated with early infantile epileptic encephalopathy and intellectual disability. Two of the variants occurred *de novo*, consistent with previously published mutations, while the third was maternally inherited. Studies are underway to confirm pathogenicity and investigate the functional consequences of these variants. Missense changes at conserved positions in *SCN8A* were observed infrequently in individuals referred for genetic screening on panels other than the ESDP. In addition to the *SCN8A* variants, we identified new pathogenic or likely pathogenic variants in 15 other genes from the ESDP, including *SCN1A*, *SCN2A*, *KCNQ2*, and *HCN1*, accounting for approximately 30% of the cases screened on the ESDP.

1134T

Identification of non-coding RNAs related to familial mesial temporal lobe epilepsy by whole exome sequencing and Sanger validation. *R. Secolin[1], P. A. O. Ribeiro[1], F. R. Torres[1], M. G. Borges[1], A. C. Coan[2], M. E. Morita[2], C. V. Soler[1], M. L. Santos[1], C. V. Maurer-Morelli[1], B. S. Carvalho[1], F. Cendes[2], I. Lopes-Cendes[1].* 1) Department of Medical Genetics, University of Campinas - UNICAMP, Campinas, SP, Brazil; 2) Department of Neurology, University of Campinas, UNICAMP, Campinas, SP, Brazil.

Background: Epilepsy is a common chronic neurological disorder that affects approximately 1% of the population worldwide. Familial mesial temporal lobe epilepsy (FMTLE) is a clinically well characterized syndrome with an autosomal dominant inheritance. Previous linkage studies identified a risk haplotype on chromosome (chr.) 18p11. 31 in one family segregating MTLE. Therefore, the identification of genetic variants within the candidate region could unveil the molecular basis of FMTLE. **Objective:**To identify variants within the candidate region 18p11. 31 in the family segregating MTLE. **Methods:** We evaluated four individuals, including three patients and one unaffected relative. We perform the whole exome sequencing (WES) using Nextera Rapid Capture Expanded Exome kit and Hiseq Illumina 2500 equipment (Illumina™). Bioinformatics analysis was performed using the GATK software package. We aligned the reads by BWA algorithm, using the Genome Reference Consortium Human Build 38 (GRCh38. p2). We annotated the variants using VariantAnnotator (GATK package). We included additional variant information, such as gene location, functional consequence and global minimum allele frequency, using the Variant Effect Predictor (VEP). We refine our search on chr. 18p11. 31. Variants found were validated by Sanger sequencing in additional 21 family members, including 14 patients sharing the affected haplotype on chr. 18p11. 31. **Results:** WES revealed a total of 272955 variants. Among them, we found three variants (rs75897598, rs79570056 and rs76231122) on chr. 18p11. 31. Further Sanger sequencing validation showed that all 14 haplotype-positive individuals share the rs75897598 and rs76231122 variants. In addition, all four individuals previously sequenced by WES presented the same genotype by Sanger sequencing. **Conclusion:** Among haplotype-positive individuals, eight have the diagnoses of FMTLE, two have had only a single seizure, two have the diagnoses of generalized epilepsy and two are asymptomatic relatives. However, all of them presented hippocampal abnormalities, including hippocampal atrophy and/or abnormal shape or axis identified on magnetic resonance imaging. The two validate variants are located in two long intergenic non-protein coding RNA (*LINC00526* and *LINC00667*). The lncRNAs were recently found to be related to several neurological diseases. Therefore, these two lncRNAS could be involved in the etiology of FMTLE. Supported by CEPID-BRAINN FAPESP, São Paulo, Brazil.

1135F

SORL1 mutations and Parkinsonian features in early onset Alzheimer's disease families. *M. L. Cuccaro[1, 2], R. M. Carney[3], B. W. Kunkle[2], J. M. Vance[1,2], P. Whitehead[2], J. R. Gilbert[1,2], B. N Vardarajan[4], R. Mayeux[4], M. A. Pericak-Vance[1,2].* 1) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 2) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 3) Department of Psychiatry & Behavioral Sciences, University of Miami Miller School of Medicine, Miami, Fl; 4) Department of Neurology, Taub Institute of Research on Alzheimer's Disease, Columbia University Medical Center, New York, NY.

Early-onset Alzheimer's disease (EOAD; onset of Alzheimer's disease (AD) prior to age 65) affects ~200,000 individuals, representing 3. 8% of all AD cases. Mutations in *PSEN1*, *PSEN2*, and *APP* can cause EOAD. We conducted a genome-wide search to identify novel mutations in a series of families enriched for EOAD using whole-exome sequencing (WES). We hypothesized that WES would identify novel mutations in these families and that these mutations would be associated with a spectrum of neurodegenerative features. Our dataset consisted of 50 families with at least one individual with EOAD. All individuals and additional family members were characterized via comprehensive medical and research record review. WES (Agilent SureSelect) was performed on individuals of interest per family; *APOE* genotypes were available as well. WES identified variants of interest in *SORL1*, a gene which previously had been primarily associated with late-onset AD. *SORL1*missensemutations were identified in 2 families. One family contained 4 affected individuals (AOO range 59-82 years; 3 with APOE genotype 3/4, 1 with 3/3) and 2 unaffected individuals (ages 81-84, both APOE 3/3) with the novel T588I*SORL1*mutation. Two individuals with *SORL1* mutations in this family also presented with Parkinsonian features. The second family carried thepreviously reported *SORL1*mutation (T749M) with 3 affected individuals with the mutation (AOO 55-84, all APOE 3/3), 1 affected individual without the mutation (AOO 76, APOE 3/4), and 1 unaffected individual with the mutation (age 79, APOE 3/3). One affected individual with in the second family with the T749M *SORL1* mutation had neuropathologic evidence of Lewy bodies without clinical Parkinsonism. A follow-up of SORL1 mutations in a separate dataset of LOAD families (Vardarajan 2015, Annals of Neurology) corroborated the presence of Parkinsonism in 4 individuals with *SORL1* mutations (3 individuals with the common rs2298813 mutation and 1 individual with a rare mutation). This study confirms *SORL1* as a gene involved in risk for EOAD. Equally compelling is our finding of Parkinson-related features and Lewy bodies without clinical Parkinsonism in association with these *SORL1* mutations. This is the first study to suggest a relationship between *SORL1*, EOAD, and Parkinsonism.

1136W

Analysis of linkage and whole exome sequencing to identify novel alcohol dependence-associated variants in extended European and African American families. *M. Kapoor[1], S. Bertelsen[1], L. Wetherill[2], JN. McClintick[2], J. Wang[1], H. Edenberg[2], T. Foroud[2], A. Goate[1], COGA Collaborators.* 1) Icahn School of Medicine at Mount Sinai, New York, NY, 10029, USA; 2) Indiana University School of Medicine, Indianapolis, IN, USA.

Genes implicated in alcohol dependence (AD) risk via linkage, candidate gene and GWAS account for only a small fraction of the overall risk, with effects varying across populations. To identify rare, high penetrance risk alleles we employed whole exome sequencing in systematically ascertained and deeply phenotyped extended families from the Collaborative Study on the Genetics of Alcoholism (COGA). We performed non-parametric linkage analysis in 2322 subjects from 118 extended European American (EA) families and observed strong linkage (lod = 2. 7) for AD symptom count on chromosome 3q24-26. Further, linkage analysis in each pedigree identified 4 families, contributing largely to the chromosome 3-linkage signal with a combined lod score of 5. 5. Twelve subjects (founder parent child trio) from these 4 families were selected for whole exome sequencing. Identity By Descent (IBD) estimates for non-sequenced subjects were generated through dense genome-wide SNP data available for each subject in the family. Sequencing and IBD based variant filtering around 3q24-26, identified 11 genes where novel functional variants segregated with AD in at least two families. We further examined the identified genes in exome array data from nearly 3000 independent EA samples and validated association for 4/11 genes, exhibiting significantly enrichment for rare variants. The most-significant association with AD was observed for WD repeat domain 52 (*WDR52*) gene (P = 8 x 10-4). Further, an improved association signal was observed (P = 3. 34 x 10-4) in meta-analysis of exome array data from EAs and nearly 1000 African American subjects. Linkage analysis in African American families in COGA resulted in a lod score of 2. 7 in the same region of chromosome 3, with AD symptom count. We obtained whole exome sequence for 40 subjects from these 10 AA families contributing to this lod score. We performed the segregation analysis using the pVAAST and prioritize several genes with functional variants. Two genes (*PPP2R3A* and *SLC41A3*) prioritized in the current analysis also overlapped with the top signals from the European American analysis. The *PPP2R3A* gene was also found to be significantly enriched for low frequency variants in the AD subjects in exome array analysis. Our initial findings suggest that multiple rare/novel variants in the same or different genes under the chromosome 3, linkage peak might be associated with AD.

1137T

Genomic analysis identifies candidate pathogenic variants in 9 of 18 patients with unexplained West syndrome. *A. Kikuchi[1], N. Hino-Fukuyo[1,2], N. Arai-Ichinoi[1], T. Niihori[2], R. Sato[1], T. Suzuki[1], H. Kudo[1], Y. Sato[1], T. Nakayama[1], Y. Kakisaka[1,3], Y. Kubota[1], T. Kobayashi[1,4], R. Funayama[5], K. Nakayama[5], M. Uematsu[1], Y. Aoki[2], K. Haginoya[6], S. Kure[1].* 1) Department of Pediatrics, Tohoku University School of Medicine, Sendai, Japan; 2) Department of Medical Genetics, Tohoku University School of Medicine, Sendai, Japan; 3) Department of Epileptology, Tohoku University School of Medicine, Sendai, Japan; 4) Division of Genomic Medicine Support and Genetic Counseling, Department of Education and Training, Tohoku Medical Megabank Organization (ToMMo), Tohoku University, Sendai, Japan; 5) Division of Cell Proliferation, United Centers for Advanced Research and Translational Medicine, Tohoku University Graduate School of Medicine, Sendai, Japan; 6) Department of Pediatric Neurology, Takuto Rehabilitation Center for Children, Sendai, Japan.

West syndrome, which is narrowly defined as infantile spasms that occur in clusters and hypsarrhythmia on EEG, is the most common early-onset epileptic encephalopathy (EOEE). Patients with West syndrome may have clear etiologies, including perinatal events, infections, gross chromosomal abnormalities, or cases followed by other EOEEs. However, the genetic etiology of most cases of West syndrome remains unexplained. DNA from 18 patients with unexplained West syndrome was subjected to microarray-based comparative genomic hybridization (array CGH), followed by trio-based whole-exome sequencing in 14 unsolved families. We identified candidate pathogenic variants in 50% of the patients (n = 9/18). The array CGH revealed candidate pathogenic copy number variations in four cases (22%, 4/18), including an Xq28 duplication, a 16p11. 2 deletion, a 16p13. 1 deletion and a 19p13. 2 deletion disrupting *CACNA1A*. Whole-exome sequencing identified candidate mutations in known epilepsy genes in five cases (36%, 5/14). Three candidate de novo mutations were identified in three cases, with two mutations occurring in two new candidate genes (*NR2F1* and *CACNA2D1*) (21%, 3/14). Hemizygous candidate mutations in *ALG13* and *BRWD3* were identified in the other two cases (14%, 2/14). Evaluating a panel of 67 known EOEE genes failed to identify significant mutations. Despite the heterogeneity of unexplained West syndrome, the combination of array CGH and whole-exome sequencing is an effective means of evaluating the genetic background in unexplained West syndrome. We provide additional evidence for *NR2F1* as a causative gene and for *CACNA2D1* and *BRWD3* as candidate genes for West syndrome.

1138F

Identification of rare variants associated with age-related macular degeneration by whole exome sequencing in 41 extended families. *R. Priya[1], M. Geerlings[2], E. De Jong[2], J. Corominas Galbany[2], L. Fritsche[3], K. Branham[4], N. Saksens[2], S. Perez[1], V. Chaitankar[1], M. Othman[4], D. Ferrington[5], E. Chew[1], G. Abecasis[3], C. Hoyng[2], A. Swaroop[1], A. den Hollander[2].* 1) 1National Eye Institute, National Institutes of Health, Bethesda, MD, USA; 2) Department of Ophthalmology, Radboud University Medical Center, Nijmegen, The Netherlands; 3) Departments of Biostatistics, University of Michigan, Ann Arbor, MI, USA; 4) Departments of Ophthalmology, University of Michigan, Ann Arbor, MI, USA; 5) Departments of Ophthalmology & Neurosciences, University of Minnesota Twin Cities, Minneapolis, MN, USA.

Age-related macular degeneration (AMD) is a common cause of visual impairment in the elderly. Genome-wide association studies have identified more than 20 susceptibility loci with common variants that explain almost half of the AMD heritability. More recently, targeted sequencing has revealed high-risk, rare coding variants at four known AMD loci. To further explain the missing hereditability, we undertook whole exome sequencing (WES) of 41 extended AMD families with multiple affected individuals. Exomes were captured using SureSelect Human All Exon kit (Agilent), and data were generated on Illumina sequencer. Reads were aligned to the human reference genome (NCBI build 37. 3, hg19) using Burrows-Wheeler Aligner, variants were called by GATK, and variants were annotatated with ANNOVAR. We identified 26 genes carrying rare, non-synonymous coding or splice variants that were shared among affected members from more than one family. Gene enrichment analysis using database for annotation, visualization and integrated discovery (DAVID) recognized genes in extracellular matrix, transmembrane transporter activity, voltage gated channel activity and non-membrane organelle cytoskeletons among the top annotation clusters. To validate their role, we have performed a gene-based burden test of the candidates using exome chip data from a large AMD case-control cohort. We have also examined human retina transcriptome data using weighted co-expression network analysis (WGCNA) to identify discrete modules of co-expressed genes that are enriched for AMD candidate genes. In conclusion, using WES, we have identified putatively family-specific AMD risk variants in genes that have not been previously associated with AMD. Our analysis highlights that family-based WES in extended AMD families may provide novel insights into the disease etiology.

1139W

Aggregated analysis of developmental disorders reveals novel risk genes and improves genetic diagnosis. *K. Retterer[1], J. Scuffins[1], D. Copenheaver[1], J. Juusola[1], W. K. Chung[2], Y. Shen[3].* 1) GeneDx, Gaithersburg, MD; 2) Departments of Pediatrics and Medicine, Columbia University Medical Center; 3) Departments of Systems Biology and Biomedical Informatics, Columbia University Medical Center.

Recent studies in developmental disorders (NDD) indicate that rare, undiagnosed developmental disorders often share risk genes or pathways with defined NDD phenotypes, such as autism and intellectual disability (ID), and aggregated analysis of *de novo* mutations can identify novel risk genes and improve the diagnostic yield of genetic analysis. We applied this strategy to a consecutive series of 1984 patients with undiagnosed developmental disorders for which we conducted whole exome sequencing of parents-proband trios. We identified *de novo* rare variants from exome data using a set of stringent empirical filters, annotated them by ANNOVAR, and predicted damaging missense (Dmis) mutations *in silico* by Meta-SVM. Using a gene-specific background *de novo* mutation rate, we identified candidate risk genes based on statistical significance of genes with at least two likely gene disrupting (LGD) or missense mutations. We identified 94 candidate risk genes with a false discovery rate (FDR) less than 0. 1. We then compiled publicly available *de novo* mutation data from ~ 5500 NDD published cases with autism, ID, epilepsy, and developmental delay, and reprocessed the mutation data using the same annotation approaches. We found that the majority (88%) of candidate risk genes have supportive evidence in the publicly available data set. There are several potentially novel genes, such as *HDAC2* and *HNRNPH2*, with plausible mechanisms in histone modification or pre-mRNA processing. We then inferred candidate risk genes in the combined data set (Meta-DD) and identified 135 genes with FDR < 0. 1, of which 60 genes are not candidates based on our cases only. Among our 358 cases carrying at least one LGD or Dmis rare variants in the Meta-DD candidate risk genes, 155 had no genetic diagnosis identified with the original clinical testing. This result represents an upper bound of a 7. 8% increase in diagnosis rate (from 25. 5% to 33. 3%). Interestingly, we also found there was an elevated rate of *de novo* damaging mutations in cases with more than one NDD phenotype. Specifically, the patients with all three major NDD phenotypes (autism, ID, epilepsy) are about twice as likely to carry damaging *de novo* rare variants in these candidate risk genes than autism patients without ID or epilepsy. This is consistent with recent findings that the *de novo* mutation burden in autism is concentrated in patients with lower IQ, and that these risk genes with *de novo* mutations are often pleiotropic.

1140T

Identification of a Nonsense Mutation in PIDD in Pakistani Families with Non-Syndromic Autosomal Recessive Intellectual Disability Using Exome Sequencing. *J. B. Vincent[1], A. mikhailov[1], T. Sheikh[1], A. Mir[2], M. Ayub[3], N. Vasli[1].* 1) Neurogenetics Sect, R30, Ctr Add/Mental Hlth,Clarke Div, Toronto, ON, Canada; 2) International Islamic University, Islamabad, Pakistan; 3) Department of Psychiatry, Queen's University, Kingston, ON, Canada.

Intellectual disability (ID) is a genetically heterogeneous neurodevelopmental disorder affecting 1-3% of the general population. It is characterized by deficits in memory skills and language development with difficulty in learning and problem solving. Today, more than 50 genes were detected in patients with ID. In an effort to identify the disease-causing mutations in patients with non-syndromic autosomal recessive ID (NS-ARID) using homozygosity mapping following by exome sequencing, we detected a homozygous nonsense mutation, c. 2587C>T, p. Gln863*, in the gene *PIDD* in two large consanguineous Pakistani families. Sanger sequencing analysis showed complete segregation within the two families with multiple affected individuals. Analysis of cryptic relatedness was performed on a set of 73,306 independent autosomal markers in PLINK version 1. 07. The pair-wise PI_HAT values between members of two families were less than 0. 10 indicating that these two families were not closely related. P53-Induced Death Domain Protein (PIDD) is a Leucine-rich repeat and death domain-containing protein expressing in brain, with a key role in DNA damage/stress response. It acts as a cell-fate switch through interacting with a prosurvivalmolecule, RIP1 and activation of NF-kB signaling to repair and survive the lesions, or via interacting with a pro-death factor, RAIDD and activation of caspase-2 and apoptotic cell death. The p. Gln863*mutation was in the death domain (DD), through which PIDD interacts with other DD proteins such as RIP1 or RAIDD. Functional studies are ongoing to establish the mechanism of pathogenicity of the variant detected. Our finding suggests the involvement of new pathways, i. e. PIDD-related pathways, in the etiology of intellectual disability.

1141F

Whole genome sequencing of autism spectrum disorder extended families reveals risk genes and variants missed by exome sequencing. *A. J. Griswold[1], H. N. Cukier[1], D. Van Booven[1], P. L. Whitehead[1], N. K. Hofmann[1], J. M. Lee[1], E. R. Martin[1,2], M. L. Cuccaro[1,2], J. R. Gilbert[1,2], J. P. Hussman[3], M. A. Pericak-Vance[1,2].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL, USA; 3) Hussman Institute for Autism, Baltimore, MD, USA.

Autism spectrum disorder (ASD) is a neurodevelopmental condition affecting 1 in every 68 individuals in the United States. A strong, but heterogeneous, genetic etiology of ASD has been demonstrated by genome-wide association studies (GWAS), copy number variation (CNV) screens, and, most recently, massively parallel sequencing. We previously reported a whole exome sequencing (WES) study of families with multiple, distantly related, ASD affected individuals and identified shared, inherited variants in ASD candidate genes previously implicated in a range of neurodevelopmental and neuropsychiatric diseases (Cukier, et al, 2014). However, in a minority of families, WES failed to reveal obvious putative risk variants based on our filtering criteria. Because WES evaluates only ~2% of the total genome coding for protein, misses certain coding exons, and is limited in detecting structural variations (SVs), we hypothesize that there are additional shared ASD risk variants in these families undetectable by WES. We have performed whole genome sequencing (WGS) on 15 affected ASD individuals across 6 of these extended families (2 or 3 individuals per family) to a depth of ~40X on the Illumina HiSeq2500. Of the ~4M single nucleotide variants (SNVs) identified per individual, 98% are noncoding and are currently being assessed for potential functional impact using bioinformatic tools such as GWAVA, CADD, and FATHHM-MKL. Among coding SNVs initially identified by WES, WGS calls are ~93% concordant across all individuals. To enrich for shared familial risk variants, we used the family structure and existing GWAS data to find variants in identical by descent (IBD) regions in each family. Depending on the family structure, the total size of the IBD regions in the six families range from 1. 5 to 500Mb. We identify 1,295 – 869,126 total SNVs within these regions per family with 15 – 6,946 in coding exons. WGS identifies ~25% more shared coding SNVs than WES per family. Furthermore, WGS reveals 0 - 300 SVs >1kb shared across individuals in each family disrupting 1-40 genes. Among these SVs is a ~1Mb deletion of the synaptic gene *SYN2* which has been implicated as an ASD candidate gene in genetic and animal model studies. Therefore, WGS is an essential tool to identify functional noncoding SNVs, SNVs not captured by WES, and SVs that might be conferring ASD risk in these unique pedigrees and were missed by other genomic technologies.

1142W

Comprehensive genomic analysis and quantitative functional assays in model systems provides evidence for the "two-hit" model in genomic disorders. *L. Pizzo[1], A. Polyak[1], J. Iyer[1], Y. Wang[1], A. Kubina[1], S. Girirajan[1,2].* 1) Department of Biochemistry and Molecular Biology. The Pennsylvania State University, University Park, PA; 2) Department of Anthropology. The Pennsylvania State University, University Park, PA.

Genomic disorders are associated with extensive phenotypic heterogeneity with individuals carrying the same genetic mutations manifesting varying levels of severity. This has led to challenges in genetic diagnosis, counseling and management. We recently described a two-hit model to explain the phenotypic variability associated with a 520-kbp deletion on chromosome 16p12. 1. A majority of these deletions were inherited from a parent and >20% of affected probands carried another large CNV elsewhere in the genome. We performed detailed phenotypic analysis of 50 probands with the deletion and identified variable frequency of neurodevelopmental features, including developmental delay (>90%), cardiac malformations (24%), epilepsy (21%), and behavioral problems (50%). Using de Vries scoring, we found that the severity of phenotypes increased through generations with carrier parents and grandparents showing subtle neuropsychiatric features such as depression and cognitive problems. We also performed exome sequencing experiments on 27 families and found an increased burden of rare inherited mutations in probands compared to parents. Several *de novo* and inherited single nucleotide variants were identified within known disease genes (such as *SETD5*, *DMD* and *LAMC3*) that potentially act in concert with the 16p12. 1 deletion genes in an additive or synergistic manner. To further assess the functional effect of the deletion itself and to validate genetic interactions, we performed phenotypic analysis by knocking down conserved genes in *Drosophila melanogaster*. We took advantage of the tissue-specific expression system conferred by UAS-Gal4 and used RNA interference to obtain the desired knockdown of putative genes in flies. We combined data from multiple phenotypic paradigms from different fly RNAi lines for the four conserved genes (*UQCRC2*, *C16ORF52*, *POLR3E*, and *CDR2*) and correlated the effect of gene dosage alteration to severity. For example, we found increased head sizes for reduced expression of *CDR2* (14. 74 SD), *UQCRC2* (9. 7 SD) and *POLR3E* (12. 37 SD) and a severe neuronal phenotype, for *POLR3E* (p<0. 0001). Using multiple quantitative tools, we also identified several enhancers and suppressors of the phenotypes due to knockdown of conserved genes within 16p12. 2. Our integrative analysis using copy-number variation and exome sequencing analysis along with quantitative functional analysis provides further evidence for the proposed two-hit model.

1143T

Meta-analysis of rare exome chip variants in 12,493 individuals identifies novel genes for white matter hyperintensities. *X. Jian[1], A. V. Smith[2], A. Teumer[3], J. C. Bis[4], E. Hofer[5], S. Hagenaars[6], L. Yanek[7], T. H. Mosley[8], M. A. Ikram[9], S. Seshadri[10], M. Fornage[1] for the neuroCHARGE Working Group.* 1) The University of Texas Health Science Center at Houston, Houston, TX; 2) Icelandic Heart Association, Kópavogur, Iceland; 3) University of Greifswald, Greifswald, Germany; 4) University of Washington, Seattle, WA; 5) Medical University of Graz, Graz, Austria; 6) The University of Edinburgh, Edinburgh, United Kingdom; 7) Johns Hopkins Medicine, Baltimore, MD; 8) The University of Mississippi Medical Center, Jackson, MS; 9) Erasmus University Medical Center, Rotterdam, Netherlands; 10) Boston University, Boston, MA.

Purpose: Cerebral white matter hyperintensities (WMH) detected on magnetic resonance imaging are common in the aging population and are part of the spectrum of vascular and neurodegenerative disease associated with brain aging. Though WMH are highly heritable, common variants identified by genome-wide association studies (GWASs) explain a small proportion of WMH variance. Functional low-frequency and rare variants within coding regions of the genome may contribute to WMH variance but have not been explored. **Methods:** Study subjects were stroke/dementia-free adults from six population-based cohorts within the CHARGE Consortium. We genotyped approximately 250,000 mostly coding variants in 11,192 European Americans (EAs) and 1,301 African Americans (AAs) using Infinium HumanExome BeadChip. Within each cohort, we used the seqMeta R package to compute race-specific score statistics for each variant and genotypic covariance matrices within predefined gene regions. These were then combined by meta-analysis to generate single-variant and gene-based tests of association. In gene-based tests, different grouping criteria were applied according to whether the variants were annotated as 'functional', 'damaging', or 'loss-of-function' by dbNSFP. **Results:** Single-variant tests replicated two GWAS significant or suggestive loci: 17q25 (rs3760128 in *TRIM65* with p = 2. 1×10−6 in EA+AA and 2. 6×10−6 in EA) and 2q33. 2 (rs2351524 in *NBEAL1* with p = 4. 66×10−7 in EA+AA and 6. 87×10−7 in EA). Gene-based tests identified two novel genes: *CNTNAP1* on 17q21 (p = 3. 76×10−6 by collapsing 10 functional variants with minor allele frequency (MAF) < 5% in EA+AA sample) is predominately expressed in brain tissue and has been suggested to have a role in the signaling between axons and myelinating glial cells; *APBB2* on 4p13 (p = 7. 49×10−7 by collapsing six damaging variants with MAF < 1% in EA sample) may modulate the internalization of beta-amyloid precursor protein and has been associated with Alzheimer's disease. **Conclusion:** We have replicated two GWAS loci and identified two novel genes harboring functional rare variants that may influence WMH. Our results suggest that both common and rare variants have an effect on this complex trait and choice of criteria for grouping rare variants may help eliminate noise and unveil novel association signals. Future replication studies are warranted to confirm our findings.

1144F

Allelic diversity of human disease genes in a large multigenerational Old Order Amish pedigree. *R. L. Kember[1], Y. Park[1], L. Hou[2], X. Ji[1], D. C. Craig[3], J. R. O'Connell[4], A. R. Shuldiner[4], D. J. Rader[1], C. D. Brown[1], F. J. McMahon[2], M. Bucan[1].* 1) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 2) Human Genetics Branch, National Institute of Mental Health Intramural Research Program, National Institutes of Health, Bethesda, MD; 3) The Translational Genomics Research Institute, Phoenix, AZ; 4) University of Maryland School of Medicine, Baltimore, MD.

Dissecting genetic components of complex diseases such as bipolar disorder (BP) can often be complicated by the presence of other medical co-morbidities. In such cases, family-based studies reduce confounding effects resulting from non-genetic components, however teasing out genetic associations to BP risk is still challenging. Recent studies suggest that variants in Mendelian disease genes can contribute to complex disease risk. Our previous work in the extended multigenerational Old Order Amish pedigree (ASMAD) revealed multiple linkage regions (in different subpedigrees) with BP risk-alleles, supporting a high degree of locus and allelic heterogeneity. This may in part be due to the co-morbid disease loci within this family that segregate in different subpedigrees. To study BP-specific genetic risk factors and reduce the genetic heterogeneity attributable to medical comorbidities, we utilize a combination of high throughput DNA and RNA sequencing and dense SNP chip genotypes. Genotype data from 394 family members were phased and imputed in reference to the whole genome sequence (WGS) for 80 family members (30 parent-child trios) and 1000 Genomes Project Phase 3 haplotype panel. Long-range phased haplotypes estimated using an extended pedigree allow us to explore the possible pleiotropy of Mendelian disease genes in relation to BP and the co-segregation between bipolar risk factors and Mendelian loci. We conducted carrier screening for all disease causing mutations in the Human Genome Mutation Database (HGMD-DM), and identified known and novel alleles in 3017 disease genes in both the heterozygous and homozygous state, many that are private to the ASMAD population. We created a complete catalog of all variants (both CNVs and SNPs) in disease genes, and compared allele frequencies between 1000 Genomes, an Amish control population, and the ASMAD sample. We can capture the distribution of these variants and follow them across multiple generations by utilizing the rich pedigree data. Strikingly, we find that burden of all variants in HGMD-DM genes is nominally elevated in individuals with BP, both in SNP and CNV analysis. Moreover, RNA-seq data for a subset of family members demonstrates the expression changes associated with these variants. These results reinforce the hypothesis of a complex genetic architecture underlying BP, and suggest that studies will benefit from investigating the role of variants in disease genes in BP and co-morbid conditions.

1145W

Combining autism and intellectual disability exome data yields insight into both disorders. *K. Roeder[1,2], A. E. Cicek[2], B. Devlin[3], J. D. Buxbaum[4], Autism Sequencing Consortium.* 1) Dept Statistics, Carnegie Mellon Univ, Pittsburgh, PA; 2) Computational Biology Dept, Carnegie Mellon Univ, Pittsburgh, PA; 3) Dept of Psychiatry, Univ of Pittsburgh School of Medicine, Pittsburgh, PA; 4) Seaver Autism Center, Icahn School of Medicine at Mount Sinai, NY, NY.

Identifying the large number of genes underlying neurodevelopmental disorders such as autism spectrum disorder (ASD) and intellectual disability (ID) is a major challenge, and interpreting how these genes contribute to risk is another. DNA from several thousand ASD and ID subjects and their parents have now been exome sequenced, from whom analysis of *de novo* mutations has proven a powerful tool for identifying risk genes. Moreover, such analyses suggest these disorders have many risk genes in common. We exploit this overlap to enhance power to discover risk genes for both disorders. Furthermore, we investigate the nature of the overlap of risk genes with respect to gene networks built for two critical spatio-temporal developmental periods in the brain. The analysis relies on recently developed statistical tools for the identification of risk genes using de novo mutations (TADA) and gene expression networks (DAWN). We apply community detection methods to DAWN networks to identify clusters of risk genes working in concert at key developmental periods. Combining ASD and ID exome data reveals 16 additional neurodevelopmental risk genes (FDR < 05), compared to earlier whole exome analyses, largely because the ASD and ID risk genes strongly overlap. Combining the risk genes identified for ASD with gene expression networks yields further insights. In the prefrontal cortex during fetal development we identify two functional clusters, one enriched for chromatin modification and another for regulation of synaptic transmission. In the mediodorsal nucleus of the thalamus and cerebellar cortex, assessed during infancy and early childhood, we identify three functional clusters enriched for chromatin modification, synaptic transmission/calmodulin binding and neuron-neuron synaptic transmission/GABA-A receptor activity. Moreover, ID genes are more strongly represented in three of the five functional ASD clusters. Thus, while ID and ASD share many risk genes, ID genes are predominantly found in a subset of the functional ASD risk gene communities, lending insight into the biological mechanisms underlying typical and atypical CNS development.

1146T

Mitochondrial genome sequence variation, mutation accumulation, heteroplasmy, and haplogroups associated with Alzheimer's Disease. *X. Wang[1,2,5], M. A. Florez[1], L. Jiang[1,2], M. Teferra[1], J. D. Mahnken[3,5], R. A. Honea[4,5], J. Burns[4,5], R. Swerdlow[4,5], M. L. Michaelis[1,2,5], E. K. Michaelis[1,2,5].* 1) Higuchi Biosciences Center, The University of Kansas, Lawrence, KS; 2) Department of Pharmacology and Toxicology, The University of Kansas, Lawrence, KS; 3) Department of Biostatistics, The University of Kansas Medical Center, Kansas City, KS; 4) Department of Neurology, The University of Kansas Medical Center, Kansas City, KS; 5) Alzheimer's Disease Center, The University of Kansas Medical Center, Kansas City, KS.

Alzheimer's Disease (AD) is the most common form of dementia, affecting one in nine people aged 65 and older, and one in three after reaching 85. Mitochondrial dysfunction has been proposed to play an important role in the etiology of AD. While the nuclear genome codes for a large number of proteins in this organelle, the mitochondrial genome codes for 37 genes, 13 of which code for protein components that are essential for bioenergetics. Because almost all current AD genetics/genomics studies are conducted on the nuclear genome, how the mitochondrial genome might be involved in AD pathogenesis is still largely unknown. To study how sequence variation and the accumulation of mutations in the mitochondrial genome might be associated with AD, we have applied Illumina HiSeq next-generation sequencing (NGS) to identify mitochondrial DNA (mtDNA) variants that are significantly associated with AD. Furthermore, the accumulation of mutations in the mitochondrial genome, especially the distribution pattern of mutations and the resultant sequence heterogeneity (i. e. , heteroplasmy), have been examined. As an extension of this analysis, mtDNA-based haplogroups have also been studied for their association with the disease. The study cohort contains over 400 AD cases and age-matched healthy controls and was recruited to the University of Kansas Alzheimer's Disease Center. Because of the small size of the mitochondrial genome, we are able to conduct highly multiplexed sequencing with 96 or more samples being sequenced in a single lane. As a result, we have identified new mtDNA polymorphic sites that are significantly associated with AD. The accumulation of mutations in mtDNA has been shown to be extensive and affect many sites in the mitochondrial genome. The incidence of AD is found to be significantly less common in certain haplogroups. The high-throughput and cost effectiveness of our NGS approach afford unprecedented opportunities for the study of AD and other diseases, including many psychiatric and neurological diseases, characterized by mitochondrial deficits.

1147F

Using zebrafish to model the role of *GEMIN5* in neurological disease. *K. Schaffer[1,2], E. Burke[1,3], A. Adomako-Ankomah[2], J. Fohtung[2], W. Bone[1], R. Sood[2], B. Carrington[2], K. Bishop[2], W. A. Gahl[3], S. Burgess[2].* 1) NIH Undiagnosed Diseases Program Translational Laboratory, Bethesda, MD; 2) Translational and Functional Genomics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 3) Section on Human Biochemical Genetics, Medical Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD.

As a participant in the NIH Undiagnosed Diseases Network (UDN), a 21-year-old male patient was seen and demonstrated signs of upper motor neuron dysfunction, deafness, and seizures, and has since declined into a state of paralysis. After SNP and exome analysis of the patient and his family, *Gem associated protein 5* (*GEMIN5*) was identified as a strong candidate disease allele. *GEMIN5* is the snRNA-binding component of the survival of motor neuron (SMN) complex. The SMN complex plays an important role in mRNA splicing by cytoplasmic assembly of spliceosomal small nuclear ribonucleoproteins that subsequently catalyze the splicing reaction. While *GEMIN5* is not currently associated with a human disease, the gene *SMN1* (and the related *SMN2*), the core protein of the SMN complex, is associated with spinal muscular atrophy (SMA) types 1-4. Reduction of *SMN1* is also associated with severe motor neuron degeneration. Therefore, due to the known interaction of the SMN complex and *GEMIN5*, we hypothesize the missense mutation found in one copy of *GEMIN5* is contributing to this patient's disease. We tested the function of *GEMIN5* in zebrafish by targeted inactivation using CRISPR-Cas9. We observed fatality between 7-21 dpf in fish homozygous for frame shifting deletions in *GEMIN5*. We tested the mutant fishes' ability to repair neural damage by performing a hair cell regeneration assay after copper exposure. The mutant animals had no regenerative response despite their developmentally normal appearance. We have created a transgenic line ubiquitously expressing wild-type human *GEMIN5* or the variant found in the patient to rescue the zebrafish mutant phenotypes. We examined the mutants' motor neuron development by sensory motor neuron staining using Znp-1 antibody. These results confirm a role for *GEMIN5* in survival and reinforce our interest in testing the association between the identified variants in the patient and the manifestation of the disease.

1148W

Mutation analysis of the *WNT7A* and interaction study with the *DISC1* in schizophrenia. *M. Cheng[1], A. Chuang[1], C. Chen[2], S. Hsu[1], H. Tsai[1].* 1) Department of Psychiatry, Yuli Branch, Taipei Veterans General Hospital, Hualien Country, Taiwan; 2) Department of Psychiatry, Chang Gung Memorial Hospital at Linkou and Chang Gung University School of Medicine, Taoyuan, Taiwan.

Schizophrenia is a devastating chronic mental illness that affects about 1 percent of general population worldwide. Family, twin and adoption studies have provided substantial evidence to support a strong genetic basis underlying the pathogenesis of schizophrenia. Several reports show that Wnts mediate the neuron synaptic function and brain development, suggesting that aberrant Wnt signaling involves the pathophysiology of schizophrenia. This study aimed to investigate whether the *WNT7A*, a member of the WNT gene family, is associated with schizophrenia. We resequenced putative promoter region and all the exons of the *WNT7A* in 570 patients with schizophrenia and 563 non-psychotic controls from Taiwan and conducted a case-control association study. We further evaluated the intergenic interaction between the SNPs of the *WNT7A* and the Ser704Cys of the *DISC1* and their relationship with schizophrenia. We totally identified 14 genetic variants of the *WNT7A* in this sample, including 5 SNPs (rs73151669, rs3749319, rs12639607, rs3762719, and c. 777G>A) and 9 rare mutations (minor allele frequency <5%). However, SNP- and haplotype-based analyses showed no association of these SNPs with schizophrenia. However, we identified several rare mutations that were predicted to have pathological effect on the WNT7A protein function or disrupt microRNAs binding sites located at 3'UTR of the *WNT7A* in the schizophrenic patients. For the epistatic interaction between the *WNT7A* and the *DISC* in schizophrenia, the A allele (A/A+C/A) at the *WNT7A* rs73151669 increased risk for schizophrenia in the context of a *DISC1* SerSer background (odds ratio = 1. 48, p value = 0. 01), but was not significant in patients carrying Cys allele (odds ratio = 0. 91, p value = 0. 7). Our results suggest that multiple rare variants of the *WNT7A* might contribute to the pathogenesis of schizophrenia in some patients and there is a genetic association of the *WNT7A* in epistasis with the *DISC1* and the risk of schizophrenia.

1149T

The Alzheimer's Disease Sequencing Project Discovery Phase: Case-Control Study Design, Progress, and Preliminary Results. *J. C. Bis[1], A. C. Naj[2], G. W. Beecham[3], J. Dupuis[4], J. E. Below[5], K. L. Hamilton-Nelson[3], H. Lin[4], Y. Ma[4], N. Gupta[6], W. Salerno[7], Y. Chen[4], V. Chouraki[4], L. A. Cupples[4], C. Cruchaga[8], G. Jun[4], B. Kunkle[3], A. Partch[2], R. Koester[4], M. Schmidt[3], B. Vardarajan[9], E. Wijsman[1], A. Goate[10], J. Haines[11], S. J. van der Lee[12], C. van Duijn[12], S. Seshadri[4], A. DeStefano[4], E. Martin[3], M. Fornage[5], L. Farrer[4] on behalf of the Alzheimer's Disease Sequencing Project Case-Control Work Group.* 1) University of Washington, Seattle, WA, USA; 2) University of Pennsylvania, Philadelphia, PA, USA; 3) University of Miami, Miami, FL, USA; 4) Boston University, Boston, MA, USA; 5) The University of Texas Health Science Center at Houston, Houston, TX, USA; 6) The Broad Institute, Cambridge, MA, USA; 7) Baylor College of Medicine, Houston, TX, USA; 8) Washington University in St. Louis, St. Louis, MO, USA; 9) Columbia University, New York, NY, USA; 10) Icahn School of Medicine at Mount Sinai, New York, NY, USA; 11) Case Western Reserve University, Cleveland, OH, USA; 12) Erasmus University Medical Center, Rotterdam, the Netherlands.

Responding to a 2012 Presidential Initiative to fight Alzheimer's disease (AD) and a request from the Director of the National Institutes of Health, the National Institute on Aging and National Human Genome Research Institute (NHGRI) jointly developed a project to analyze the genomes of a large number of well-characterized individuals with or without AD using $25M in funding already committed to the NHGRI's Large-scale genome Sequencing and Analysis Centers (LSACs). The overarching goals of this effort, **the Alzheimer's Disease Sequencing Project (ADSP)**, include identifying new (1) genes involved in AD, (2) genomic variation contributing to AD risk or protection, (3) insights as to why some individuals with risk factor genes escape AD, and (4) potential avenues for therapies and AD prevention. The ADSP Discovery Phase includes two components: the **Family Study** conducted whole genome sequencing on multiplex AD families and the **Case-Control Study** sequenced whole exomes of unrelated participants from the Alzheimer's Disease Genetics Consortium (ADGC) and Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium. Here, we describe the design, progress, results, and public data availability for the Case-Control Study. On the basis of *a priori* risk – determined by age, sex, *APOE*, and pathology – we selected **5,107 low-AD-risk cases** and **4,976 high-AD-risk controls** (who had 'escaped' the disease). To enrich the risk profile and diversity of the study, we also sequenced unrelated cases from AD families (N=512) as well as Dominican-Hispanic cases (N=173) and controls (N=171). Sequencing has recently been completed by three LSACs (Baylor College of Medicine, Washington University-St. Louis, and the Broad Institute); BAM files have been deposited in dbGaP/SRA; novel QC methods have integrated genotype calls from two algorithms (*GATK*, *ATLAS*); and we have performed sequence-based principal components ancestry analyses. Our ongoing analyses focus on both individual common variants (minor allele count > 10) using Score tests and Firth bias-corrected regression as well as rare functional variants (including Loss-of-Function), aggregated within genes using burden tests and SKAT. In addition to identifying important coding variant associations in known AD genes, we anticipate that our analyses will also reveal promising new loci for AD risk and protection. We expect to present preliminary findings and plans for the next phase of ADSP.

1150F

Novel ABCC1 variant associated with frontotemporal dementia. *M. De Both, A. Siniard, R. Richholt, K. Ramsey, MJ. Huentelman.* Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ.

Frontotemporal dementia (FTD) is among the most common forms of primary dementia, second only to Alzheimer's disease in early-onset cases. It is characterized by the degeneration of neurons in the frontal and temporal lobes of the brain, which can manifest as severe changes to personality, decreased social awareness, and impaired ability to comprehend words. While it often presents in families, heritable causes of the disease are not fully understood. FTD can be grouped into three general categories: behavioral variant, primary progressive aphasia (PPA), and movement disorder type. Here, we studied a familial case of PPA. In ten siblings, 6 females were diagnosed with this form of FTD while 1 female and 3 males were not. Exome sequencing (Illumina TruSeq 62Mb Exome Enrichment Kit, sequenced on Illumina HiSeq 2000) was conducted on all surviving members (3 affected females, 1 unaffected female, 1 unaffected male) with 75 million paired reads per sample, from which we identified a novel, putatively dominant acting variant in ABCC1 (chr16:16216007 A>G, CADD = 14. 32, unobserved in ExAC) common to all 3 cases and not present in the 2 unaffected individuals. As ABCC1 has previously been found to facilitate the clearing of amyloid-b across the blood-brain barrier and into the bloodstream, we suspect our FTD cases to present with increased levels of amyloid-b in the brain, a phenotype not typically observed with FTD. However, we have not as of yet been able to examine the amyloid load in any of the cases in our study family. While amyloid-b accumulation of the brain is indicative of Alzheimer's disease, it is generally not strongly observed in cases of FTD. Treatment with thiethylperazine, already FDA-approved to relieve nausea, has been shown to reduce amyloid-b accumulation in mouse models expressing ABCC1 but exhibits no effect in knockout mice lacking ABCC1, suggesting the drug functions by inducing ABCC1 expression or by facilitating its biochemical actions. Such amyloid-b pathology could represent a novel mechanism for FTD and a therapeutic target in ABCC1. Studies are currently under way to examine the role of this newly identified ABCC1 variant on amyloid production in vitro and in vivo.

1151W

Rare Loss of Function Variants from Immune System Genes and Brain Over-expressed Genes are Enriched in Alzheimer's Disease Patients Compared to Centenarians. *Y. Freudenberg-Hua[1,5], W. Li[2], V. Vacic[3], A. V. Abhyankar[3], D. Ben-Avraham[4], E. Christen[1], J. Koppel[1], B. Greenwald[5], N. Barzilai[4], S. Germer[3], R. B. Darnell[3], J. Freudenberg[2,6], G. Atzmon[4], P. Davies[1].* 1) Litwin-Zucker Research Center for the Study of Alzheimer's Disease, The Feinstein Institute for Medical Research, North Shore-LIJ, Manhasset, NY; 2) Robert S. Boas Center for Genomics and Human Genetics, The Feinstein Institute for Medical Research, North Shore-LIJ, Manhasset, NY; 3) New York Genome Center, New York, NY; 4) Institute for Aging Research Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY; 5) Division of Geriatric Psychiatry, Zucker Hillside Hospital, North Shore-LIJ, Glen Oaks, NY; 6) Translational Genetics, Regeneron Genetics Center, Tarrytown, NY.

We sequenced the whole genomes of 45 patients with Alzheimer's diseases (AD) and 53 cognitively healthy centenarians of Ashkenazi Jewish ancestry. Variants were quality filtered and annotated based on RefSeq transcripts. Despite the relatively small sample size, we replicated associations of the known AD variants *APOE4* (OR=8. 3, P=3. 6e-08) and p. Arg47His in *TREM2* (OR=7. 6, P=0. 048), indicating the power of our approach to compare AD cases with centenarian super-controls. We restricted our subsequent analysis to rare single-nucleotide and indel variants not present in the 1000 Genomes database and with a minor allele frequency below 2% in our Ashkenazi sample. We compared the burden of predicted loss-of-function variants (LoFs, defined as nonsense, frame shift, initiation codon (INIT) and splice site variants) and protein altering variants (PAVs, defined as LoFs plus missense, in frame indels, stop-loss) between cases and controls, as normalized by the level of rare synonymous variants. Notably, for the 895 rare LoFs, we found an increased burden among AD cases (OR=1. 2, Fisher's exact P=0. 004). No enrichment was observed for the 11,845 rare PAVs (OR=1. 01, P=0. 662). To probe the role of the immune system in AD, we next limited the analysis to 4,723 immune response genes. We found a moderately increased burden among AD cases for the 2,617 PAVs (OR=1. 13, P=0. 027) as well as the 190 LoFs (OR=1. 32, P=0. 038). The largest burden was found for INIT variants (OR=2, P=0. 040 overall; OR=3. 23, P=0. 076 among immune genes). Immune function genes with INIT mutations exclusive to AD patients are *DLC1, NR1H3, DAPK2* and *BPI*. We furthermore observed enrichment among cases for the 73 LoFs located in genes over-expressed in the human brain (OR=1. 49, P=0. 043), and this enrichment appears to be stronger when restricted to orthologs of mouse essential genes (OR=3. 23, P=0. 021). In contrast, we observed no enrichment for the 185 LoFs from genes being annotated as neither essential nor over-expressed in the brain (OR=1. 02, P=0. 897). Essential brain over-expressed genes that are mutated exclusively in our AD patients are *CTNNA2, TCF7L2, CC2D1A, ZGPAT, PNPLA8, PPM1B, CLCN7, AVIL*, and *CNTNAP1*. In conclusion, we found the burden of rare LoFs to be increased among AD patients, with significant contributions coming from immune system genes as well as brain over-expressed genes. Our findings support a role of rare variants from immune and brain over-expressed genes in the etiology of AD.

1152T

Exome sequencing identifies novel genes associated with Multiple System Atrophy. *A. Siniard, I. Schrauwen, J. Corneveaux, K. Ramsey, M. Huentelman.* Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ.

Multiple System Atrophy (MSA) is a rare aggressive neurodegenerative disorder with adult onset. It can be characterized by a mixture of autonomic failure, ataxia, and parkinsonism. The two main disease subtypes can be distinguished clinically as "MSA-P" (predominant parkinsonism) or "MSA-C" (dominant cerebellar features). Both are associated with significant decline in dopamine transporter levels during disease progression. The cardinal pathology of MSA is glial inclusions enriched for alpha-synuclein protein. The cause of MSA is unknown but is likely due to the combination of heritable and environmental factors. To investigate the heritable risk factors associated with MSA we studied nuclear families with one affected individual using exome sequencing with Illumina's TruSeq Exome Kit. By using a selective filtering and variant prioritization approach, we attempted to identify variants that met following criteria: (1) family specific variants or ones present at a very low frequency in the population (ExAC MAF <0. 05), (2) variants predicted to have a damaging effect on the protein (CADD >11), and (3) familial segregation consistent with a digenic heterozygous inheritance model. In two families, variants were found in *LRRK2* and *SLC6A3* (aka DAT, dopamine transporter) or *LRRK2* and *SLC6A2* (aka NET, norepinephrine transporter). Mutations in *LRRK2* have previously been associated with Parkinson's disease (PD) risk. Additionally, LRRK2 is known to regulate synaptogenesis and dopamine receptor activation. In the third family, we identified two variants in different voltage-gated calcium channel subunits, CAV1. 3 and CAV2. 1. *CACNA1D,* encoding CAV1. 3, has been reported to be involved in neurodegenerative mechanisms associated with the development of PD. *CACNA1A,* encoding CAV2. 1*, w*as previously associated with spinocerebellar ataxia type 6, which shares several clinical features with MSA-C. This study is the first report of a nuclear family analysis of MSA in a Caucasian cohort. While small, our findings suggest the presence of rare variants with significant heterogeneity as a possible mechanism of disease. Interestingly, although each family studied had a different combination of variants they coalesce on a common biological theme related to dopamine processing in the brain, a previously hypothesized disease mechanism. Future work is necessary to further generalize these findings but also to investigate the functional ramification of these newly identified variants.

1153F

Rare coding mutations identified by sequencing of Alzheimer's disease GWAS loci. *B. N. Vardarajan[1,2,3,5], M. Ghani[8], A. Kahn[3], S. Sheikh[1,3], C. Sato[8], S. Barral[1,2,3], J. L. Lee[1,2,6], R. Cheng[2,3], C. Reitz[1,2,3], R. Lantigua[3,7], D. Reyes-Dumeyer[2,3], M. Medrano[10], I. Z. Jimenez-Velazquez[11], E. Rogaeva[8], P. St George Hyslop[8,9], R. Mayeux[1,2,3,4,6].* 1) Neurology, Columbia University, New York, NY; 2) The Taub Institute for Research on Alzheimer's Disease and the Aging Brain, Columbia University, New York, NY; 3) The Gertrude H. Sergievsky Center, Columbia University, New York, NY; 4) Psychiatry, Columbia University, New York, NY; 5) Systems Biology, Columbia University, New York, NY; 6) Epidemiology, Mailman School of Public Health, Columbia University, New York, NY; 7) Medicine, Columbia University, New York, NY; 8) Tanz Centre for Research in Neurodegenerative Diseases, Deparment of Medicine,Toronto, Canada; 9) Cambridge Institute for Medical Research, Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK; 10) Medicine, Pontificia Universidad Catolica Madre y Maestra, Santiago, Dominican Republic; 11) Medicine, Geriatrics Program, School of Medicine, University of Puerto Rico, San Juan, Puerto Rico.

Objective: To detect rare coding variants underlying loci detected by genome-wide association studies (GWASs) of late-onset Alzheimer's disease (LOAD). **Methods**: We conducted targeted sequencing of *ABCA7, BIN1, CD2AP, CLU, CR1, EPHA1, MS4A4A/MS4A6A* and *PICALM* in three independent LOAD cohorts: 176 patients from 124 Caribbean Hispanics families, 120 patients and 33 unaffected individuals from the 129 NIA-LOAD Family Study; and 263 unrelated Canadian individuals of European ancestry (210 sporadic patients and 53 controls). Rare coding variants found in at least two datasets were genotyped in independent groups of ancestry matched controls. Additionally, the Exome Aggregation Consortium (ExAC) was used as a reference dataset for population-based allele frequencies. **Results:** Overall we detected a statistically significant 3. 1-fold enrichment of the non-synonymous mutations in the Caucasian LOAD cases compared with controls (p=0. 002) and no difference in synonymous variants. A stopgain mutation in *ABCA7* (E1769X) and missense mutation in *CD2AP* (T374A) were highly significant (p=5. 3e-04 and 5. 3e-08 respectively) in Caucasian LOAD cases, and mutation in *EPHA1* (P460L) was significantly associated in Caribbean Hispanics (p=8. 64e-05), with LOAD. P460L in EPHA was also nominally significant in the Caucasians (p=0. 03). The *EPHA1* variant segregated completely in an extended Caribbean Hispanic family and was also nominally significant in the Caucasians. Additionally, in *BIN1* K358R segregated in two of the six families where the mutations were discovered. **Interpretation:** Targeted sequencing of confirmed GWAS loci revealed an excess burden of deleterious coding mutations in LOAD with the greatest burden observed in *ABCA7* and *BIN1*. Identifying coding variants in LOAD will facilitate the creation of tractable models for investigation of disease related mechanisms and potential therapies.

1154W

The percentage of sporadic amyotrophic lateral sclerosis patients with a known genetic etiology of disease is less than previously described. *J. M. Downie[1], S. B. Gibson[2], K. P. Figueroa[2], M. B. Bromberg[2], S. M. Pulst[2], L. B. Jorde[1].* 1) Department of Human Genetics, University of Utah School of Medicine, 15 North 2030 East Room 5100, Salt Lake City, UT 84112; 2) Department of Neurology, University of Utah School of Medicine, 175 North Medical Drive East, Salt Lake City, UT 84132.

Amyotrophic lateral sclerosis (ALS) is a fatal disease in which upper and lower motor neurons degenerate, leading to eventual paralysis. More than 16 genes have now been implicated in ALS pathogenesis. 90% of ALS cases occur with no previous family history and are termed sporadic ALS (SALS). However, twin studies have yielded an SALS heritability estimate of 61%. There is substantial disagreement about the proportion of SALS cases caused by mutations in known ALS genes, with estimates ranging between 11% and 28%. These estimates are largely derived from studies calculating the proportion of SALS patients with a rare, protein-altering mutation in a designated panel of genes. However, these studies relied on rarity as an indicator of deleteriousness instead of considering the likelihood that these variants affect normal protein function. Defining the degree to which ALS genes contribute to SALS pathogenesis will help to determine how much heritability remains to be explainedand will allow for a better understanding of SALS pathogenesis. To this end, we incorporated *in-silico* predictions of variant deleteriousness in 96 SALS patients who were exome sequenced using the Illumina HiSeq platform. Reads were aligned to GRCh37 using BWA, and variants were called using GATK (v3. 3). *ATXN2* and *C9orf72* repeat expansions were detected via PCR. Predictions of variant deleteriousness were obtained using 11 different algorithms contained in dbNSFP v2. 9. To prevent population stratification effects, we focused our study specifically on Europeans with ALS. Nine non-Europeans, who were identified using principal components analysis, were removed from further analysis. We calculated the proportion of SALS patients who had at least one rare (European MAF < 0. 01; 1000 Genomes + ESP6500), deleterious (based on *in-silico* predictions) amino-acid altering variant in any of fourteen ALS genes (*ANG*, *DAO*, *DCTN1*, *EWSR1*, *FIG4*, *FUS*, *OPTN*, *SETX*, *SOD1*, *SQSTM1*, *TAF15*, *TARDBP*, *VAPB*, *VCP*) or possessed a pathogenic repeat expansion in either *ATXN2* or *C9orf72*. We found that across the 11 *in-silico* algorithms that 14. 6%, on average (SD = 0. 0277), of SALS patients had at least one rare, deleterious variant or a pathogenic repeat expansion in one of these genes. Our estimate, which is based on both variant rarity and deleteriousness, is lower and likely more accurate than previous frequency-based estimates.

1155T

Identification of SPG11/*KIAA1840* mutations in patients with autosomal recessive axonal Charcot-Marie-Tooth disease. *L. Pedace[1], C. Montecchiani[1], T. Lo Giudice[1,2], A. Casella[1], M. Mearini[1], F. Gaudiello[1], C. Terracciano[2], J. L. Pedroso[3], R. Massa[2], O. G. P. Barsottini[3], P. H. St George-Hyslop[4,5,6], T. Kawarai[7], A. Orlacchio[1,2].* 1) Laboratorio di Neurogenetica, CERC - IRCCS , Rome, Italy; 2) Dipartimento di Medicina dei Sistemi, Université di Roma "Tor Vergata", Rome, Italy; 3) Department of Neurology, Universidade Federal de São Paulo, Brazil; 4) Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Ontario, Canada; 5) Department of Medicine, University of Toronto, Toronto, Ontario, Canada; 6) Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom; 7) Department of Clinical Neuroscience, Institute of Biomedical Sciences, Tokushima University Graduate School, Tokushima, Japan.

Background: Mutations in the SPG11/*KIAA1840* gene are common cause of autosomal recessive hereditary spastic paraplegia with *thin corpus callosum* and peripheral axonal neuropathy and accounts for about 40% of autosomal recessive juvenile amyotrophic lateral sclerosis. The overlap between axonal Charcot-Marie-Tooth disease and both these diseases, as well as the common autosomal recessive inherited findings of *thin corpus callosum* and axonal Charcot-Marie-Tooth in three related patients, prompted us to sequence the SPG11/*KIAA1840* gene in affected individuals with autosomal recessive axonal Charcot-Marie-Tooth disease. **Methods:** The study focused on 28 unrelated pedigrees with autosomal recessive Charcot-Marie-Tooth type 2 and without genetic assessment, originated in Italy, Brazil, Canada, England, Iran, and Japan. The diagnosis was based on clinical findings and familiar history. Clinical and instrumental functional analysis consist of neurological assessment, diagnostic imaging, electroneurographic assay, and sural nerve biopsy. Molecular studies include linkage analysis, Sanger sequencing, RFLP analysis and bioinformatics. **Results:** Linkage study of all families showed homozygous haplotypes and produced positive logarithm of odds score in all affected subjects. Sanger sequencing identified 15 SPG11/*KIAA1840* mutations in 12 families. Two sequence changes were never reported before and *in silico* analysis predicted their pathogenetic effect. All mutations seemed to be pathogenic since they co-segregated with the disease in all pedigree and were absent in controls. All known axonal Charcot-Marie-Tooth autosomal recessive loci (CMT1F/*CMT2E/NEFL,* CMT2A/CMT2B1/*LMNA,* CMT2A2/*MFN2,* CMT2B2/*MED25,* CMT2F/*HSPB1*/*HSP27,* CMT2K/*GDAP1,* CMT2P/*LRSAM1,* CMT4C/*KIAA1985,* CMTRID/*COX6A1,* NMAN/*HINT*), genes causing autosomal recessive hereditary spastic paraplegia with *thin corpus callosum* and axonal peripheral neuropathy (*SPG7, SPG15*, and *SPG21*), as well as the causative gene of peripheral neuropathy with or without agenesis of the corpus callosum (*SLC12A6*) were screened out. An additional locus for autosomal recessive Charcot-Marie-Tooth disease type 2H on chromosome 8q13-21. 1 was excluded by linkage analysis. **Conclusions:** Our results indicate that SPG11/*KIAA1840* is the causative gene of a wide spectrum of clinical features, including autosomal recessive axonal Charcot-Marie-Tooth disease.

1156F

Uniparental disomy causing hereditary spastic paraplegia: Paternal disomy in FA2H causing homozygous SPG35 in two non-consanguine families. *A. S. Soehn[1], T. W. Rattay[2,3], S. Beck-Woedl[1], K. Karle[2,3], S. Wiethoff[2,3], J. Reichbauer[2,3], U. Gaiser[4], M. Doebler-Neumann[4], P. Bauer[1], S. Zuechner[5], R. Schuele[2,3,5], L. Schoels[2,3].* 1) Institute of Medical Genetics and Applied Genomics, University of Tuebingen, Tuebingen, Germany; 2) Center for Neurology, Department of Neurodegenerative Diseases, Hertie-Institute for Clinical Brain Research, University of Tuebingen, Tuebingen, Germany; 3) German Center of Neurodegenerative Diseases (DZNE), Tuebingen, Germany; 4) Department of Neuropediatrics, Tuebingen University School of Medicine, Tuebingen, Germany; 5) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL, USA.

Hereditary spastic paraplegias (HSPs) are degenerative diseases of upper motor neurons characterized by a progressive spastic gait disorder. They are genetically highly heterogenous. SPG35 is an autosomal-recessive (AR) HSP caused by mutations in the fatty acid 2-hydroxylase (*FA2H*), which is essential for maintenance of the neuronal myelin sheath. Thus far approx. 20 mutations have been described in 13 SPG35 families with 36 affected patients. Here we describe two novel homozygous *FA2H* mutations in two non-related non-consanguine families detected by whole exome sequencing or a HaloPlex assay providing high coverage of all known HSP genes ("HSP panel"), respectively. Segregation analysis revealed both fathers being heterozygous mutation carriers whereas both mothers did not carry *FA2H* mutations. A macro deletion within SPG35, that could have caused a hemizygous genotype, was excluded by MLPA. Finally, a microsatellite array revealed paternal disomy in both families leading to homozygous SPG35 mutations. UPD has rare been described as causative mechanism in neurodegenerative diseases. This is the first report of UPD as a cause for HSP. Interestingly, we identified this rare mode of inheritance in two families with the rare genotype SPG35. We are not aware of any mechanism how SPG35 mutations may trigger UPD. Since UPD seems to be a relevant factor in AR HSP we recommend segregation analysis especially in non-consanguine homozygous index cases to unravel further UPD cases in this field.

1157W

A study on the Association of monoamine and glutamate gene polymorphisms with alcohol dependence in patients from North India. *R. Gupta[1], R. Singh[1], S. Sahni[1], T. Grover[1], A. Ambekar[2], M. Vaswani[2], A. Sharma[1].* 1) Department of Anatomy, AIIMS, New Delhi, Delhi, Indi; 2) Department of Psychiatry, AIIMS, New Delhi, Delhi, India.

Alcohol dependence (AD) is a complex disorder influenced by monoamine and glutamate pathways which play a role in normal brain functions. Disturbances in these pathways are implicated in many psychiatric disorders including alcohol dependence and polymorphisms present in these pathway genes are reported to increase the risk of developing alcohol dependence. **Aim:** To identify polymorphisms in the Monoamine and Glutamate pathway genes and correlate their presence with alcohol dependence. **Method:** The study conformed to declaration of Helsinki and was approved by the Institute ethics committee. Total of 200 AD patients (DSM IV criteria) from outpatient department of Psychiatry and 200 healthy volunteers from the general population were recruited. Genomic DNA from blood samples was screened for polymorphisms in the DRD4 120bp duplication, DRD3 Ser9Gly, DRD2 Taq1A, COMT rs4680, -287A>G, HT1B G861C, 5-HTTLPR, STin2 genes of monoamine and GRIN2A rs11866328, rs1071502, rs1375067, rs1530669 of glutamate pathway. Difference in genotype frequencies was assessed through chi-square test and results were correlated with duration of alcohol use and dependence, age at onset of dependence, quantity of alcohol consumed, WHO ASSIST score and liver function tests of SGPT, SGOT, GGT. **Result:** Age at first use of alcohol was 21. 84±0. 4711, WHO ASSIST score was 42. 34±0. 7103, SGOT, SGPT, GGT were 84. 94±5. 426, 68. 98±4. 529 and 208. 5±28. 43 respectively. COMT rs4680, GRIN2A rs11866328 and rs1071502 showed a significant difference between patients and controls ($p < 0.05$). Higher frequency of HT1B (G861C) C allele and GRIN2A rs1375067 was observed in cases but the difference was not significant. The remaining markers did not show any change between cases and controls. The COMT rs4680 GG genotype showed association with amount of alcohol consumed (95. 53±13. 49 gms/day) as compared to GA (89. 55±6. 99 gms/day) and AA (61. 73±7. 33 gms/day) genotypes. The 5-HTTLPR short allele SS was significantly associated with increased alcohol intake, WHO ASSIST score and duration of dependence. GRIN2A markers rs11866328 (A allele) and rs1071502 (C allele) showed significant correlation with raised levels of SGOT (77. 63±11. 16; 103. 5±22. 51) and GGT (215±49. 78; 219. 3±55. 04) respectively. **Conclusion:** This is the first report on association of COMT and GRIN2A with alcohol dependence from India. Significantly higher frequency of the markers in patients seems to confer risk for severity of AD in them.

1158T

Identification of ADHD candidate genes in large pedigrees combining linkage analysis and whole-exome sequencing. *J. Corominas-Galbany[1], M. Klein[1], T. Zayats[2], S. Gross-Lesch[3], A. Henkes[1], C. Jacob[3], A. Reif[4], M. Romanos[5], A. Arias Vasquez[1,2,6], K. P. Lesch[3], B. Franke[1,7], IMpACT consortium.* 1) Department of Human Genetics, Radboud university medical center, Nijmegen, The Netherlands; 2) Department of Biomedicine, Bergen University, Bergen, Norway; 3) Department of Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany; 4) Department of Psychiatry, Psychosomatics and Psychotherapy, University of Frankfurt am Main, Frankfurt, Germany; 5) Department of Child Psychiatry, Psychosomatics and Psychotherapy, University of Wuerzburg, Wuerzburg, Germany; 6) Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Radboud university medical center, Nijmegen, The Netherlands; 7) Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud university medical center, Nijmegen, The Netherlands.

Attention-Deficit/Hyperactivity Disorder (ADHD) is a common neuro-psychiatric and multifactorial disorder characterized by inattention, and/or motor hyperactivity and impulsivity. This disorder has a complex genetic background hampering the identification of underlying genetic risk factors. The study of multi-generation pedigrees with multiple affected individuals can point towards novel ADHD-related genes. By combining haplotype analysis and whole exome sequencing (WES), we aim to identify genes carrying rare and/or common variants that contribute to ADHD etiology. Three German pedigrees with multiple (adult) ADHD-affected members were used in this study, in which linkage analysis was carried out on the major part of the individuals and WES data from two or more patients were obtained to detect shared genetic variants. Genotyping was performed on Affymetrix Genome-Wide Human SNP Array 6. 0 and SNP data was used to identify haplotype blocks segregating with disease. WES was carried out using Agilent SureSelect All Exon 50Mb Target Enrichment kit and single-end sequencing on the 5500xl SOLiDTM System. Linkage analyses were performed using a Multi-Point analysis implemented in Superlink online SNP 1. 1, for each family independently and by combining them. All regions with a maximum LOD score higher than 2 were selected as a candidate regions. Genes in these regions containing rare and common genetic variants shared among all sequenced individuals were selected as candidate genes. A common problem in multifactorial diseases is that the effect of individual markers is too weak to be detected; consequently, multiple genetic markers may be simultaneously analyzed to increase explained variance. Therefore, gene-set analyses were performed in an independent exome-chip genotyping dataset from 1846 adults with ADHD and 7519 controls (IMpACT consortium). The gene-set from one of the three families indeed showed a significant association with adult ADHD. As no individual gene/variant caused this finding, this suggests that the variants in several of the genes might act together in increasing ADHD risk in the pedigree and the population. The analysis strategy followed here might provide new knowledge for a better understanding of ADHD's genetic basis.

1159F

Novel L1 retrotransposon discovery in schizophrenia exomes. *C. E. Krebs[1], L. Olde Loohuis[2], R. S. Kahn[3], S. de Jong[2,4], R. A. Ophoff[1,2,3].* 1) Department of Human Genetics, UCLA, Los Angeles, CA; 2) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, UCLA, Los Angeles, CA; 3) Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands; 4) Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK.

Despite being a highly heritable disease, much of the genetic architecture of schizophrenia remains unknown. In line with a growing body of evidence that neuron-specific LINE-1 (L1) retrotransposition exists in brain and is increased in schizophrenia subjects relative to controls, we utilized an exome sequencing set of 19 schizophrenia pedigrees (n = 78 subjects) to look for novel *de novo* L1 retrotransposon insertions. Using the transposable element discovery tool Retroseq we were able to detect novel L1s within whole-blood exomes of the schizophrenia pedigrees (total = 2862, mean = 29. 8 per subject). The ability to call novel L1s directly depends on the depth of sequencing around the insertion site and indeed L1 count and overall average exome coverage were highly correlated (Kendall's rank correlation, p = 1. 202x10-5). Therefore, in order to properly compare L1 counts across subjects, we subsampled the aligned reads from each subject to normalize coverage at 60x. After repeating the Retroseq analysis on the normalized data, we found no difference between the novel L1 count of probands and unaffected siblings (Wilcoxon rank sum, p = 0. 337). As has been previously reported, L1 insertion sites were biased towards 3' UTRs and depleted in coding exons (χ^2, p = 2. 2x10-16), suggesting negative selection against highly disruptive insertions. Interestingly, as many as half of all novel L1s discovered may be shared between at least two different samples, with some occurring in as many as 78% of all samples, suggesting that there are many L1s that are common in the population but not present in the reference human genome. Future studies with larger sample sizes and more relevant tissue types will be needed to determine if indeed novel L1 retrotransposition plays a role in schizophrenia etiology.

1160W

Exome Sequencing Points to Roles of *De Novo* and Rare Transmitted Mutations in Bipolar Disorder. *N. Matoba[1,2], M. Kataoka[1,3], A. Kazuno[1], K. Fujii[1,4], K. Matsuo[5], JC. Roach[6], A. Takata[1], T. Kato[1].* 1) Lab. for Molecular Dynamics of Mental Disorder, RIKEN BSI, Wako, Saitama, Japan; 2) Dept. of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan; 3) Dept. of Child Neuropsychiatry, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan; 4) Dept. of Psychiatry, Dokkyo Medical University School of Medicine, Tochigi, Japan; 5) Div. of Neuropsychiatry, Department of Neuroscience, Yamaguchi University of Graduate School of Medicine, Ube, Japan; 6) Institute for Systems Biology, Seattle, Washington, USA.

Although numerous genetic studies have been conducted for bipolar disorder (BD), thegenetic architecture of BD remains elusive. Here we performed a trio-based exomesequencing study for BD to investigate the potential roles of *de novo* and transmittedmutations in the disease etiology. We identified 71 de novo point mutations in 79 BDprobands with a trend toward excess of protein-altering mutations when compared tocontrol subjects in published studies. We also suggest the roles of calcium signaling-relatedgenes from the analysis of variants transmitted to BD probands. We will further discussproperties of genes hit by *de novo* and rare transmitted mutations and potential rolesof these mutations in the pathophysiology of BD.

1161T

***CNTN6* mutations are risk factors for abnormal neurite outgrowth and autism spectrum disorders.** *o. Mercati[1,2,3], G. Huguet[1,2,3], A. Danckaert[4], G. André-Leroux[4], A. Maruani[5], M. Bellinzoni[4], L. Gouder[1,2,3], M. Mathieu[1,2,3], J. Buratti[1,2,3], M. Benabou[1,2,3], J. Van-Gils[1,2,3], A. Beggiato[5], M. Konyukh[1,2,3], JP. Bourgeois[1,2,3], C. Delépine[6], A. Boland[6], B. Régnault[7], M. Francois[8], T. Van Den Abbeele[8], AL. Mosca-Boidron[9], L. Faivre[9], Y. Shimoda[10], K. Watanabe[10], D. Bonneau[11], M. Rastam[12], M. Leboyer[13,14,15,16], C. Gillberg[17], R. Delorme[1,2,3,5], I. Cloëz-Tayarani[1,2,3], T. Bourgeron[1,2,3,16].* 1) Human Genetics and Cognitive Functions Unit, Institut Pasteur, Paris, France; 2) CNRS UMR 3571 : Genes, Synapses and Cognition, Institut Pasteur, Paris, France; 3) Université Paris Diderot, Sorbonne Paris Cité, Human Genetics and Cognitive Functions, Paris, France; 4) Institut Pasteur, Unité de Microbiologie Structurale, 75724, Paris, France; CNRS UMR 3528, 75724, Paris, France; Université Paris Diderot, Sorbonne Paris Cité, Microbiologie Structurale, 75724, Paris, France; 5) Assistance Publique-Hôpitaux de Paris, Robert Debré Hospital, Department of Child and Adolescent Psychiatry, Paris, France; 6) Centre National de Génotypage, Evry, France; 7) Eukaryote Genotyping Platform, Genopole, Institut Pasteur, Paris, France; 8) Assistance Publique-Hôpitaux de Paris, Robert; 9) Département de Génétique, CHU Dijon et Université de Bourgogne, Dijon, France; 10) Department of Bioengineering, Nagaoka University of Technology, Nagaoka, Niigata 940-2188, Japan; 11) Département de Biochimie et Génétique, Centre Hospitalier Universitaire, Angers, France; 12) Department of Clinical Sciences in Lund, Lund University, Lund, Sweden; 13) INSERM U955, Psychiatrie Génétique, Créteil, France; 14) Université Paris Est, Faculté de Médecine, Créteil, France; 15) Assistance Publique-Hôpitaux de Paris, Hôpital H. Mondor and A. Chenevier, Département de Psychiatrie, Créteil, France; 16) FondaMental Foundation, Créteil, France; 17) Gillberg Neuropsychiatry Centre, University of Gothenburg, Gothenburg, Sweden.

Contactin genes *CNTN5* and *CNTN6* code for neuronal cell-adhesion molecules that promote neurite outgrowth in the sensory-motor neuronal pathway. Deleterious mutations of *CNTN5* and *CNTN6* have previously been reported in individuals with autism spectrum disorder (ASD), but their role in the disorder remains unknown. In this study, we screened a cohort of patients with ASD and controls for the presence of deleterious variants in *CNTN5* and *CNTN6* genes. The patients showed enrichment of *CNTN6* deletions (6/1534 ASD vs. 1/8936 controls; P=6x10-5) and private *CNTN6* coding-sequence variants (9/212 ASD vs. 2/217 controls; P= 0. 03). Two *CNTN6* deletions were transmitted from fathers with ASD and one CNTN6P770L variant was *de novo*. *CNTN5* deleterious variants were also identified including a patient carrying five copies of *CNTN5*, but not significantly enriched in ASD. Using a co-culture cellular assay and protein structure modeling, we found that several CNTN5 and CNTN6 variants either decreased or increased neurite length and branching compared with wild-type proteins, and could induce changes in the local structure of the protein. Clinical investigation of the patients carrying CNTN5/6 variants showed that they suffer from hypersensitivity to sounds (hyperacusis) and display with subtle changes in wave latency within the auditory pathway. These results reinforce the hypothesis of abnormal neurite outgrowth in the pathophysiology of ASD and shed new light on the genetic susceptibility to abnormal sensory perception in ASD.

1162F

Identification of five obsessive-compulsive disorder genes utilizing animal models, evolutionary constraints and regulatory information. *H. Noh[1], R. Tang[1,2], E. Karlsson[1,3], R. Swofford[1], J. Flannick[4], C. O'Dushlaine[4], J. Johnson[1], M. Koltookian[1], D. Howrigan[4], E. Grünblatt[5,6], E. Andersson[7], D. Djurfeldt[7], P. Patel[8], G. Genovese[4], J. Moran[4], S. McCarroll[4], C. Pato[9], C. Hultman[10], G. Hanna[8], S. Stewart[11], J. Knowles[9], C. Mathews[12], C. Rück[7], S. Walitza[5,6], D. Cath[13], J. Scharf[14], G. Feng[2,4], K. Lindblad-Toh[1,15].* 1) Vertebrate Genome Biology, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 2) McGovern Institute for Brain Research at MIT, Cambridge, MA, USA; 3) Bioinformatics and Integrative Biology, UMass Medical School, Worcester, MA, USA; 4) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) University Clinic of Child and Adolescent Psychiatry, University of Zurich, Zurich, Switzerland; 6) Neuroscience Center Zurich, University of Zurich and ETH Zurich, Switzerland; 7) Dept. Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden; 8) Dept. Psychiatry, University of Michigan Health System, Ann Arbor, MI, USA; 9) Dept. Psychiatry & Behavioral Sciences, USC, Los Angeles, CA, USA; 10) Dept. Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 11) BC Mental Health & Addictions Research Institute, UBC, Vancouver, BC, Canada; 12) Dept. Psychiatry, UCSF, San Francisco, CA, USA; 13) Dept. Clinical and Health Psychology, Utrecht University, Utrecht, Netherlands; 14) Dept. Neurology, Massachusetts General Hospital, Boston, MA, USA; 15) Science for Life Laboratory, IMBIM, Uppsala University, Uppsala, Sweden.

Genome-wide association studies (GWAS) have identified many risk loci for mental disorders, yet the heritability is largely unexplained. It is now evident that numerous variants contribute to the susceptibility, hampering the identification of all the risk genes by typical GWAS and with attainable sample sizes. Here, we employ a systematic model-driven approach that reduces the genetic search space for obsessive-compulsive disorder (OCD). OCD manifests in repetitive behaviors and intrusive thoughts, affecting 1-3% of humans. It is highly heritable (h2=0. 45-0. 65), yet no genes have been confidently found. Natural canine OCD is an ideal model for human OCD. As in humans, canine OCD is multigenic, and manifests in repetition of normal behaviors, e. g. tail chasing and flank sucking derived from predation and suckling, respectively. Drug response and onset age are also equivalent. Moreover, the limited genetic background within breeds aids gene mapping, allowing us to identify GWAS loci using <100 cases. Murine models are less similar as it is monogenic. Still, neurological examination consistently implicates the cortico-striatal-thalamo-cortical (CSTC) circuit in compulsive grooming. We sequenced coding and noncoding regions of 608 genes from canine OCD GWAS and the CSTC circuit in ~600 DSM-IV OCD and ~600 healthy individuals of European ancestry. We performed gene-based tests of all, coding and regulatory variants separately, and found five significantly associated genes for OCD. We found *NRXN1* and *HTR2A* with excess coding variants, *CTTNBP2* and *REEP3* with excess regulatory variants, and *LIPH* with overall excess variants, suggesting different genes harbor different types of OCD variation. Overall, we found a burden of variants in synaptic adhesion and maintenance pathways for OCD. Gel shift assays on candidate variants in *CTTNBP2* and *REEP3* identified six that alter DNA-protein binding in neuroblasts. We are now validating the associations for 90 SNPs in an independent cohort. Our method utilizing evolutionary and regulatory information is being assessed on the whole genome sequence (WGS) from other disorders for broader application. In conclusion, coding and noncoding variants are both critical to OCD, but may affect different genes. Our work provides a powerful framework for coding and noncoding variant analysis that is vital to WGS data. Lastly, our comparative approach, guided by natural and artificial models, identified novel genes and pathways for OCD.

1163W

Exome Sequencing in Multiplex Families with Bipolar Disorder Reveals Circadian Signaling Pathways. *S. Ramdas[1], A. B. Ozel[2], K. W. Walker[2], M. Burmeister[2], J. Z. Li[2].* 1) Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI.

Bipolar disorder (BPD) is a psychiatric disease characterized by alternating manic and depressive episodes. Twin and family studies have shown that BPD is highly heritable, with a heritability estimate of greater than 0. 7. However, extensive genetic and phenotypic complexity remains a major barrier to identifying its genetic basis. We hypothesize that multiplex BPD families manifest a highly penetrant subset of bipolar cases and may transmit one or more high-impact coding variants that alter the function of key neurodevelopmental or neural signaling genes. In this study we focus on 34 multi-generational multiplex families in the NIMH repository with BPD Type I. We have genotyped all available members in these families, and performed exome sequencing in 90 affected individuals representing first cousins or more distant relatives, aiming to identify functional variants that are rare in the general population but are shared by affected relatives. In the event of locus heterogeneity, linkage analysis has reduced power due to ineffective accrual of positional signals across families. However we test the hypothesis that there are shared functional perturbations across families and that such evidence accumulates quickly even in the presence of locus heterogeneity. We first filter variants by population frequency, functional annotation, frequency in control databases, and segregation in each family to obtain a list of 302 putative genes, or an average of 9 genes per family. We then analyze recurrence patterns of genes across families and find some previously implicated genes mutated across multiple families, including PCNT and TNXB. Pathway analysis of the 302 genes reveals enrichment of the circadian rhythm signaling pathway, which has been previously implicated in BPD. We also analyze publicly available brain expression data to identify co-expressed pairs of genes in different brain regions at different stages of development. Among our 302 genes, we find a suggestive enrichment of co-expressed gene pairs in infancy. Our results point to a functional convergence in the genetic signal across families, though further studies are needed to narrow in on the causal genes in each family.

1164T

Unique Psychiatric and Non-Psychiatric Phenotypic Characterizations in relation to Copy Number Variants (CNVs) in a Schizophrenia Cohort. *V. Sriretnakumar[1, 2], C. Zai[1, 4], M. Maciukiewicz[1, 4], J. So[1, 2, 3], J. Kennedy[1, 4].* 1) Neurogenetics Lab, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada; 3) The Fred A. Litwin Family Centre in Genetic Medicine, University Health Network and Mount Sinai Hospital, Toronto, Ontario, Canada; 4) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada.

Increasing evidence supports the significance of copy number variants (CNVs) in the genetic contribution to psychiatric illnesses, particularly schizophrenia (SCZ). This study utilizes a robust approach to uncovering genetic variants within the heterogeneity of SCZ by delineating correlations between CNV data and extensive phenotypic data in SCZ patients. Phenotypic and CNV data of 348 SCZ patients were collected from medical history records and Affymetrix SNP Array 6. 0 assay, respectively. The number of autosomal, X-linked and total CNV counts were plotted against phenotypic characteristics in probands and proband's family history. Independent t-tests, Fischer's exact test, Pearson/Spearman correlations, and Bonferroni multiple testing corrections were performed among the various categories to identify significant associations.

The presence of head injury in probands is associated with the number of autosomal CNV duplications (p=0. 0449) and deletions (p=0. 0490). Suicide attempt in probands was also associated with the number of autosomal CNV duplications (p=0. 0088), deletions (p=0. 0450), the total number of CNVs (p=0. 0274). Genetic associations of head injury and suicide attempt in probands did not reach statistical significance after correction for multiple testing. However, there are significant associations between suicide attempt in proband's family history and the number of autosomal CNV duplications (p=0. 0017), deletions (p= 0. 0016), and the total number of CNVs (p=0. 0130). Trends were also seen between X-linked CNVs and age of onset of SCZ in probands; history of substance abuse in proband's family history; and digestive system disorders in probands. All X-linked associations did not survive sex correction.

These results suggest a strong association between CNV burden and specific phenotypic presentations in the SCZ patient population. This is compatible with the ever-increasing number of microdeletions and microduplications found to be associated with neurodevelopmental disorders, and our findings may contribute to expanding the neuropsychiatric phenotypes associated with these genetic variants. Sample size and power will be increased by genotyping the remaining individuals in this cohort to detect smaller effect sizes. Further analyses will be undertaken to define specific CNVs and genes contained within the implicated CNV regions to better characterize potential genetic effects on the phenotypic presentation of SCZ patients.

1165F

Using exome sequencing to identify rare, *de novo* mutations in Tourette Syndrome families. *D. Yu[1,2], A. Huang[3], K. Samocha[1,2], V. Ramensky[3], M. Lek[1,2], L. K. Davis[4], D. MacArthur[1,2], N. J. Cox[6], N. B. Freimer[3], C. A. Mathews[5], B. M. Neale[1,2], G. Coppola[3], J. M. Scharf[1,2], Tourette Syndrome International Consortium for Genomics.* 1) Massachusetts General Hospital, Boston, MA; 2) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Semel Institute, University of California at Los Angeles, Los Angeles, CA; 4) University of Chicago, Chicago, IL; 5) Langley Porter Psychiatric Institute, University of California at San Francisco, San Francisco, CA; 6) Vanderbilt University, Nashville, TN.

Tourette Syndrome (TS) is a highly-heritable, childhood-onset neuropsychiatric disorder characterized both by neurological and psychiatric symptoms. Despite its strong familiality, identification of definitive TS susceptibility genes has been challenging. We and others have previously demonstrated that a subset of individuals with TS can harbor large, rare *de novo* deletions or duplications in regions of the genome previously associated with recurrent, pathogenic copy number variation in neurodevelopmental disorders that overlap phenotypically with TS, including ADHD and autism-spectrum disorder (ASD). Here, we examine the role of *de novo* coding variation in TS pathogenesis using exome sequencing in 180 TS parent-proband trios. Exome sequencing libraries were generated using the Agilent SureSelect Broad v1. 1 targeted capture kit for 149 parent-proband trios at the Broad Institute and Nimblegen EZ-Cap v3. 0 for 37 trios at UCLA. Paired-end sequencing was conducted using the Illumina HiSeq at both institutions to a minimal read depth of 90% at 10x. Sequence processing for all samples was conducted at the Broad Institute using Picard, and joint calling of all 558 samples was performed using GATK 3. 1. After removing poor quality samples, 180 trios remained for analysis. *De novo* variants were identified and evaluated using a probability of de novo (Prob_dn) metric based on allele balance, allele frequency, read depth, Phred score, and expected *de novo* rate in the generation population. Subsequent QC filters were applied on proband allele balance (AD>0. 2 and <0. 8), allele count (AC<2 if AD<0. 3 or AD>0. 7), andProb_dn (>0. 5). Nineteen *de novo* loss of function (LoF) mutations were identified after QC, including 8 SNV and 11 indel *de novo* LoFs which were validated by Sanger sequencing; this *de novo* rate was not significantly higher than expectation (0. 106 per trio vs. 0. 087 per trio, p=0. 24). 122 missense *de novo* variants (120 SNVs and 2 indels) and 51 synonymous SNV *de novo* variants were called, and no enrichment above expectation was found. Enrichment of *de novo* LoF and missense variants from the 180 TS trios was examined for 107 previously reported ASD genes, FMRP interacting genes, genes enriched at the post-synaptic density, and genes under high evolutionary constraint, with no significant enrichment observed. Combined analysis with a parallel sample of 216 independent TS trio exomes sequenced at Yale University will be presented.

1166W

Novel gene identification in familial Alzheimer's disease. *J. Rehker[1], R. Nesbitt[1], A. Patowary[1], B. K. Martin[2], E. A. Boyle[2], J. Shendure[2], Z. Brkanac[1].* 1) Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA; 2) Department of Genome Sciences, University of Washington, Seattle, WA.

Alzheimer's disease (AD) is the most common cause of dementia in US, affecting approximately 5% of individuals older than the age of 70. In addition to highly penetrant rare variants in APP, PSEN1 and PSEN2, rare variants in TREM2, PLD3 and ABCA7 have recently been recognized as contributing to AD. The ApoE4 allele, which is relatively common, is a significant contributor to AD. Whole genome association studies have identified approximately a dozen common low-risk genes associated with the disease that collectively have impact smaller than APOE4. It is likely that additional rare variants contributing to AD remain to be found. To identify additional rare variants contributing to AD, our approach was focused on 14 families from NIA and NIMH AD collections with 3 or more affected individuals. For each family, one or more of the affected subjects were second-degree relatives. Our analytical approach was tiered, with the aim to evaluate private variants first, followed by analysis of rare variants. As a consequence of this approach we obtained exome sequence on the most distantly related cases in each family and genomewide genotyping for all cases. We previously reported the analysis of private variants shared among cases (J. Rehker et all. , Rare highly penetrant variants of late onset Alzheimers disease; Program #295. ASHG annual meeting, October 25 2013, Boston, MA). In order to expand our analysis to rare variants of 1% or less, we employed linkage analysis to narrow regions shared by all affected cases. We have focused on protein changing variants (missense, frameshift, stop, and splice). For genotyping in families we used Illumina Human Exome arrays that include a backbone of linkage markers. The linkage analysis and haplotyping were performed using MERLIN and HAPLOPAINTER. We have used a tiered approach to identify candidate variants. If all affected subjects did not share any rare variants, we excluded subjects with APOE4/4 genotype and subjects without pathology confirmation of AD diagnosis. This resulted in the identification of 10 novel candidate genes. In order to provide a statistical support for involvement of the identified candidate genes with AD, we are performing targeted sequencing and gene based association analysis of 980 familial cases and elderly controls. The results of our linkage and association analyses will be presented at the conference.

1167T

Mutations in *PTRH2* cause infantile-onset multisystem neurologic, endocrine, and pancreatic disease (IMNEPD). *C. Hu[1,2], M. Matter[3], L. Issa-Jahns[4], S. Picker-Minh[4], M. Jijiwa[3], N. Kraemer[4], L. Musante[2], M. Vega[3], O. Ninnemann[4], D. Schindler[5], N. Damatova[5], K. Eirich[5], M. Sifringer[4], S. Schrötter[4], B. Eickholt[4], L. Heuvel[6], C. Casamina[3], G. Stoltenburg-Didinger[4], H. Najmabadi[7], K. Kahrizi[7], H. Ropers[2], T. Wienker[2], C. Hübner[4], A. Kaindl[4].* 1) Guangzhou Women and Children's Medical Center, Guangzhou, China; 2) Max-Planck Institute for Molecular Genetics, Berlin, Germany; 3) The University of Hawaii Cancer Center, Honolulu, Hawaii; 4) Charite-Universitàtsmedizin Berlin, Berlin, Germany; 5) University of Würzburg, Würzburg, Germany; 6) Radboud University Medical Center, Nijmegen, The Netherlands; 7) University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.

We characterized a consanguineous family of Yazidian-Turkish descent with IMNEPD. The two affected children suffer from intellectual disability, postnatal microcephaly, growth retardation, progressive ataxia, distal muscle weakness, peripheral demyelinating sensorimotor neuropathy, sensorineural deafness, exocrine pancreas insufficiency, hypothyroidism, and show signs of liver fibrosis. We performed whole-exome sequencing followed by bioinformatic analysis and Sanger sequencing on affected and unaffected family members. We identified a homozygous frameshift mutation in the previously not disease-associated peptidyl-tRNA hydrolase 2 (*PTRH2*) gene. *PTRH2* encodes a primarily mitochondrial protein involved in integrin-mediated cell survival and apoptosis signaling. We show that *PTRH2* is highly expressed in the developing brain and is a key determinant in maintaining cell survival during human tissue development. Moreover, we link *PTRH2* to the mTOR pathway and thus the control of cell size. The pathology suggested by the human phenotype and neuroimaging studies is supported by analysis of mutant mice and patient fibroblasts. The diagnosis of this family was recently substantiated by another three cases with similar manifestations and molecular etiology.

1168F

Modeling the association between loss-of-function *GBA1* mutations and synucleinopathies with *in vivo* mouse models. *N. Tayebi[1], A. Gonzalez[1], B. Giasson[2], B. Berhe[1], M. Brooks[2], N. Trivedi[3], A. Elkahloun[4], R. Petralia[5], E. Sidransky[1].* 1) Molecular neurogenetics section/MGB/NHGRI, National Institutes of health, Bethesda, MD; 2) Department of Neuroscience, Center for Translational Research in Neurodegenerative Diseases, University of Florida, Gainesville, FL, 32610, USA; 3) Computational and Statistical Genomics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland, USA; 4) Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, NIH, Bethesda, Maryland, USA; 5) Advanced Imaging Core, National Institute on Deafness and Other Communication Disorders, NIH, Bethesda, Maryland, USA.

Gaucher disease (GD), the recessively inherited deficiency of glucocerebrosidase (GCase) is caused by mutant *GBA1.* Missense and loss-of-function mutations in *GBA1* are common genetic risk factors for Parkinson disease (PD). To model this association, we crossed transgenic mice hemizygous (hemi) or homozygous (homo) for human mutant A53T α-synuclein (A53Tα-syn) with mice carrying a loss-of-function null *gba* allele (wt/χMGC). The resulting wt/χMGC//A53Tα-syn mice overexpressing A53Tα-syn and carrying a null *gba* allele, were compared to control groups (wt/wt, wt/A53Tα-syn and wt/χMGC) to elucidate the effect of GCase deficiency on the progression of neurological impairment in PD. Mice were followed for up to 2 years and were euthanized with their respective controls when symptomatic**.** Weight loss was the initial sign of illness, followed by arching of the back, impaired axial rotation, and limb paralysis. Survival analysis of 84 mice showed that wt/χMGC//A53Tα-syn hemi and homo mice exhibit symptoms on average 9. 76 weeks and 4. 04 weeks earlier than A53Tα-syn hemi and homo controls, respectively (p-values 0. 023, 0. 0030). Further, the wt /χMGC//A53Tα-syn mice had a more rapid disease progression than A53Tα-syn mice (p-value <0. 0001). Quantification of α-syn aggregates in the pons of 17 mice showed no significant effect of carrying the χMGC allele. GCase protein levels and activity in wt/χMGC//A53Tα-syn mice reflected haploinsufficiency of *gba,* but no added effect due to A53Tα-syn was seen and α-syn levels in A53Tα-syn mice with and without χMGC were similar. To identify other genes involved in this association, a whole transcriptome array was performed on brains of the crossed lines, revealing differences in PRNP (p-value: 3. 22*10-12; fold change: 3. 6), SERPINA3 (1. 71*10-6; 4. 7), and MMRN1 (1. 53*10-7; 4. 7), confirmed by real-time PCR. Interestingly, H2-T10 (3. 37*10-12; 6. 98) was up-regulated in only wt/χMGC//A53Tα-syn mice, suggesting its contribution to the earlier and more severe phenotypes observed. This gene is homologous to *MHC-1* in humans. Evaluation of LIMP-2, PSAP, P62 and Cathepsin-D RNA expression showed no differences. Ultrastructural and immunohistochemical evaluations of striatum, substantia nigra and midbrain indicate more microgila activation and astrogliosis in wt/χMGC//A53Tα-syn mice. These data suggest that haploinsufficiency of GCase impacts disease onset and severity in the A53Tα-syn model, and may be related to acute brain inflammation.

1169W

Transcriptomic analysis reveals convergent molecular pathways during the development of neurons derived from patients with idiopathic autism. *B. A. DeRosa[1,2,3], K. C. Belle[1,2,3], C. Garcia-Serje[1,3], H. N. Cukier[1,3], J. M. Lee[1,3], M. L. Cuccaro[1,2,3], J. P. Hussman[4], J. M. Vance[1,2,3], M. A. Pericak-Vance[1, 2, 3], D. M. Dykxhoorn[1,2,3].* 1) John P. Hussman Institute for Human Genomics; 2) Dr. John T. MacDonald Foundation Department of Human Genetics; 3) University of Miami Miller School of Medicine, Miami, FL; 4) Hussman Institute for Autism, Baltimore, MD.

Recent studies have shown that genes harboring autism spectrum disorder (ASD) risk loci are highly enriched in sets of genes expressed during early neocortical development and genes encoding proteins that function in specific biological pathways involving regulation of transcription, chromatin remodeling, cell adhesion, signaling complexes, and synapse function. These findings raise the possibility that common molecular etiologies underlie pathogenesis in subsets of individuals with idiopathic autism. Patient-specific induced pluripotent stem cells (iPSCs) present a unique opportunity to examine the hypothesis that heterogeneous ASD loci converge on specific molecular pathways during early neuronal development. Therefore, we generated patient-specific iPSC lines from 6 unrelated ASD individuals with rare variants identified through exome sequencing in extended multiplex families. The patient-specific iPSCs, in addition to iPSCs derived from unrelated control individuals, were differentiated into cortical neurons and transcriptome analysis using RNA-seq was performed on the neurons at three time points over a 135 day time course of their *in vitro* development. Pathway and gene ontology (GO) analysis was performed on the set of differentially expressed (DE) genes identified in the transcriptome data. Our results implicate disturbances in the regulation of transcription, WNT signaling, chromatin remodeling, cell adhesion and migration, and synapse development across all time points analyzed. Interesting patterns emerge from the differences in expression between analyzed time points. To highlight some of these findings, specific pathways enriched in early neurons (day 35 *in vitro*) include those involving WNT/MET signaling ($p = 4.95\times10\text{-}36$), and collagen catabolism ($p = 7.41\times10\text{-}09$). DE genes in midpoint neurons (day 85 *in vitro*) map to pathways involving cell migration ($p = 1.34\times10\text{-}06$) and GABAergic neuron signaling transmission ($p = 2.86\times10\text{-}09$). In later time point neurons (day 135 *in vitro*), DE genes are enriched in pathways involving WNT-mediated axon guidance ($p = 4.71\times10\text{-}05$), calcium signaling ($p = 8.92\times10\text{-}20$), and chromatin remodeling ($p = 1.49\times10\text{-}24$). These results provide functional evidence that common molecular etiology underlies pathogenesis in ASD and suggest a temporal order to the disturbance of cellular functions during neuronal development.

1170T

Dynamic regulation of RNA-editing in human brain development. *T. Hwang[1,2], J. H. Shin[2], D. R. Weinberger[2].* 1) Johns Hopkins University, Baltimore, MD; 2) Lieber Institute for Brain Devleopment, Baltimore, MD.

RNA-editing is the co-transcriptional process perturbing RNA sequences. In humans, the most prevalent form of RNA-editing is the single nucleotide change of adenosine to inosine ("A-to-I editing"), which is known to play a critical role in ion channel genes such as GRIA2 (GluR2) and HTR2C (5-HT2CR). In this study, we pursued comprehensive RNA editing profiles in human brain, the tissue with the most active RNA-editing activity, by applying next generation sequencing technologies to post-mortem tissues spanning from fetus to adult life. We have also developed robust computational tools to identify RNA-editing events. More than 50,000 reliable RNA editing events were identified and random examples were validated with both of DNA and RNA sequencing, including known and novel RNA-editing in protein-coding sequences and untranslated regions of mRNA. The genome-wide landscape of RNA-editing also reveals dynamic regulation of RNA-editing, which consists of three patterns: consistently high-edited, consistently low- edited and increasingly-edited across cortical development. Surprisingly, the dynamic pattern is brain-specific and accompanies temporally the growth and elaboration of cortical circuitry. The secondary structures of RNA and the temporal pattern of A-to-I RNA editing enzyme, ADAR (adenosine deaminase acting on RNA) expression determine cis- and trans- regulatory mechanisms of RNA-editing patterns, respectively. Finally, we confirm that developmentally-regulated RNA-editing is specifically involved in the development of cortical layers through glutamate signaling and vesicle/organelle membrane genes. Overall, our findings provide new insight into the understanding of RNA-editing, one of mechanisms expanding sequence diversity, in human brain development.

1171F

Longitudinal RNA-seq analysis of Parkinson's disease patient-derived dopaminergic neurons. *K. C. Belle[1 2 3], A. Mehta[1], K. Nuytemans[1 3], B. A. DeRosa[1 2 3], J. M. Vance[1 2 3], D. M. Dykxhoorn[1 2 3].* 1) Human Genetics and Genomics, Univeristy of Miami, Miami, FL; 2) Dr. John T. MacDonald Foundation Department of Human Genetics; 3) University of Miami Miller School of Medicine, Miami, FL 33136, United States.

Parkinson's disease (PD) is a progressive neurodegenerative disease characterized by the loss of dopaminergic neurons in the substantia nigra. While disease symptoms present themselves later in life, and progress in severity, it remains an open question when the antecedent dysfunction that leads to cell death and loss of motor function takes place. To initiate our investigation of this question, we differentiated induced pluripotent stem cells (iPSCs) derived from two Parkinson's patients (α-synuclein triplication and idiopathic patient) and two aged-matched controls, into midbrain dopaminergic neurons (mDAns) and performed RNA sequencing at different time points. *W*e hypothesized that as the differentiation of these lines progresses, a convergence on to common pathways and networks that drive the pathogenesis of PD phenotypes would be observed. RNA from differentiated cultures was harvested at three time points: 20 days (d20), 45 days (d45), and 120 days (d120) post-differentiation. RNA sequencing data was analyzed on a case vs control basis at each time point generating a list of differentially expressed genes that was then analyzed for enrichment in biological pathways or functions through gene set enrichment analysis *(GSEA).* GSEA of the d20 time point sequence noted differential expression between cases and controls in the neuronal developmental pathways, including Sonic Hedgehog, Notch, and TGF-b, all of which are involved in the differentiation of mDAns. In our next time point, d45, we observed similar differences in developmental pathways with the addition of the cytokine-cytokine interaction pathway. Initial pathway analysis of the differentially expressed genes between cases and controls at time point d120 revealed the involvement of multiple disease-related pathways with nominal significance; including mitochondrial protein targeting, mitochondria signaling of apoptosis, cell structure disassembly during apoptosis, and caspase cascade in apoptosis. These results suggest that next generation sequencing of RNA from iPS neuronal cultures has the potential to help understand how genetic variation associated with PD affects function and leads to disease. Additional samples and investigation of older iPS culture times are currently underway.

1172W

Disruptions to the miRNA regulatory pathway may cause an increased rate of schizophrenia in individuals with 22q11. 2 DS. *W. Manley[1], S. Siecinski[1], S. Ryan[1], V. Coulibaly[1], S. Buyske[1,2], L. M. Brzustowicz[1].* 1) Department of Genetics, Rutgers University, Piscataway , NJ; 2) Department of Statistics & Biostatistics, Rutgers University, Piscataway, New Jersey.

Schizophrenia is a complex and poorly understood disease caused by the interplay of many environmental and genetic factors. The 22q11. 2 deletion syndrome (DS) is a disorder that is caused by the microdeletion of part of chromosome 22 leading to a 25-fold greater chance of developing schizophrenia in affected individuals versus the general population. The missing 22q11. 2 region contains DGCR8, which is required for the initial step of miRNA biogenesis. However, the 22q11. 2 deletion itself is not directly the cause of schizophrenia, since 75% of individuals with this deletion do not develop the disease. We hypothesize that the 22q11. 2 deletion increases the risk of schizophrenia through alterations in miRNA regulatory networks via depletion of several miRNAs. This may serve as a protective buffer against the accumulation of deleterious mutations at other schizophrenia risk loci. We have derived human neural stem cells (NSCs) from individuals with the 22q11. 2 deletion and schizophrenia. The NSCs were generated from iPSCs using Neural Induction Media (Life Technologies). In order to ensure the presence of the 22q11. 2 deletion, a FISHprobe for 22q11. 2 (TUPLE) was used (Cell Line Genetics). Also, the levels of DGCR8 gene expression were quantified using the 7900HT Fast Real-Time PCR System (Applied Biosystems) along with Taqman Gene Expression Assays and Copy Number Assays to ensure DGCR8 reduction in 22q11. 2 DS NSCs relative to otherwise healthy control NSCs. We have characterized disruptions to the miRNA regulatory network in the NSC lines using Taqman Array Human Microarray Cards Version 3. 0. Here we will present the miRNAs that we have identified to be differentially expressed in otherwise healthy control NSCs versus NSCs with 22q11. 2 DS and schizophrenia. Our data shows that of the 377 tested miRNAs, nearly 30% have at least a 2 fold expression change in deletion samples. We predict that these miRNAs could be involved in the elevated risk of developing schizophrenia in individuals with 22q11. 2 deletion syndrome versus the general population. Future studies will focus on determining the biological targets of these miRNAs using miRNA target prediction software.

1173T

Neurodevelopmental disorders linked to damages in ARX-KDM5C path. *A. Padula[1], L. Poeta[1], A. Ranieri[1], C. Shoubridge[2], C. Trimarco[1], J. Gecz[2], J. Dupont[3], C. Schwartz[4], H. vanBokhoven[5], M. G. Miano[1].* 1) Institute of Genetics and Biophysics, CNR, Naples, Naples, Italy; 2) Department of Paediatrics, University of Adelaide, South Australia 5006, Australia; 3) Hospital de Santa Maria, Lisbon 1649-035, Portugal; 4) Greenwood Genetic Center, Greenwood, South Carolina, 29646, USA; 5) Radboud University Nijmegen Medical Centre, 6500 HB, Nijmegen, The Netherlands.

Studying molecular convergence in neurodevelopmental disorders caused by mutations in specific disease-related genes permits us to define druggable molecular pathways. The purpose of our study is to assess the degree of damage associated with the ARX-KDM5C interaction and to establish a correlation between similar phenotypes and same cellular functions. Mutations in Aristaless-related homeobox gene (*ARX*), a homeotic transcription factor with a key role in interneuron maturation, have been found in a spectrum of X-chromosome phenotypes including cortical malformations, chronic Epilepsy and X-Linked Intellectual Disabilities (XLID). About Lysine-specific demethylase 5C (*KDM5C*), its mutations have been reported as an important cause of XLID. Its protein is a histone demethylase acting as transcriptional repressor during brain development. Here we summarize *in vitro* functional analysis of two classes of ARX mutations by studying the impact on the stimulation of *KDM5C*, already reported by us as an ARX disease-target gene: 1. PolyAlanine elongations affecting the first and the second PolyA tracts, frequently found in patients with Epilepsy or XLID; and 2. Missense mutations clustered in the paired-type homeodomain (HD) found in patients with Lissencephaly with abnormal genitalia (XLAG). We have proven that PolyAlanine elongations are partial loss-of-function mutations that impair the activation of *KDM5C* transcription; while HD missense mutants are loss-of function alterations, which abolish the transcriptional activity attributable to the WT protein. Moreover, in patient-derived LCLs mutated in the ARX-KDM5C path, we have found a KDM5C reduction in coupling with a global increase of H3K4me3 signalling, potentially due to a compromised KDM5C activity. Since H3K4me3 is the hallmark of open chromatin, ARX-dependent KDM5C defects could compromise chromatin remodelling. Taken in consideration these results, we propose a *"fault-disease model"* showing that the degree of spectrum of KDM5C defects correlate with the severity of the neurophenotypes associated with ARX mutations. Starting from these data, we reason that modelling ARX and KDM5C defects in the identical stem cell line might allow us to assess the degree of molecular convergence caused by altered dosage of these two disease-related genes and the validation of potential drugs that could compensate KDM5C-H3K4me3 deregulation.

1174F

Human neuronal stem cell line validated as *in vitro* genomic model for studying neurodevelopmental disorders. *A. P. S. Ori[1], M. H. M. Bot[1], R. T. Molenhuis[2], R. A. Ophoff[1,3].* 1) Center for Neurobehavioral Genetics, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, California, USA; 2) Department of Translational Neuroscience, Brain Center Rudolf Magnus, University Medical Center Utrecht, The Netherlands; 3) Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California, USA.

Large-scale genetic studies of neuropsychiatric disorders have identified hundreds of susceptibility alleles. The functional characterization and elucidation of pathways underlying these diseases present a major next challenge, especially for neurodevelopmental disorders such as schizophrenia and autism spectrum disorder. Therefore in order to address this, we established and examined an *in vitro* model to study neurodevelopment that allows for standardized gene manipulation and functional characterization using genomic approaches. We differentiated H9 human neuronal stem cells towards mature neurons over a period of 30 days and collected gene expression profiles at seven time points in triplicates. Using a machine learning approach we demonstrate that the observed *in vitro* gene expression profiles significantly match *in vivo* human neurodevelopmental stages and cellular laminae of human cortical regions. Using an empirical Bayes approach, we identified 25% of genes to be significantly differentially expressed during the 30 days of development. By subsequent fuzzy c-means clustering with bootstrapping, 10 high confidence clusters of genes with distinct expression patterns were distinguished. Biological annotation of these clusters highlight specific pathways and mechanisms important for neuronal differentiation. As expected, clusters representing neuronal processes show a clear increasing expression pattern over time while cell cycle and DNA replication processes show the opposite. Additional clusters with distinctive patterns are enriched for transcription factors, DNA binding proteins, and immune and defense response processes. The latter cluster highlights potential limitations of known pathway databases as genes can have multiple functions depending on time and location of expression and unknown functions cannot be annotated. Finally, genes differentially expressed over time are significantly enriched for genes associated to psychiatric neurodevelopmental disorders, such as schizophrenia (by permutation, P = 0. 048). Taken together, our results validate *in vitro* modeling of human *in vivo* neurodevelopment and identify key biological pathways and overall emphasize its relevance to study disease mechanisms.

1175W

BRAINCODE: How does the Human Genome Function in Specific Brain Neurons? *X. Dong[1], Z. Liao[1], D. Gritsch[1], B. Guennewig[2], Y. Hadzhiev[3], Y. Bai[1], G. Liu[1], B. Zheng[1], C. Blauwendraat[4], Q. Hu[1], C. H. Adler[5], J. C. Hedreen[6], M. P. Frosch[7], R. Faull[8], P. T. Nelson[9], J. Locascio[1], C. Vanderburg[7], P. Rizzu[4], A. A. Cooper[2], P. Heutink[4], T. G. Beach[10], F. Mueller[3], J. S. Mattick[2], C. R. Scherzer[1].* 1) Harvard Medical School and Brigham & Women's Hospital, USA; 2) Garvan Institute of Medical Research, Australia; 3) University of Birmingham, UK; 4) German Center for Neurodegenerative Diseases, Germany; 5) Mayo Clinic, USA; 6) McLean Hospital, USA; 7) Massachusetts General Hospital, USA; 8) University of Auckland, New Zealand; 9) University of Kentucky, USA; 10) Banner Sun Health Research Institute, USA.

The human brain comprises ~86 billion neurons whose function is central to human biology. How does the human genome program high performing neurons and neural networks in response to experience? What subprograms does the genome express in physiologically and morphologically distinct brain cells? The goal of the **BRAIN C**ell encycl**O**pe**D**ia of transcribed **E**lements Consortium (**BRAINCODE**) is to provide a comprehensive map of actively transcribed elements, both protein-coding and non-coding, from specific cell types, not in culture, but directly isolated from human brains. Going beyond traditional mRNA sequencing, *all* polyadenylated *and* non-polyadenylated transcripts over 50bp were ultra deeply sequenced using ribo-depleted total RNA from 50,000 neurons laser-captured from more than 130 human post-mortem brains yielding 23 terabytes of reads. Three prototypical neuron types, dopamine neurons, pyramidal neurons, and Betz cells, were prioritized because of their key biological roles and differential vulnerability to important neurodegenerative diseases such as Parkinson's or Alzheimer's disease. Genetic variation between individuals was examined for correlation with differences in transcribed sequences to identify genomic regions that influence whether, how, and how much a transcript is expressed in specific cell types in human brains. Initial results indicate a vast universe of annotated and novel non-coding RNAs expressed in brain cells and suggest a more diverse and much more complex transcriptional architecture than previously imagined. Support: NINDS U01 NS082157; U. S. Department of Defense; Michael J. Fox Foundation; U24 NS072026; P30 AG19610; P30 AG028383; NHMRC 631668.

1176T

The expression of long non-coding RNA *ANRIL* do not affected by Vitamin D treatment in Multiple Sclerosis patients. *M. Pahlevan Kakhki, M. Behmanesh.* Tarbiat Modares University, Tehran, Tehran, Iran.

Multiple sclerosis (MS [MIM 126200]) is an inflammatory disease in the central nervous system (CNS) which cause some disabilities in the young adults. Many family and twin studies showed that environmental and genetics factors are involved in the pathogenesis of MS. Antisense noncoding RNA in the *INK4* locus (*ANRIL*[MIM 613149]) gene is located at chromosome 9p21 which its product is a functional RNA molecule that interacts with polycomb repressive complex-1 (PRC1) and -2 (PRC2), leading to epigenetic silencing of other genes. This region is a significant genetic susceptibility locus for cardiovascular disease, and has also been linked to a number of other pathologies, including several cancers, intracranial aneurysm, type-2 diabetes, periodontitis, Alzheimer's disease, endometriosis, frailty in the elderly, and glaucoma. Here we are tried to assess the expression level of *ANRIL* in RRMS patients after 8 weeks treatment with vitamin D. 30 Iranian RRMS patients were clinically diagnosed according to McDonald's criteria. Also, we selected 30 age and gender matched controls. All patients received 50,000 IU vitamin D3 every week as an intra-muscular injection for 8 weeks. The vitamin D levels were estimated by 25-hydroxyvitamin detection kits of ELISA before and after supplementation. Five ml whole blood were obtained and PBMCs harvested using Ficoll solution, RNA extracted by Trizol reagent, and real time PCR was performed by Step One ABI system using specific primers. ΔΔct method was used for the analysis of gene expression results. The investigation was performed in keeping with the Helsinki declaration on research with human participant. Also, this trial was approved by the institutional ethical committee and was registered with Iranian Registry of Clinical Trials (ID: IRCT2014011216181N1) and medical ethics committee of Tarbiat Modares University. Our data showed that the expression level of *ANRIL* does not affected by the treatment of vitamin D after 8 weeks (p=0. 9). Moreover, we don't find any differences in the expression level of *ANRIL* gene in RRMS patients before and after vitamin D treatment (p=0. 217 & p= 0. 226 respectively) in comparison with healthy controls. In conclusion, based on our best knowledge, this is the first report which evaluate the expression level of long non coding RNAs specially *ANRIL* in multiple sclerosis patients under vitamin D treatment which is under elaborate investigation in our laboratory.

1177F

Oligogenic inheritance in families with amyotrophic lateral sclerosis and frontotemporal dementia. *S. M. K. Farhan[1], M. J. Strong[2], R. A. Hegele[1].* 1) Biochemistry, Robarts Research Institute, Western University, London, Ontario, Canada; 2) Department of Clinical Neurological Sciences, Western University, London, Ontario, Canada.

Amyotrophic lateral sclerosis (ALS) is a devastating neurodegenerative disease that currently affects ~3,500 people in Canada and has a global prevalence of >500,000 cases. Although ALS is classified as a rare disease, it is the most common motor neuron disease in adults and is characterized by degeneration of motor neurons, which often leads to progressive weakening of limb, bulbar, and respiratory muscles and is ultimately fatal, typically within 3-5 years of onset. There are two forms of ALS: familial and sporadic, which differ in their etiology but are clinically indistinguishable. Approximately, 5-10% of ALS cases are familial however, this number seems to be growing as more pedigrees are studied. We have independently ascertained four large, unique families with familial ALS and FTD across five generations. In one family, affected individuals present with primarily ALS while in the others, ALS is accompanied with FTD and in some individuals, severe psychiatric anomalies such as agitation, hypersexuality, and psychosis. In one family, five distinct phenotypes are observed: ALS, FTD, Parkinson's disease, Alzheimer's disease, and myelopathy with degenerative disc disease. A full family history, blood and when available, brain and spinal cord samples were collected. All participating individuals were assessed using a screening neurological examination and the Montreal Cognitive Assessment. Because next-generation sequencing (NGS) allows for efficient genetic variant detection, we have used a NGS based assay enriched for neurodegenerative disease genes to identify variants contributing to the phenotypes observed in these families. We have identified several, distinct mutations in these families that may explain the spectrum of neurodegenerative phenotypes observed. We are currently investigating whether these genetic variants work synergistically to give rise to the phenotype or modify large-effect mutations consistent with an oligogenic inheritance disease model. Finally, we plan to study some of the genetic variants in biological systems to recapitulate aspects of the disease providing further evidence for gene based disease expression. We present one of the first examples of an oligogenic inheritance disease model in familial ALS and FTD using a targeted NGS method. These approaches allow us to explain the complex inter-and intrafamilial variability in familial ALS and FTD and other overlapping neurodegenerative diseases.

1178W

Duplication of 1q21. 3-q22 in a child with developmental delay, dysmorphic features and obesity. *I. O. Focsa[1], A. Tutulan-Cunita[1], S. M. Papuc[1], M. Budisteanu[1,2], D. Le Tessier[3], A. Lebbar[3], J. -M. Dupont[3,4], A. Arghir[1].* 1) Medical Genetics Laboratory, Victor Babes National Institute of Pathology, Bucharest, Romania; 2) Prof. Dr. Alex. Obregia Clinical Hospital of Psychiatry, Bucharest, Romania; 3) APHP Hôpitaux Universitaires Paris Centre - Cochin, Paris, France; 4) Université Paris Descartes, Faculté de Médecine, Paris, France.

Objective: Array-based comparative genomic hybridization (aCGH) stands as the first tier genetic investigation in a wide range of human conditions including neurodevelopmental delay associated with dysmorphic features. Here, we describe the genetic anomalies identified in a 7 years-old female patient with global developmental delay, dysmorphic features and obesity. Material and methods: The patient is the only child of healthy, non-consanguineous parents, born at term. At presentation she exhibited: obesity - weight of 30 kg (>Pc 97), height - 109 cm (Pc 3), OFC - 48. 5 cm (Pc 3); dysmorphic features (low-set, malformed ears, hypotelorism, strabismus, thick eyebrows, short nose, hypoplastic ala nasi, long philtrum, thin upper lip, high palatal arch, micrognathia) global developmental delay, IQ 48; attention deficit; behavior problems with hyperphagia, hyperkinesia. Cranial CT was normal. aCGH was performed on Agilent platforms, according to manufacturer' recommendations. Results: The genomic profile obtained by aCGH (105K platform) revealed a duplication of ~3. 1 Mb at 1q21. 3-q22. The results were confirmed by a higher resolution aCGH (180K Agilent platform) and by FISH, with genomic boundaries at 153,099,623-156,258,645 (hg 19). The duplicated region contains over 150 genes, among which several OMIM genes associated with neurodevelopment: *GATAD2B* (OMIM 614998) is a transcriptional repressor with a high expression in brain; its heterozygous loss-of-function mutations were associated with autosomal dominant developmental delay. *EFNA1* (OMIM 191164), *EFNA3* (OMIM 601381), and *EFNA4* (OMIM 601380) are members of the ephrin (EPH) family that mediate development-related events, especially in the nervous system. *KCNN3* (OMIM 602983) gene encodes a membrane voltage-independent calcium-activated channel thought to regulate neuronal excitability. Conclusion: Few genomic imbalances of this region associated with cognitive impairment and dysmorphic features were reported till now. A similar duplication (2. 49 Mb) was reported previously by Gunn et al. (2011) in a patient with growth retardation and developmental delay; approximately 15 patients with similar size duplications that partially overlap the region detected in our patient, were reported in DECIPHER database. Our patient brings new evidence regarding 1q21. 3-q22 duplication and its impact upon neurodevelopment. AcknowledgementsThe present work was supported by the National Research Project PN 09. 33. 02. 03.

1179T

Targeted analysis of whole exome sequencing in early onset epilepsy. *I. Guella[1], M. K. Demos[2], S. E. Buerki[2], E. Toyota[2], S. Adam[3], M. Van Allen[3], G. Sinclair[4], G. Horvath[4], C. D. Van Karnebeek[4], P. Eydoux[5], D. Evans[1], M. McKenzie[1], A. Datta[2], C. Boelman[2], L. Huh[2], A. Michoulas[2], B. Bjornson[2], T. N. Nelson[6], M. Connolly[4], M. J. Farrer[1].* 1) Centre for Applied Neurogenetics, University of British Columbia, Vancouver, British Columbia, Canada; 2) Division of Neurology, Department of Pediatrics, University of British Columbia and BC Children's Hospital, Vancouver, BC, Canada; 3) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 4) Division of Biochemical Diseases, Department of Pediatrics, University of British Columbia and BC Children's Hospital, Vancouver, BC, Canada; 5) Department of Pathology and Laboratory Medicine, University of British Columbia and British Columbia Children's Hospital, Vancouver, British Columbia, Canada; 6) Department of Pathology and Laboratory Medicine, Children's and Women's Health Center of British Columbia, Vancouver, British Columbia, Canada.

Advances in genomic technologies, including targeted next generation sequencing and whole exome sequencing (WES), enables identification of pathogenic variants in 10–78 % of patients with unexplained epilepsy. The clinical impact is significant including earlier diagnosis and specific treatment interventions. Here, we report the results of WES on 50 patients with early onset epilepsy (≤ 5 years) of unknown cause, evaluating diagnostic yield and time to diagnosis. Between December 2014 - June 2015 WES was performed in 43 retrospective (epilepsy > 6 months) and 7 prospective (epilepsy < 6 months) patients. Detailed clinical data was abstracted in a REDCap database. WES was performed using Ion Proton™ System within 2 weeks of receiving samples. Average coverage was >80x across all samples. Reporting was restricted to 557 genes previously implicated in epilepsy. Putative causative mutations were validated by Sanger sequencing and by parental testing. A definite diagnosis was made in 13 patients (26%: 5 prospective; 8 retrospective). This included known pathogenic and novel variants in *SCN1A, ATP1A2, ALG13, STXBP1, POLG, SCN1B, KCNQ2 x 3, PAFAH1B1, TUBB2B, GABRA1* and *CACNA1H*. A possible diagnosis was identified in 5 additional retrospective patients for whom supporting evidence is pending (*GABRB3, ARHGEF9, CHD2, KCNQ2, SCN3A*). An additional retrospective patient was identified to have a diagnosis which did not fully explain her phenotype (2 variants in *BTD* recognized to cause partial biotinidase deficiency, with biochemical support of low biotinidase). Eight patients (16%: 2 prospective, 6 retrospective) had a diagnosis with possible treatment implications (*ATP1A2, SCN1A, SCN1B, BTD, KCNQ2* x 3, *POLG*). The patient with compound heterozygous mutations in *POLG* was on valproic acid which was stopped due to study results. Her liver enzymes were increasing at time of diagnosis and normalized with valproic acid removal. The clinical utility of using WES in this cohort is supported by timely potential diagnoses in 19/50 patients (38%: 5 prospective; 14 retrospective) within a mean time of 48 days (21–105 days). WES also facilitated multi-locus variant identification that may further impact the phenotype. Within the 557 gene set, sequence kernel association testing will now explore the contribution of additional genetic variants to disease. Genome-wide WES from patients/families with no genetic diagnosis will also be analyzed to implicate novel genes in epilepsy.

1180F

The NINDS Repository collection of patient-derived biospecimens available for neurological disorders research. *C. A. Pérez[1], K. Reeves[1], J. Santana[1], S. Heil[1], A. Green[1], A. Amberson[1], D. Huber[1], R. Zhang[2].* 1) NINDS Repository, Coriell Institute for Medical Research, Camden, NJ; 2) National Institute of Neurological Disorders and Stroke, Bethesda, MD.

Neurological disorders are a major public healthcare concern and their pathophysiology remains largely unknown. The limited availability of reliable disease models hinders progress in identifying reliable genetic and molecular biomarkers for disease onset, diagnosis, progression and therapeutic response. The National Institute of Neurological Disorders and Stroke (NINDS) Repository was established in 2002 with the mission of facilitating research in neurological diseases by providing high quality biospecimens. The NINDS Repository collects biospecimens and de-identified clinical data from patients with neurological disorders as well as neurologically normal controls. The NINDS Repository features collections of patient-derived DNA, fibroblasts and induced pluripotent stem cells (iPSCs) and long term longitudinal collections of biofluids such as plasma, serum, cerebrospinal fluid, saliva and urine. Many samples are annotated with well-defined mutations. Since the NINDS Repository inception, biomaterials from more than 46,000 individuals with cerebrovascular diseases, Parkinsonism, motor neuron diseases, epilepsy, Tourette syndrome, dystonia, Huntington's disease, frontotemporal degeneration and neurologically-normal controls have been banked and are available at through its online catalog at http://catalog. coriell. org/1/NINDS The NINDS Repository has established validated standard operating procedures and rigorous quality control assessments that span the life cycle of all biospecimens collected to provide premium samples. The NINDS Repository aims to ensure and implement standardization for collecting and processing across all samples while protecting patient safety and privacy. In addition, the NINDS Repository utilizes secure and integrated laboratory information management systems to monitor inventory, sample processing, storage, and distribution of biospecimens, to collect , store and share de-identified clinical data associated to those samples and to facilitate sample-data association by cross-referring with other NIH resources such as dbGaP, DMR, etc. The development of such a centralized collection of human biospecimens and their associated de-identified clinical data allows the NINDS Repository to provide a vital resource for research designed to discover and validate genetic and molecular biomarkers relevant for the study and treatment of neurological disorders prevalent in our society.

1181W

The alternative expression of GAD1 in human prefrontal cortex. *R. Tao[1], K. Davis[1,6], C. Li[1], Y. Gao[1], J. Shin[1], M. Gondré-lewis[6], D. Weinberger[1,2,3,4,5], J. Kleinman[1], T. Hyde[1,2,3].* 1) Lieber Institute for Brain Development, Baltimore, MD; 2) Department of Psychiatry, Johns Hopkins School of Medicine, Baltimore, MD; 3) Department of Neurology, Johns Hopkins School of Medicine, Baltimore, MD; 4) Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD; 5) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 6) Department of Anatomy, Howard University College of Medicine, Washington, D. C.

γ-Aminobutyric acid (GABA), the major inhibitory neurotransmitter in mammalian brain, has been implicated in both schizophrenia and normal brain development by postmortem and genetic studies. GABA levels and signaling in part are regulated by glutamic acid decarboxylase (GAD), the rate-limiting enzyme in GABA synthesis. Previous association studies suggested the *GAD1* promoter variants might play a role in schizophrenia and bipolar disorder. Genetic variation in *GAD1* has been associated with GAD67 expression. The decrease of GAD67 mRNA is one of the most replicated findings in the multiple brain regions of patients with schizophrenia. Schizophrenia is a highly heritable disorder. Many schizophrenia risk genes have particular transcripts that are highly expressed during fetal development. Moreover, some of the allelic variants derived from clinical association studies are associated with abnormal expression of some of these fetal-predominant transcripts. With respect to *GAD1*, during rodent embryonic development, two exclusive cassette exons are dominant in the brain, leading to two enzymatically inactive proteins (GAD25 and GAD44) derived from *GAD1*. These two truncated transcripts are highly expressed during fetal brain development, with a rapid decline after birth. In contrast, the full length transcript, GAD67, is expressed at low levels in fetal brain and dramatically rises after birth. To better understand how *GAD1* may be involved in the pathology of schizophrenia, we conducted a series of experiments to characterize the alternative splicing of *GAD1* in human prefrontal cortex. Our microarray and RNA sequencing data both suggest there are multiple splicing events of *GAD1* in human prefrontal cortex. Using PCR, we have identified 8 novel transcripts for *GAD1* in the human brain. There are another two novel exclusive cassette exons for *GAD1* beside I80 and I86. The *GAD1* transcripts produced by those four exclusive cassette exons are all highly expressed in human fetal prefrontal cortex. Our Western protein results showed that these four transcripts produce 3 truncated *GAD1* proteins with a truncation at the C-terminal and 1 truncation of the N-terminal. Further experiments are needed to characterize the sub-cellular locations of these proteins, and potential genetic regulatory components for these 4 fetal dominant transcripts. Our results provide further clues regarding the molecular mechanisms underlying schizophrenia during early brain development.

1182T

Functional impact of *SNX14* mutations that cause cerebellar atrophy and intellectual disability. *D. T. Bryant[1], C. Demetriou[1], E. Peskett[1], M. Ishida[1], D. Jenkins[1], A. Thomas[1], G. Anderson[2], R. Robertson[2], R. Scott[1,2], M. Bitner-Glinzicz[1,2], S. Sousa[3], G. E. Moore[1], P. Stanier[1].* 1) Genetics and Genomic Medicine, University College London, Institute of Child Health, London, United Kingdom; 2) Great Ormond Street Hospital, London, United Kingdom; 3) Medical Genetics Unit, Hospital Pediátrico, Centro Hospitalar e Universitério de Coimbra, Portugal.

A large number of conditions have been described where severe intellectual disability and ataxia are found in patients with cerebellar hypotrophy. Individually, most types are rare and without known molecular pathology. Mutations in *SNX14* have recently been reported to cause a distinct autosomal-recessive cerebellar ataxia with moderate to severe intellectual disability, early-onset cerebellar atrophy, sensorineural hearing loss and coarsened facial features. Patient-derived fibroblasts were acquired from individuals of three unrelated consanguineous families in Turkey and Portugal. These cell lines have mutations that result in either truncation or loss of the SNX14 protein as evidenced by western blot. The SNX14 protein contains a Phox (PX) domain and a regulator of G protein signalling (RGS) domain. In one patient, loss of a single exon (containing the PX-domain) alone was sufficient to result in the associated pathology. This suggests that the PX domain is critical for normal SNX14 function. While the precise function of SNX14 remains unknown, accumulation of vesicular inclusions in the patient cells suggest a disturbance in protein metabolism and/or vesicle mediated transport. Therefore, we are currently investigating these processes in order to identify the precise role of SNX14 and how mutations may lead to the neurological deficit.

1183F

Use of mutant mice to explore a hypothesis of neuronal/synaptic carnitine deficiency as a cause of autism. *J. Ge[1], E. Y. Kim[1], B. L. Zhang[1], J. R. Seavitt[1], F. M. Vaz[2], A. L. Beaudet[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Laboratory of Genetic Metabolic Disorders, Academic Medical Center, Amsterdam, The Netherlands.

The discovery that *TMLHE* deficiency leads to inability to synthetize carnitine in 1 in 350 healthy adult males and is a risk factor for autism has led us to propose a neuronal/synaptic carnitine deficiency hypothesis as a cause of autism especially in males with a normal physical exam and normal structural brain imaging. We have initiated studies in mutant mice to analyze the role of carnitine metabolism in neuronal and synaptic development to explore the hypothesis that carnitine deficiency in the brain disturbs synaptic development. Mitochondria are abundant at the synapse. The intent is to obtain null mutants for all four steps in carnitine biosynthesis and in numerous carnitine-related transporters and enzymes. Null mutations in *Bbox1* and *Aldh9a1*, both putative carnitine biosynthesis enzymes, and in X-linked *Slc6a14*, a putative blood-brain barrier carnitine transporter, were obtained from the Knockout Mouse Project (KOMP). Studies are most advanced for the *Bbox1* mice, and some data are available for the *Bbox1*/*Slc6a14* double null and for the *Aldh9a1* null mice. When fed with commonly used mouse chow which has extremely low carnitine content, *Bbox1*-/- mice are subviable with ~50% of homozygotes dying prior to weaning, but the reduced viability is corrected by supplementation with 0. 2% carnitine in the drinking water. The *Bbox1*-/- mice have 5-10% of normal levels of carnitine in plasma, liver, and brain. While the *Bbox1*-/- mice have no easily observable phenotype other than the subviability, studies including learning and social behavior and seizure susceptibility are in progress. Hemizygous *Slc6a14* -/Y male mice and female -/- mice are viable and fertile and show no obvious phenotype. The double null *Bbox1*/*Slc6a14*mice are available, and the level of carnitine in the brain compared to *Bbox1* null mice should determine whether the already very low levels of carnitine in the brain are further lowered in the double null mice. These studies should determine whether there is physiologically significant evidence that *Slc6a14* is indeed a blood-brain barrier transporter of carnitine in living mice.

1184W

Glutamate receptor interacting proteins modulate AMPA receptor phosphorylation and social behaviors in mice. *M. Han[1], R. Mejías Estévez[1], R. Rose[1], S. Chiu[2], A. Adamczyk[1], R. Huganir[2], T. Wang[1].* 1) Institute of Genetic Medicine, School of Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Neuroscience, School of Medicine, Johns Hopkins University, Baltimore, MD.

Severe deficits in social interactions are a key feature for several neuropsychiatric disorders including autism and schizophrenia. Neural mechanisms that modulate social behaviors are poorly understood. Glutamate receptor interacting protein 1 and 2 (GRIP1/2) neural-enriched scaffolding proteins that bind the c-terminal of AMPA glutamate receptors 2/3 (GluA2/3) and play an important role in the regulation of GluA2/3 trafficking and neural plasticity. Gain-of-function mutations of GRIP1 were shown to correlate with increased deficits in the reciprocal social interactions in patients with autism. To study the role of GRIP1/2 in the regulation of social behaviors, we produced neuron-specific *Grip1/2* double knockout (DKO) mice by crossing *Grip2* conventional KO mice with *Grip1* conditional (neuron-specific deletion via nestin-cre expression) KO mice. Compared to wild type (WT) control mice matched for age, sex, and strain background, DKO mice show a significant increase in sociability, preference for social novelty, and dyadic male-male social interactions. Loss of GRIP1/2 expression results in decelerated GluA2 recycling while autism-associated gain-of-function GRIP1 mutations result in an accelerated GluA2 recycling in neurons. We performed immunoblot of key neural signaling proteins including AMPAR, mGluR, mTOR, and GABA and identified a significant increase in the phosphorylated GluA2 at serine 880 in brain cortex and cerebellum but not striatum. Phosphorylation at S880 is known to affect GluA2 interactions with several neural scaffolding proteins including GRIP1 and PICK1 and alter receptor recycling and AMPA synaptic strength. These data suggest that AMPA-glutamate signaling play a role in modulating social behaviors in mice and disturbance of this signaling pathway contribute to autism-related social deficits.

1185T

Functional Analysis of the Autism and Intellectual Disability Gene *PTCHD1* Reveals Hedgehog Receptor-Like Functions and PDZ-Binding Domain-Specific Regulation of *CNTNAP1* and *NLGN1*. *K. Mittal, B. Degagne, T. Sheikh, J. Vincent.* Molecular Neuropsychiatry and Development Laboratory, Centre for Addiction and Mental Health, Toronto, Ontario, Canada.

This study is focused on investigating the complex functional aspects of a recently identified gene -- PTCHD1, and how its disruption leads to Autism Spectrum Disorder and/or Intellectual Disability. Sonic hedgehog (Shh) signaling plays a pivotal role in the pattern formation of many embryonic tissues and also in homeostasis and regeneration of adult tissues. PTCHD1 shows sequence homology to the Shh receptors PTCH1 and PTCH2, and has previously shown similar Gli repression activity to PTCH1 and 2. To establish the involvement of PTCHD1 in Hedgehog (Hh) pathway, transcription analysis was performed with Hh pathway genes and putative PTCHD1 partners. We also tested for a truncated construct lacking the C-terminal four amino acids, Ile-Thr-Thr-Val (ITTV) of PTCHD1, which is predicted to interact with the PDZ domains of proteins. PTCHD1 over-expression revealed increased levels of NLGN1 and CNTNAP1 mRNA, which suggests that interaction with proteins at the synapse (NLGN1) or at nodes of Ranvier or axo-glial junctions (CNTNAP1) may have a regulatory effect on these genes. The transcript levels were reduced with the PTCHD1 truncated construct, reversing the effect of PTCHD1 over-expression. These results suggest either a regulatory or a downstream effect on NLGN1 and CNTNAP1 genes via a PDZ-domain containing protein. As DLG4 (PSD-95) interacts with K+-voltage-gated channels KCNA1 and KCNA2 (both known interactors with CNTNAP1), and interacts with NLGN1, we hypothesize that PTCHD1 may have a synaptic role mediated by PSD-95. Alternatively, DLG3 (MPP3), SHANK1 or SHANK3 may mediate regulation of NLGN1. We showed high levels of *Shh*, and its putative receptor, *Ptchd1*, a well as *Smoothened* (*Smo*) transcripts in mouse brains between E12 and P2. To assess the expression pattern of PTCHD1 and Smo in post mitotic neurons, we immunolabeled PTCHD1 and Smo in cultured hippocampal neurons- a model system that has been widely used to study signaling pathways in neuronal growth. Positive PTCHD1 immunolabeling was visible in the hippocampal neurons, and Smo immunolabeling was bright and intense in the distal sections of dendrites. We hypothesize that PTCHD1 localization to hippocampal neurons could inhibit the Hh pathway by excluding Smoothened and also allows cilia to function as chemo sensors for the detection of extracellular Shh, similar to PTCH1, during neuronal development and synapse formation. Preliminary results also suggest localization of PTCHD1 in cilia.

1186F

Chromatin remodelers in autism: deciphering regulatory networks that contribute to autism risk. *R. A. Muhle[1], JL. Cotney[2], W. Niu[3], S. Abdallah[3], S. J. Sanders[4], A. J. Willsey[4], W. Liu[4], J. Yin[3], S. K. Reilly[3], A. T. Tebbenkamp[5], C. Tebbenkamp[5], M. Pletikos[5], N. Sestan[5], M. W. State[4], J. P. Noonan[3].* 1) Child Study Center, Yale University, New Haven, CT; 2) Department of Genetics & Genome Sciences, UConn Health, Farmington, CT; 3) Department of Genetics, Yale School of Medicine, New Haven, CT; 4) Department of Psychiatry, University of California, San Francisco, CA; 5) Department of Neurobiology, Yale School of Medicine, New Haven, CT.

Advances in next generation sequencing technologies have greatly enhanced the field of autism genetics. Recent gene discovery efforts in autism spectrum disorder (ASD) have identified chromatin modifiers, such as the chromodomain helicase *CHD8*, as important new contributors to ASD pathogenesis. These and additional ASD risk-associated genes are co-expressed in human midfetal cortex, suggesting that ASD risk genes may converge in regulatory networks that are perturbed in ASD. To investigate the factors shaping this regulatory network, we have undertaken studies to globally map regulatory targets of ASD risk-associated chromatin modifiers during neurodevelopment. We have found that CHD8 regulates other ASD risk genes during human neurodevelopment. We have mapped the binding sites of CHD8 in the developing human and mouse brain using chromatin immunoprecipitation (ChIP-seq), and have characterized global dysregulation of CHD8 targets following *CHD8* knockdown utilizing shRNAs followed by RNAseq. We find that CHD8 gene targets in human and mouse developing brain are significantly enriched in ASD risk genes bearing one or more *de novo* loss of function mutations, and that genes found in ASD risk-associated spatiotemporal co-expression networks during human brain development are more likely to be targeted by CHD8. Knockdown of *CHD8* expression levels by shRNA leads to significant dysregulation of ASD risk genes. CHD8 binds the ASD risk genes *CHD2*, *SUV420H1*, *ARID1B*, *POGZ*, *MLL5/KMT2E*, *ASH1L*, and *SETD2*, among others. In further studies, we will characterize the gene targets of CHD8-targeted ASD-associated chromatin modifiers using *in vivo* and *ex vivo* systems to elucidate common regulatory pathways. These studies provide an initial view of ASD-associated regulatory networks in the developing brain. We will build on these results by identifying regulatory targets of other chromatin modifiers conferring risk for ASD. Mapping regulatory targets of CHD8 and other ASD-associated chromatin modifiers in the developing brain will identify biological pathways underlying ASD etiology. This will provide the means to interpret ASD-associated coding and noncoding variatiants, and offer potential avenues for drug discovery by revealing specific mechanisms contributing to ASD.

1187W

Sonic hedgehog signaling involvement in bipolar disorder is supported by linkage disequilibrium near the EVC and HHIP genes for haplotypes of the D4S2949, D4S397 genotypes according to EVC genotypes among the Old Order Amish. *Y. E. Song[1], M. Galdzicka[2], J. A. Egeland[3], E. I. Ginns[4], R. C. Elston[1].* 1) Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, OH; 2) Department of Pathology, University of Massachusetts Medical School, Shrewsbury, MA; 3) Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, FL; 4) Departments of Psychiatry and Pathology, University of Massachusetts Medical School, Shrewsbury, MA.

Our study of cosegregation of Ellis-van Creveld syndrome (EvC) and bipolar affective disorder (BPAD) in a large, multigenerational Old Order Amish pedigree led to the identification of significant association between absence of bipolar I (BPI) and presence of EvC (Ginns et al, Molecular Psychiatry 2014). This discovery suggested that the molecular change causing EvC confers protection (that is, mental health wellness) from bipolar I (BPI) cases, as well as suggesting a more general protection against the spectrum of affective disorders. The prior evidence of genetic linkage between BPAD and two loci at the opposite ends of chromosome 4, D4S2949 and D4S397, coupled with the fact that EvC is the result of disrupted sonic hedgehog (Shh) signaling, suggested the functional involvement of Shh signaling as an underlying molecular mechanism for susceptibility and protection in BPAD. The estimated percent of the total sample variance that can be attributed to EVC segregation for various clinical category contrasts provided statistical evidence that disruption of sonic hedgehog signaling in EvC confers protection from BPI and perhaps against other affective disorders. In this present study, using the same Amish sample, additional genotype data were analyzed to provide the missing link from the original article. Specifically, we examined the association between alleles at D4S2949 and D4S397 respectively near the EVC and HHIP genes, at the opposite ends of chromosome 4. By examining this association, additional corroborating evidence was obtained for sonic hedgehog involvement in BPAD, both in all subjects and in those with the EVC G/G genotype. The best contrasting clinical categories explained by this protective effect of EvC, after BPI, are those classified with BPI, BPII or other disorders in the BPAD spectrum, versus those who are definitely mentally well. The new data thus support a more general protection against the spectrum of affective disorders in these families. Understanding how disrupted Shh signaling protects against BPI could uncover variants in the Shh pathway that cause or increase risk for bipolar and related mood disorders. We conclude that studying special populations, like the Amish, may be a better strategy to detect complex disease causing genes and pathways than the pursuit of ever larger case-control samples.

1188T

Genetic ablation of the long non-coding RNA *Pantr2* affects cortical neuron migration and striatal gene expression in mice. *M. Feyder[1,2], A. Groff[3,4], M. Sauvageau[3,4], S. Lodato[3,4], D. B. Sanchez-Gomez[3,4], M. Morse[3,4], C. Gerhardinger[3,4], J. L. Rinn[3,4,5], P. Arlotta[3,4], L. A. Goff[1,2].* 1) Johns Hopkins University, Baltimore, MD; 2) Department of Neuroscience, Johns Hopkins School of Medicine, Baltimore, MD; 3) Stem Cells and Regenerative Medicine, Harvard University, Cambridge, MA; 4) The Broad Institute of MIT and Harvard, Cambridge, MA; 5) Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA.

Long non-coding RNAs (lncRNAs) mediate a diverse number of cellular processes and can regulate cellular differentiation. These conclusions have been principally derived from cell culture systems. The physiological relevance of lncRNAs *in vivo*, however, is less known and represent a significant limitation to our current understanding because many are tissue-specific. The lncRNA *Pantr2* is highly expressed in neural stem cells and exhibits a moderate expression in the adult brain, suggesting a potential role in the development of the central nervous system. Accordingly, brain tissue from *Pantr2* knockout (KO) mice exhibits attenuated indices of proliferation in progenitor cell populations, altered gene expression associated with cellular proliferation and neuronal maturation, and abnormal cortical lamination (Sauvageau *et al.* , 2013). We extend these initial findings and report that these same KO animals also exhibit disordered cortical neuron migration and alterations in striatal gene expression. In the embryonic brain of KO mice, neurons expressing Tuj1, a neuronal-specific tubulin, are found in terminal zones, where they are normally absent. Similarly, in the adult brain, neurons immunostained for Cux1, a marker of upper cortical laminae, were present in lower cortical laminae, where they are also normally absent. Taken together, these results suggest that a loss of *Pantr2* may impair the migration of cortical neurons. Additionally, we uncovered transcriptional abnormalities in the striatum, a large subcortical nuclei of the basal ganglia that receives widespread projections from the cortex. RNA-seq of the embryonic telencephalon demonstrated that *Ppp1r1b* (*DARPP-32), Ngef* and *Pcp4L1*, all of which are highly expressed in the striatum, were increased in KO mice, consistent with the notion of precocious differentiation in *Pantr2* KO mice. An increase in striatal DARPP-32 was confirmed by immunostaining. Differentially expressed transcripts between WT and KO mice were associated with disorders of the basal ganglia, such as Huntington's disease, Parkinson's disease and dyskinesia. These data demonstrate that disruption of *Pantr2* expression perturbs both the cortex and the striatum. As corticostriatal connectivity may underlie behavioral disorders such as obsessive-compulsive disorder and autism, these data reveal an *in vivo* role for lncRNAs with potential relevance to clinical syndromes.

1189F

Role of ANK2 in autism. *J. Li, M. Wakahiro, B. Fregeau, G. Gente, E. Sherr.* Dept Neurology, Univ California, San Francisco, San Francisco, CA.

ANK2 encodes a member of the ankyrin family of proteins (Ankyrin B) that link integral membrane proteins (e. g. L1CAM) to the underlying spectrin-actin cytoskeleton. Ankyrins play key roles in activities such as cell motility, activation, proliferation, cell-cell contact and the maintenance of specialized plasma membrane domains. Mutations in ANK2 have been found to cause autosomal dominant long QT syndrome 4 and cardiac arrhythmia syndrome in humans. Recent next generation DNA sequencing projects in Autism cohorts (the Simons Simplex Collection) have revealed several heterozygous de novo ANK2 missense and nonsense mutations marking ANK2 as a 'high-confidence' autism candidate gene. Additionally, our lab has identified multiple individuals who not only have cognitive and behavioral deficits, but are also missing the corpus callosum (agenesis of the corpus callosum; ACC), which is comprised of nearly 200 million axons connecting the two cerebral hemispheres. These genetic findings highlight the importance of ANK2 in neurodevelopment, but how mutation of ANK2 leads to these brain disorders is still unknown. As a first step to address this question, we reported an expected and significant decrease in expression of ANK2 mRNA and protein in patient compared to control PBMC's. We also investigated the abundance of the neuronal adhesion protein, L1CAM. This protein has been shown to bind to ankyrin B, and the protein half-life was dependent on ankyrin B "anchoring" L1cam to the plasma membrane. Thus, we found that L1CAM protein abundance is decreased in patients with ANK2 mutations (one with a balanced chromosomal translocation through ANK2 t (4;8)(q25; q23), and in two other Autism patients with nonsense mutations in ANK2) and that L1CAM expression is rescued after transient transfection of ANK2 into a cell line from one patient carrying a nonsense mutation in ANK2. Our results suggest that disruption of the normal interaction and balance between ANK2 and L1CAM may play an important role in Autism and other neurodevelopmental disorders such as ACC. Moreover, this linkage between ACC and ASD has implications for the mechanistic overlap for these two groups of disorders.

1190W

Fine-mapping of regulatory variants in schizophrenia. *K. Farh[1], M. J. Daly[2], B. Bernstein[2].* 1) Illumina, San Diego, CA; 2) Broad Institute, Cambridge, MA.

Genome-wide association studies have now identified over a hundred risk loci for schizophrenia, over 97% of which are noncoding, indicating that the great majority of genetic risk can be attributed to differences in gene regulation between individuals. To understand the underlying molecular mechanisms, we are employing a combination of statistical and biochemical methods to identify genetic variation that disrupts *cis*-regulatory elements in disease-relevant cell types. By measuring allele-specific differences in gene regulation on the two chromosomes of heterozygous individuals, we can effectively control for effects on gene regulation due to non-genetic factors, such as age, sex, environment, and stochastic influences.

1191T

Pathway analysis of genome-wide SNP data for Gilles de la Tourette Syndrome shows enrichment in genes expressed in nervous system tissues. *F. Tsetsos[1], J. Alexander[1], D. Yu[2], J. H. Sul[3], G. Coppola[4], I. Zelaya[4], P. Drineas[5], C. Mathews[6], J. M. Scharf[2], P. Paschou[1], TSA-ICG, GGRI.* 1) Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Evros, Greece; 2) Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetics Research, Departments of Neurology and Psychiatry, Massachusetts General Hospital, Boston, MA, USA; 3) Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 4) Semel Institute for Neuroscience and Human Behavior, Departments of Neurology and Psychiatry, The David Geffen School of Medicine at University of California Los Angeles, Los Angeles, CA, USA; 5) Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY, USA; 6) Department of Psychiatry, University of California, San Francisco, CA, USA.

Gilles de la Tourette Syndrome (GTS) is a childhood-onset neuropsychiatric disorder that is characterized by a multitude of vocal and motor tics. Recent advances in the genetics of GTS are starting to shed new light on its genetic aetiology. We have recently completed a second Genome Wide Association Study (GWAS) of GTS on 2,859 cases/3,855 controls. A subsequent meta-analysis with existing GTS GWAS data (final sample size of 4645 cases and 9,750 controls) generated multiple LD-independent SNPs with p-value less than 10^{-5}. These SNPs implicate novel loci on Tourette Syndrome pathogenesis.

Here, we focused on the top scoring SNPs from the GTS GWAS meta-analysis for downstream pathway analysis. For the pathway analysis, we used DEPICT, a novel tool for the interpretation of GWAS results, and additional pathway analysis software, including INRICH, FORGE and DAPPLE. Furthermore, we also applied a set-based association approach on the Second Genome Wide Association Study case and control data. In order to perform the set-based association, we collected pathway gene sets from KEGG, REACTOME, Molecular Signatures Database, OMIM and constructed sets of SNPs located at genes present on the pathway gene sets. The set-based association was run using PLINK and 1 million permutations.

DEPICT analysis showed significant enrichment among top GTS associated SNPs in genes expressed in nervous system tissues, including the parietal lobe and the basal ganglia. INRICH analysis hints towards the involvement of neuronal transcription factor binding sites, with the top scoring being TCF3. We also report the results from the set based association analysis, which further illuminate candidate pathways for investigation, implicating a complex network of factors. Our study aims to untangle the complex interaction networks of genes implicated in GTS and points to candidate etiological pathways for further investigation in future studies.

1192F

Structural variants and neurodegenerative diseases in aging: regulatory and causality consequences. *O. Chiba-Falek[1], M. W. Lutz[1], R. Saul[2], A. D. Roses[1, 3].* 1) Department of Neurology, Duke University, Durham, NC; 2) Polymorphic DNA Technologies, Alameda, CA; 3) Zinfandel Pharmaceuticals, Chapel Hill, NC.

Over the last decade large genome-wide association studies (GWAS) identified many loci associated with neurodegenerative disease risk, however the precise causal genetic variants and the molecular mechanisms underlying those genetic associations remain largely unknown. In GWAS, SNPs have taken the spotlight in analyses of human complex diseases and expression traits (eQTL). Structural Variants (SVs) such as, short tandem repeats (STRs), homopolymers and indels, were underrepresented, mainly due to GWAS platforms that enabled high density SNP genotyping and the inability to accurately sequence long contiguous genomic sequence using current next-generation sequencing. We embarked on an effort to identify noncoding functional SVs within loci in the genome that have been associated with neurodegenerative phenotypes. We have studied the functional consequences of intronic and intragenic SVs and their role in susceptibility to a broad spectrum of dementia and Lewy body (LB) disorders. We employ bioinformatics tools including a novel software (dbSV) to annotate SV, and phased sequencing approach to discover candidate SVs and haplotypes associated with neurodegenerative diseases and pathologies and to evaluate the *cis*-effects of disease-risk SVs on gene expression and splicing in human brain tissues from disease-affected and healthy individuals. Using phylogenetic analyses based on phased sequence data, we identified a variable intronic poly-T, in the *TOMM40* gene that is associated with risk for Alzheimer's (AD) and age of onset. We demonstrated using AD-affected and normal brain tissues and a luciferase reporter system that this highly-variable poly-T site regulates the transcripts levels of both *TOMM40* and its neighboring gene, *APOE*. Previously we also discovered that Rep1, a SV associated with increased PD-risk, regulates *SNCA* transcription in human brain and confirmed our findings using luciferase reporter assay and humanized mouse model. In our new studies we have recently identified four distinct haplotypes within a highly-polymorphic-low-complexity CT-rich region with specific haplotype conferred risk to develop LBV/AD. We further demonstrated that the risk haplotype was significantly associated with elevated levels of *SNCA*-mRNA in human brain tissues and suggested that the CT-rich site acts as an enhancer element. Collectively, our research has demonstrated the importance of noncoding SVs in the etiology of sporadic neurodegenerative diseases in adulthood.

1193W

Layered genetic control of DNA methylation and gene expression: locus of multiple sclerosis in healthy individuals. *J. Shin[1,12], C. Bourdon[1,12], M. Bernard[1], M. Wilson[1], E. Reischl[2], M. Waldenberger[2], B. Ruggeri[3], G. Schumann[3], S. Desrivieres[3], A. Leemans[4], M. Abrahamowicz[5], L. Richer[6], L. Bouchard[7,8], D. Gaudet[8,9], T. Paus[10,11], Z. Pausova[1], The IMAGEN Consortium, The SYS Consortium.* 1) The Hospital for Sick Children, University of Toronto, Toronto, ON, Canada; 2) Research Unit of Molecular Epidemiology, Helmholtz Zentrum Munchen, Munich, Germany; 3) MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom; 4) Image Sciences Institute, University Medical Center Utrecht, Utrecht, the Netherlands; 5) Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Canada; 6) Department of Psychology, Université du Québec à Chicoutimi, Chicoutimi, Canada; 7) Department of Biochemistry, Université de Sherbrooke, Sherbrooke, Canada; 8) ECOGENE-21 and Lipid Clinic, Chicoutimi Hospital, Chicoutimi, Canada; 9) Department of Medicine, Université de Montréal, Montréal, Canada; 10) Rotman Research Institute, University of Toronto, Toronto, Canada; 11) Child Mind Institute, New York, NY, USA; 12) Authors with equal contribution.

DNA methylation may contribute to the etiology of complex genetic disorders through its impact on genome integrity and gene expression; it is modulated by DNA-sequence variants, named *methylation quantitative-trait loci* (meQTLs). Most meQTLs influence methylation of a few CpG dinucleotides within short genomic regions (<3kb). Here we identified a layered genetic control of DNA methylation at numerous CpGs across a long 300-kb genomic region. This control involved a single *long-range* meQTL and multiple *local* meQTLs. The *long-range* meQTL explained >50% of variance in methylation of CpGs located over extended areas of the 300-kb region. The meQTL was identified in four samples (p=2. 8x10-17, 3. 1x10-31, 4. 0x10-71, 5. 2x10-199), comprising a total of 2,796 individuals. The *long-range* meQTL was strongly associated not only with DNA methylation but also with mRNA expression of several genes within the 300-kb region (p=7. 1x10-18–1. 0x10-123). The associations of the meQTL with gene expression became attenuated when adjusted for DNA methylation (causal inference test: p=2. 4×10-13–7. 1×10-20), indicating coordinated regulation of DNA methylation and gene expression. Further, the *long-range* meQTL was found to be in linkage disequilibrium with the most replicated locus of multiple sclerosis, a disease affecting primarily the brain white matter. In middle-aged adults free of the disease, we observed that the risk allele was associated with subtle structural properties of the brain white matter found in multiple sclerosis (p=0. 02). In summary, we identified a *long-range* meQTL that controls methylation and expression of several genes and may be involved in increasing brain vulnerability to multiple sclerosis.

1194T

De novo BCL11A variants in neurodevelopmental disorder disrupt multiple aspects of protein function. *S. B. Estruch[1], S. A. Graham[1], P. Deriziotis[1], S. E. Fisher[1,2].* 1) Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; 2) Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands.

The rare chromosome 2p16. 1-p15 deletion syndrome (MIM 612513) is characterized by intellectual disability, dysmorphic features and microcephaly. A high proportion of affected individuals have autism, behavioral problems and language deficits. Skeletal and organ abnormalities have also been reported. Comparison of the chromosomal regions deleted in different cases suggests that haploinsufficiency of the zinc finger transcription factor BCL11A (MIM 606557) may underlie the neurological features of the syndrome. Although the role of BCL11A in brain development is poorly understood, the gene is expressed in the developing cortex, hippocampus, basal ganglia and cerebellum, and has recently been reported to play a role in the specification of subcortical projection neurons by repressing expression of TBR1 (MIM 604616), a gene recurrently mutated in cases of autism. In addition to its function in the brain, BCL11A has a key role in mediating the switch from the fetal to the adult form of hemoglobin, and persistence of fetal hemoglobin has been reported in several patients with BCL11A deletions. The first disorder-associated missense variants in BCL11A have recently been identified in three unrelated infants with developmental delay. The variants in these patients cluster within the N-terminal portion of the protein, and lie outside the zinc finger DNA-binding domains, in a region of unknown function. We therefore sought to characterize the effects of these variants on protein function in order to confirm their etiological role in disorder in these three patients, and to illuminate the molecular mechanism of disorder. We found that all three variants disrupt the localization of BCL11A within the nucleus and abolish protein dimerization and transcriptional regulatory activity. Our results therefore strongly support a causal role for BCL11A variants in the developmental delay in these patients, and add to the growing evidence that BCL11A is a recurrently mutated gene in neurodevelopmental disorder. Furthermore the characterization of these variants reveals a key role for the N-terminal region of BCL11A in mediating protein dimerization and regulation of transcription.

1195F

Deciphering the genetic basis of intellectual disability using 1000 knockouts. *B. Yalcin[1,3,6], A. Mikhaleva[1], V. E. Vancollie[2], M. Kannan[3], C. Wagner[3], A. Edwards[4], H. Whitley[3], J. Estabel[2], C. J. Lelliott[2], U. K. Sanger Mouse Genetics Project[2], D. A. Keays[5], J. K. White[2], R. Ramirez-Solis[2], D. J. Adams[2], Y. Herault[3], A. Reymond[1].* 1) Center for Integrative Genomics, University of Lausanne, Switzerland; 2) Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1HH, UK; 3) Institute of Genetics and Molecular and Cellular Biology, Illkirch, 67404, France; 4) Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, UK; 5) Research Institute of Molecular Pathology, 1030 Vienna, Austria; 6) Corresponding author.

Intellectual disability (ID) affects 1-3% of the general population. Genetic mutations account for about half of the currently undiagnosed cases, and it has been suggested that up to 2,000 further genes remain to be identified. To identify genes involved in brain malformation and potentially associated with ID, we are collaborating with the Sanger Mouse Genetics Project (MGP), allied to the International Mouse Phenotyping Consortium (IMPC), to systematically study the neuroanatomy of the MGP/IMPC knockout mouse strains. We are measuring a standardized set of 78 brain parameters across 22 brain regions to detect a variety of mechanisms that underlie brain disorders, such as aberrant cell proliferation, neuronal migration defects or elevated cell death rates. So far, we have assessed brain defects in 992 knockout mouse genes that are randomly distributed throughout the mouse genome. These preliminary data yielded success with the identification of 91 genes including 18 well-known ID genes such as *Aspm*, *Cdk5rap2*, *Ctcf*, *Dlg3* and *Smc3*, demonstrating the pertinence of our approach. The remaining 73 genes are novel, and examples include *Bach2*, *Cbx6*, *Daam2, Kif21b, Os9, Pik-3cb*, *Rnf10*, *Selk* and *Wdr47*. Interestingly, one fifth of these novel genes have been reported in non-recurrent forms of severe ID (Gilissen *et al* Nature 2014), thus are good candidate genes for ID. Strikingly, these novel genes showed a significant enrichment in both cellular components and molecular functions pertaining to the cytoskeleton, and were mostly affecting the formation of the commissures and the ventricles. Half of these novel genes have a global effect throughout the brain, for example *Rnf10* and *Kif21b*, while the other half have specific impact on brain structure, for example *Bach2*. Our study is the largest screen of brain morphology from the MGP/IMPC. It shows the importance of an unbiased screen as we can detect severe brain abnormalities in about 8% of knockout genes of previously unknown brain function, and that these novel genes are a rich resource for investigating brain disorders. Our resource offers a complementary approach to human genetic studies.

1196W

Disassembly of MICOS complex by *CHCHD10* mutations promotes lossof mitochondrial cristae with defectsin mitochondrial genome maintenance and apoptosis. *E. Genin[1], M. Plutino[1], S. Bannwarth[1,2], E. Villa[3], E. Cisneros-Barroso[4], M. Roy[5], B. Ortega-Vila[4], K. Fragaki[1,2], F. Lespinasse[1], E. Pinero-Martos[4], G. Augé[1,2], D. Moore[6,7], F. Burté[6,7], S. Lacas-Gervais[8], Y. Kageyama[5], P. Yu-Wai-Man[6,7], H. Sesaki[5], J. E. Ricci[3], C. Vives-Bauza[4], V. Paquis-Flucklinger[1,2].* 1) IRCAN, UMR 7284, INSERM U1081, Nice Sophia-Antipolis University, Nice, France; 2) Department of Medical Genetics, National Centre for Mitochondrial Diseases, Nice, France; 3) INSERM U1065, Centre Méditérranéen de Médecine Moléculaire (C3M), Nice Sophia-Antipolis University, Nice, France; 4) Research Health Institute of Palma (IdISPa), University Hospital, Palma de Mallorca, Spain; 5) Department of Cell Biology, Johns Hopkins University, School of Medicine, Baltimore, MD, USA; 6) Wellcome Trust Centre for Mitochondrial Research, Institute of Genetic Medicine, Newcastle University, Newcastle, UK; 7) Newcastle Eye Centre, Newcastle, UK; 8) Joint Center for Applied Electron Microscopy, Nice Sophia-Antipolis University.

Recently, we described *CHCHD10* as a novel gene responsible for mtDNA instability disorder by studying a large family with a late-onset phenotype including motor neuron disease, cognitive decline looking like frontotemporal dementia (FTD), cerebellar ataxia and mitochondrial myopathy with accumulation of multiple mtDNA deletions. Subsequently, other groups and ourselves identified *CHCHD10* mutations in cohorts of patients with frontotemporal dementia-amyotrophic lateral sclerosis (FTD-ALS), pure familial or sporadic ALS, late-onset spinal motor neuropathy (SMAJ) or Charcot Marie Tooth 2A (CMT2A).

Here, we show that CHCHD10 is a component of the "mitochondrial contact site and cristae organizing system" (MICOS) complex. The expression of *CHCHD10* mutant allele leads to MICOS complex disassembly and loss of mitochondrial cristae in patient fibroblasts. The abnormalities of the inner membrane are responsible for nucleoid disorganization leading to defect in mtDNA repair after oxidative stress, which explains the multiple mtDNA deletions found in patient muscles. Interestingly, the expression of *CHCHD10* mutant alleles inhibits apoptosis by preventing cytochrome*c* release. This result supports previous studies suggesting that, in some ALS models, motor neuron death can occur *via* caspase-independent apoptotic mechanism.

In conclusion, we show for the first time that mutations in a gene encoding a MICOS component are responsible for human disorder. Dissecting the cellular pathways disrupted by the expression of *CHCHD10* mutant alleles represents, therefore, a golden opportunity to gain powerful insight into the sequence of events that link mitochondrial dysfunction with neurodegenerative disorders.

1197T

***Polr3a* KI and KI/KO mice do not develop hypomyelinating leukodystrophy.** *K. Choquet[1,2,3], S. Durrieu[4], R. Lariviere[1], M. -J. Dicaire[1], G. Bernard[5], T. E. Kennedy[1], M. Teichmann[4], C. L. Kleinman[2,3], B. Brais[1,2].* 1) Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada; 2) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 3) Lady Davis Institute, Jewish General Hospital, Montreal, Quebec, Canada; 4) INSERM U869, Institut Europeen de Chimie et Biologie, Universite Bordeaux Segalen, Bordeaux, France; 5) Department of Pediatrics, Neurology and Neurosurgery, Division of Pediatric Neurology, Montreal Children's Hospital, McGill University Health Center, Montreal, Quebec, Canada.

Hypomyelinating leukodystrophies (HLD) are a heterogeneous group of neurodegenerative disorders characterized by impaired myelin formation in the central nervous system (CNS). Recessive mutations in the ubiquitously expressed *POLR3A* and *POLR3B* genes cause one of the most frequent forms of childhood-onset HLD: POLR3-related leukodystrophy. *POLR3A* and *POLR3B* encode the two catalytic subunits of RNA Polymerase III (POLR3), which is responsible for the transcription of transfer RNAs and a large array of other small non-coding RNAs. We hypothesized that mutations in *POLR3A* and *POLR3B* would alter POLR3's transcription efficiency, leading to a decreased expression of key transcripts for the development of CNS white matter. Comparison of the expression levels of POLR3 transcripts in control and patient-derived fibroblasts identified only one transcript decreased in a subset of patients, BC200 RNA. This lack of global deregulation of POLR3 transcripts in patients' fibroblasts suggested that the impact of *POLR3A* and *POLR3B* mutations would be restricted to CNS tissues. Therefore, we developed and characterized mouse models of POLR3-related HLD. Since homozygous *Polr3a* knock-out (KO) mice could not be conceived, we generated knock-in (KI) mice homozygous for the French Canadian founder mutation c. 2015G>A (p. Gly672Glu) in *Polr3a*, as well as compound heterozygous mice for one *Polr3a* KI and one KO allele (KI/KO). Both KI and KI/KO mice are viable, able to reproduce and do not display a grossly abnormal phenotype. To establish if they manifest a subtler motor phenotype, WT, KI and KI/KO mice (n=15 per group) were submitted to rotarod, inverted grid, beam and open field tests at 40 and 90 days old. The three groups performed comparably on all tests, indicating that the KI and KI/KO mice have normal balance, coordination, strength and general locomotion. Quantification of CNS myelin proteins in brain and cerebellum of 30 and 90 days old mice did not detect differences between the three groups, suggesting that myelination is not impaired in *Polr3a* KI and KI/KO mice. Finally, we compared the expression levels of POLR3 transcripts in the brains of KI and WT mice using RNA-Seq but no differences were uncovered. In conclusion, our results suggest that *Polr3a* KI and KI/KO mice do not recapitulate the severe childhood-onset HLD seen in humans with *POLR3A* mutations, although we cannot exclude that they may develop a later-onset phenotype.

1198F

Evaluating the expression of the long non coding RNA *lnc- IL7R* and two isoforms of *IL7RA* gene in the PBMCs of relapsing-remmiting multiple sclerosis patients. *P. Bina, M. Pahlevan Kakhki, M. Behmanesh.* Biological Sciences, Tarbiat Modares University, Tehran, Tehran, Iran.

Multiple sclerosis (MS [MIM 126200]) is an inflammatory-mediated demyelinating disease of the human central nervous system (CNS). Many studies showed that the environmental and genetics factors are involved in the pathogenesis of MS. Interleukin 7 receptor alpha (*IL7RA* [MIM 146661]) (also known as *CD127*), encoded by Interleukin 7 receptor (*IL7R* [MIM 146660]), is a member of the type I cytokine receptor family and forms a receptor complex with the common cytokine receptor gamma chain. It has been showed that rs6897932 SNP [MIM 612595] on *IL7R* affects alternative splicing of exon 6, leading to increased skipping of the entire exon 6 and corresponds to the soluble form of the ligand specific IL7RA subunit. *LncRNAs* are a group of non coding RNAs which are longer than 200 nt and play important roles in various biological processes, such as proliferation, differentiation, and development through various modes of action. *Lnc-IL7R*, is 1427 nt in length which overlaps with the 3'untranslated region (3'UTR) of the human *IL7R* gene, regulates the inflammatory responses of LPS-treated cells. The majority of lnc-IL7R existed in the nucleus. The increase in *lnc-IL7R* expression after lipopolysaccharide stimulation indicates that *lnc-IL7R* is involved in the early immune responses. RR-MS patients were selected based on some exclusion and inclusion criteria. The study was approved by the ethnics commette of Tarbiat Modares university of IRAN and informed consent were obtained from all participants. five ml of whole blood were obtained and PBMCs harvested using Ficoll solusion, RNA extracted by Trizol reagent, and real time PCR was performed by Step One ABI system using specific primers. χχct method was used for the analysis of gene expression results. Our pilot studies on LncRNAs showed interesting evidences from lncRNAs key roles in the different physiological conditions of MS. Also, we checked the expression of *Lnc-IL7R* using RT-PCR method in the diabetic chronic arthery disease (CAD+) patients which showed the significant elevated expression of this LncRNA in comparison with healthy subjects. Analysis of the expression level of this Lnc-RNA in the diabetic and MS patients are under investigation in our laboratory by RT-PCR and Real time PCR methods. We hope that, our study shed further light to the precise mechanism of gene expression regulation in MS patients. In summery, this study supports the view that *lnc-IL7R* may be involved in the pathogenesis of RR-MS.

1199W

Divergent transcription of the *BDNF* gene in human and mouse brain: relevance for schizophrenia. *G. Ursini[1,2], G. Punzi[1,2], J. H. Shin[1], K. Maynard[1], E. Radulescu[1], A. Jaffe[1], J. E. Kleinman[1], T. M. Hyde[1], Y. Jia[1], K. Martinowich[1], D. R. Weinberger[1].* 1) Lieber Institute for Brain Development, Baltimore, MD; 2) Group of Psychiatric Neuroscience, Aldo Moro University, Bari, Italy.

Background: Non-coding RNA (ncRNA) transcription by RNA polymerase II often initiates from bidirectional promoters that synthesize mRNA and ncRNA in opposite directions. These divergent ncRNAs may be particularly important in CNS function as cells of neural origin are highly transcriptionally active and show robust ncRNA expression. *BDNF* is a complex and highly regulated gene: several untranslated 5' exons can be spliced with a 3' coding exon providing bipartite or tripartite transcripts and multiple splice variants. Here we detected and analyzed previously unknown divergent transcription in this locus. Methods: PolyA enriched RNA was extracted from post-mortem human and mouse prefrontal cortex (PFC) and then purified and enriched with PCR to create a cDNA library for high throughput sequencing using the Illumina HiSeq2000. We used Race-PCR, followed by PacBio Sequencing of the PCR amplicons, in order to detect the structure of the divergent transcripts. We performed a KCl-induced depolarization experiment on cultured E14 mouse cortical neurons. A functional magnetic resonance imaging analysis of brain physiology *in vivo* was performed in 326 healthy volunteers, 56 patients with schizophrenia and 84 siblings. Results: In the human and mouse transcriptome, novel antisense divergent ncRNAs (divRs) are initiated from bidirectional promoters located in the region of *BDNF* exons I, II and III and have different 3' terminations, also located 50kb 5' of *BDNF*. Expression of these divRs is low at baseline, but increases 20-fold following induced neural activity ($p<0.01$). We found that the expression of divRs is significantly greater in patients with schizophrenia compared with controls ($F1, 200=3.8$, $p=.04$) and genetic variants associated with expression of these divRNAs ($p<0.01$) show association with schizophrenia ($p<1e-4$) and predict PFC activity measure with fMRI during working memory in patients with schizophrenia and their healthy siblings. Conclusions: We demonstrate multiple antisense divRs in the *BDNF* locus, revealing an additional complexity in the biology of *BDNF* which may be differentially regulated in schizophrenia as well as other disorders. Divergent transcription might contribute to the complex regulation of the *BDNF* gene, in human and in mouse, specifically after neuron depolarization. More in general, when mapping a phenotype to a certain genomic locus, it is necessary to consider ncRNAs and bidirectional promoters as causes for phenotypic variability.

1200T

Gene expression profiles of cultured neural progenitor cells from patients with schizophrenia provide insight into the disease. *C. Armoskus, V. Spitsyna, T. Souaiaia, C. Walker, J. Nguyen, A. Camarena, J. Herstein, K. Wang, J. A. Knowles, O. V. Evgrafov.* Zilkha Neurogenetic Institute, USC, Los Angeles, CA.

Gene expression profiling by RNA-Seq is increasingly being used to investigate a variety of mental disorders, including schizophrenia (SZ). However, most studies examine post-mortem brain, which may not be an adequate tissue for investigation of brain development, or the blood transcriptome, which may not be fully representative of neural gene expression. In this study, we investigated gene expression in cultured neural progenitor cells derived from olfactory neuroepithelium (CNON) from 82 patients with SZ and 72 controls. Cells of this type divide, migrate, and differentiate into neurons, demonstrating the same developmental processes that define the size and structure of components of the nervous system. Libraries were made using TruSeq Stranded Total RNA library preparation kits (Illumina) and sequenced on HiSeq2000 (Illumina). Reads were aligned and counted with GT-FAR against hg19/GenCode 19. Differential expression (DE) analysis using DESeq2 revealed 105 genes DE in SZ (FDR < 0. 05). Several DE genes, such as FGFR2, PPARGC1A, VAV3, NRXN3, ADRA2A, ABCA1, and PLAT have previously been implicated in SZ based on GWAS or candidate gene association studies. Other DE genes, including CCL8, CCL11, CNTNAP3, PDE8B, and TGFB2, have previously been reported as upregulated in SZ, matching our observations. HTR2B, a serotonin receptor, is over-expressed in SZ and has not been previously associated with SZ but is a functionally plausible candidate. IPA canonical pathway analysis showed enrichment for TR/RXR and PPARα/RXRα activation pathways. Notably, both pathways were found significantly enriched in a network created based on GWAS data. Enrichment for biological processes from Gene Ontology included “cell migration”, “cell adhesion”, “tissue development”, and “positive regulation of cell proliferation”, which are processes involved in neurodevelopment, and “positive regulation of MAPK cascade”. The KEGG pathway “Cytokine-cytokine receptor interaction” was also enriched, supporting the hypothesis of cytokine involvement in SZ etiology. Results of transcriptome analysis of CNON samples are strongly supported by other studies using different methods and different cellular or animal models. However, the CNON model has the advantages of allowing genetic modification or pharmaceutical treatment of individual cell lines. CNON cell lines also have the potential to be differentiated into neurons to study disease-specific changes in synaptic machinery.

1201F

Regulatory Function of Schizophrenia-Associated Variants in *CACNA1C*. *N. Eckart[1], R. Wang[1], R. Yang[2], Q. Song[3], H. Zhu[3], D. Valle[1,4], D. Avramopoulos[1, 2].* 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Psychiatry, Johns Hopkins University School of Medicine, Baltimore, MD; 3) Department of Pharmacology and Molecular Sciences, Johns Hopkins University School of Medicine, Baltimore, MD; 4) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD.

Schizophrenia is a chronic psychiatric disorder with 60-80% heritability. Several GWAS have repeatedly identified the SNP rs1006737, an intronic variant in the gene *CACNA1C*, to be strongly associated with disease risk. We and others have previously shown a correlation between genotype at this SNP and steady state levels of *CACNA1C* mRNA in human post mortem brains, suggesting it tags a regulatory variant at this locus. Here we report on our search among variants in high linkage disequilibrium (r2 > 0. 8) for those that may be functionally relevant. We consistently observe that the risk allele of rs4765905, a SNP tagged by rs1006737, shows significantly reduced enhancer activity (p=1. 5 *10-6) in dual luciferase reporter assays on human neuroblastoma SK-N-SH cells. The two alleles of this SNP also show different affinity for proteins or complexes extracted from SK-N-SH nuclei in electrophoretic mobility shift assays (EMSA). Using protein microarrays we show allele-specific binding for rs4765905 to a number of proteins, including transcription factors such as ZKSCAN5 and HR. This evidence suggest that rs4765905 might have regulatory function. While only rs4765905 shows differences in luciferase assays, 13 of 16 SNPs examined by EMSA show allele-specific protein binding. Based on the observed electrophoretic shift, it appears that it may be the same protein complexes that bind most of these variant sequences. Our data from the protein microarrays also show that 15 of the 4215 proteins bind 4 or more of the 9 SNPs examined thus far. These observations suggest the possibility that complex interactions involving multiple SNPs in strong LD might regulate *CACNA1C* expression. *CCAT* is an alternative *CACNA1C* transcript starting from the exon 46 and encodes a transcription factor hypothesized to negatively regulate *CACNA1C* transcription. We applied circularized chromatin conformation capture with next-generation sequencing (4C-seq) from the *CCAT* promoter and found interacting fragments approximately 20kb downstream of the transcription termination site in both HEK293 and SK-N-SH cells, but interaction evidence in the area of the schizophrenia associated SNPs was weak. We are currently interrogating the *CACNA1C* promoter by 4C-seq. Our data help elucidate the molecular mechanism by which one of the best-supported risk loci contributes to schizophrenia through regulation of the *CACNA1C*.

1202W

A human-specific isoform of AS3MT regulated by a human-unique VNTR explains susceptibility to psychiatric illness associated with the 10q24. 32 locus. *M. Li[1], R. Tao[1], A. E. Jaffe[1], J. H. Shin[1], Y. Wang[1], Q. Chen[1], C. Li[1], Y. Jia[1], K. Ohi[1], B. J. Maher[1], R. E. Straub[1], N. J. Brandon[2], A. Cross[2], J. Chenoweth[1], D. Hoeppner[1], H. Wei[1], T. M. Hyde[1], R. McKay[1], J. E. Kleinman[1], D. R. Weinberger[1].* 1) Lieber Institute for Brain Development, Baltimore, MD; 2) AstraZeneca Neuroscience iMED, Cambridge, MA, USA.

Background: Genome-wide association studies (GWAS) have reported many SNPs associated with psychiatric disorders, but knowledge is lacking regarding causative genomic sequences and specific molecular mechanisms. **Methods:** We performed deep RNA-Sequencing in the dorsolateral prefrontal cortex (DLPFC) of human brain from healthy controls and psychiatric patients (N=604). Further analyses were carried out using molecular cloning, qPCR, luciferase assays, DNA methylation analysis, protein characterization in brain and in neuroblastoma and human pluripotent stem cells. **Results:** SNPs across the 10q24. 32 locus showing GWAS association with several psychiatric disorders and spanning multiple genes are associated selectively with the expression of a previously uncharacterized human-specific truncated AS3MT isoform (AS3MTd2d3) in both Caucasian and African American samples of cases and controls (p=1. 15×10-27). These SNPs are in high linkage disequilibrium (LD) with a human unique variable number tandem repeat (VNTR) polymorphism in the first exon of AS3MT; the VNTR risk allele is associated specifically with increased expression of AS3MTd2d3 (p=1. 99x10-30), which is also more abundant in patients than controls (p<0. 0007). VNTR genotype predicts AS3MT promoter activity in luciferase assays and DNA methylation in the spliced 3rd exon (p=9. 31×10-39). AS3MTd2d3 encodes a protein lacking 102 amino acids from the methyltransferase domain, which in brain is comparably abundant to the full-length protein. AS3MTd2d3 is expressed early during human stem cell differentiation toward neuronal fates and lacks arsenic methyltransferase activity. **Conclusions:** These results provide a molecular explanation for the prominent 10q24. 32 locus association, implicating a novel and evolutionarily recent protein involved in early brain development and risk for psychiatric illness.

1203T

Comparative transcriptomic and metabolomic profiling of psychiatric disorders across three brain regions. *R. Ramaker[1,2], K. Bowling[1], B. Lassiegne[1], J. Gertz[1], K. Varley[1], N. Davis[1], P. Cartegena[4], M. Vawter[4], D. Walsh[4], B. Bunney[4], J. Barchas[5], W. Bunney[4], S. Watson[3], H. Akil[3], S. Cooper[1], R. Myers[1].* 1) Department of Genetics, University of Alabama at Birmingham, Birmingham, AL; 2) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 3) University of Michigan, Ann Arbor, MI; 4) University of California at Irvine, Irvine, CA; 5) Weill Cornell, New York, NY.

Neuropsychiatric disorders comprise 13% of the global burden of disease and account for a staggering 184 million disability-adjusted life years (DALYs) worldwide. Schizophrenia (SZ), bipolar disorder (BPD), and major depression disorder (MDD) are multigenic diseases with complex etiology that combine to account for over 50% of these mental health related DALYs. Furthermore, current therapies for these psychiatric disorders are not effective in many patients and may treat only a subset of an individual patient's symptoms. Approaches targeting the underlying molecular pathologies within and across these disorders are necessary to address the enormous psychiatric component of global disease burden and improve care for the millions of Americans diagnosed with these conditions. We performed gene expression profiling by RNA sequencing on dissected post-mortem tissues from the anterior cingulate cortex (AnCg), dorsolateral prefrontal cortex (DLPFC), and nucleus accumbens (nAcc) from four well-documented cohorts of 24 patients each with SZ, BPD, MDD and 24 matched controls (CTL). We also conducted untargeted metabolomic profiling via two-dimensional gas chromatography with mass spectrometry (2D-GCMS) in AnCg tissue from the same subjects across each disorder. The most significant gene expression signals were brain region-specific with the most robust disease differences seen in the AnCg of SZs compared to CTLs although more subtle changes were observed in the DLPFC and nACC. Our identification of overlapping RNA expression profiles between BPD and SZ suggest that a subset of patients with these disorders share a common molecular signature. Combined transcriptomic and metabolomic analyses provide evidence to further previous work demonstrating that dysregulated γ-aminobutyric acid (GABA) and glutamate metabolism are major drivers in a subset of SZ and BPD patients. Examination of transcripts previously associated with individual neuronal cell types indicates different cell populations are present between brain regions and disease groups. Moreover, we show that SNPs identified by GWAS and exome sequencing are overrepresented near genes differentially expressed in BPD and SZ. This study highlights the power of pairing transcriptomic and metabolomic profiling to better understand complex psychiatric disorders.

1204F

Identifying enhancer variants at the *POU3F2* locus associated with both cognition and neuropsychiatric disease. *S. Q. Shen, J. S. Kim-Han, J. C. Corbo.* Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO.

Expansion of the neocortex, especially of the upper layers, is thought to underlie the advanced cognitive capacities of humans, as well as their unique susceptibility to neuropsychiatric disease. Several large genome-wide association studies (GWAS's) have recently identified non-coding variants associated with both increased cognitive performance and increased risk for bipolar disorder. This region is located ~1 Mb upstream of *POU3F2*, a transcription factor important for the development of the upper layers of the cerebral cortex. We hypothesize that the causal variant at this locus exerts its effect by altering *POU3F2* expression levels. Using computational and experimental approaches, we have identified two neurodevelopmental enhancers within this region. These enhancers contain two common human variants that are in strong linkage disequilibrium with the tagging SNPs in the GWAS's, and thus may represent causal variants. These variants affect highly conserved nucleotides and have high predicted pathogenicity. We have generated transgenic reporter mice, as well as mutant CRISPR-Cas mice with germline deletions of the orthologous enhancers, for detailed *in vivo* and *in vitro* phenotyping. In this way, we hope to elucidate the mechanistic link between cognition and neuropsychiatric disease at a molecular level.

1205W

A polymorphism in human α-synuclein affects 3'UTR mRNA isoform expression. *E. S. Barrie, S. K. Handelman, J. T. Frater, M. Kataki, D. Scharre, W. Sadee.* The Ohio State University, Columbus, OH.

Aggregation of alpha-synuclein, encoded by the gene *SNCA,* is the primary structural component of Lewy bodies, the hallmark of Parkinson's disease and frequently found in Alzheimer's disease. Its physiological functions involve dopamine signaling and regulation. The purpose of this study was to identify *cis*-acting regulatory variants in the *SNCA* gene region that affect transcription and mRNA processing. To detect these variants, we use measurements of allelic mRNA ratios in human brain RNA samples. Allelic ratios deviated significantly from unity when measured at a SNP in the middle of the 3χ untranslated region (UTR), but not at its start, suggesting genetic regulation of RNA isoforms with different length 3χUTRs. Scanning of the *SNCA* gene locus in multiple brain samples identified a SNP in the 3χUTR of *SNCA* strongly associated with instances of allelic imbalance. While this SNP does not overlap a microRNA binding site, a non-coding RNA running antisense to the 3χUTR expressed in these brain samples could also play a role. To assess the clinical impact of this variant, we queried the results from multiple neurocognitive genome-wide association studies (GWAS). Although not directly implicated, this candidate regulatory SNP from our study was in linkage disequilibrium with multiple SNPs significantly associated with Parkinson's disease in GWAS, all part of a long haplotype apparently implicated in pathophysiology. Our results indicate that a variant regulates expression of *SNCA* RNA isoforms and may be implicated in disease risk.

1206T

Elucidating the molecular pathways in X-linked dystonia-parkinsonism (XDP): Expression profiling in blood and fibroblasts followed by target transcript analysis in iPSC-derived neurons. *K. Freimann[1], A. Domingo[1,2], A. Westenberger[1], D. Amar[3], L. Lee[2], R. Rosales[2], R. Jamora[2], E. dela Paz[2], R. Shamir[3], P. Seibler[1], C. Klein[1].* 1) Institute of Neurogenetics, University of Lübeck, Lübeck , Germany; 2) XDP Study Group, Manila, Philippines; 3) Edmond J. Safra Center for Bioinformatics, Tel Aviv University, Tel Aviv, Israel.

Objective: Identification of dysregulated genes that might be causative of X-linked dystonia-parkinsonism (XDP) in blood as well as fibroblasts and subsequent validation in iPSC-derived neurons. Background: Genetic studies have suggested that TAF1 (TATA-binding protein associated factor 1), a component of RNA polymerase II-based transcription, plays a critical role in the pathogenesis of XDP. Given that several gaps remain in the exact molecular mechanism of XDP pathogenesis, it is plausible that also further transcriptional differences may be causative of the clinical phenotype. Methods: RNA was extracted from blood and from cultured fibroblasts of XDP patients and controls and subjected to microarray-based genome-wide expression profiling. Fibroblasts were reprogrammed and characterized with respect to pluripotency markers. Three patient and three control iPS cell lines were differentiated into cortical neurons for 66 days. Identified dysregulated transcripts were validated in blood and fibroblasts by quantitative PCR and further analyzed in iPSC-derived neurons. Results: In contrast to blood and fibroblasts, the canonical transcript of *TAF1* was not underexpressed in iPSC-derived neurons of patients versus those of controls (fold change 1. 08). Out of the other analyzed transcripts we saw underexpression of *Synaptotagmin-Like 2* (*SYTL2*, coding for a protein that is involved in exocytosis and vesicle docking) and *Chemokine-Like Receptor 1* (*CMKLR1*, coding for a receptor involved in several signaling cascades) with fold changes of -1. 3 and -5. 7 respectively. Conclusions: In our study, we show the first transcriptional analysis of dysregulated genes in a neuronal model of XDP. Taking clues from microarray analysis in blood and fibroblasts, we found genes related to vesicular function and intracellular signaling to be downregulated in iPSC-derived neurons. Nevertheless, the comparison between the different tissues clearly indicates differential expression strengthening the fact that a cellular model, recapitulating the disease phenotype, of an otherwise inaccessible tissue is needed. Considering that neuronal transcripts of genes other than *TAF1* might be involved in the pathogenesis of XDP, this present cellular model may serve as a valuable tool for future studies.

1207F

Transcriptional networks disentangle cognitive decline from Alzheimer's neuropathology. *S. Mostafavi[1], C. Gaiteri[2], J. Xu[3], C. White[3], S. Sullivan[3], C. McCabe[4], E. Bradshaw[3], T. Young-Pearse[3], L. Yu[2], L. Chibnic[3], N. Pochet[3], J. Schneider[2], D. A. Bennett[2], P. L. De Jager[3,4].* 1) University of British Columbia, Vancouver; 2) Rush University; 3) Harvard Medical School; 4) Broad Institute.

Age-related cognitive decline is Alzheimer's disease (AD) is associated with multiple neuropathologies that are linked to impairment of neuronal function and eventual neuronal loss. These neuropathologies accumulate slowly during the long asymptomatic prodromal phase of AD. Only 21 variants have been associated with AD susceptibility despite large GWAS, and the genomic architecture of this disease and age-related cognitive decline remains poorly characterized. Here, we take a complementary approach and leverage RNA-sequencing data from dorso-lateral prefrontal cortex of 507 subjects from a longitudinal aging cohort, in order to resolve the transcriptional changes related to cognitive decline from those that are related to brain pathology. In particular, in order to tease apart merely correlated from likely consequential associations related to cognitive decline, we developed a probabilistic model for jointly analyzing multiple AD-related traits, confounding factors such as cell type composition, and modules of co-expressed genes. Our approach identifies co-expressed gene modules relevant to AD-pathology and cognitive decline and learns a directed Bayesian network that relates module-level gene expression to AD-related traits and cell type composition. From 47 modules observed in the human brain, this approach identifies 6 modules of co-expressed genes that are directly associated (p<1e-6) with cognitive decline or AD pathology. Inflammation-related gene modules appear to play a modest, indirect role in this network, and, using four different approaches and AD GWAS results, we find that today's limited genetic architecture of AD susceptibility does not strongly influence our network structure. Rather, it highlights the importance of Module 109, enriched for genes related to cell cycle regulation, as being proximal to cognitive decline in the sequence of molecular events leading to dementia (p<1e-8). While it has an effect on amyloid pathology, its effect on cognitive decline is largely mediated through non-amyloid processes. We replicate this finding in both an additional set of 96 subjects from our cohort (p<0. 005) and in an independent microarray dataset from frontal cortex of 300 individuals (p<1e-6). shRNA and ORF perturbation of elements of module 109 in human astrocytes supports the coherence of this module*in vitro*form a preliminary basis to characterize its effects on AD-relevant outcomes such as APP processing or synaptic function.

1208W

Transcriptome analysis of human induced pluripotent stem cells and neural progenitor cells in Tourette's Disorder. *N. Sun[1,2], RP. Hart[2,3,4], L. Deng[1,2], JC. Moore[1,2,3], JA. Tischfield[1,2,3], GA. Heiman[1,2].* 1) Genetics, Rutgers University, Piscataway, NJ; 2) Human Genetics Institute of New Jersey, Piscataway, NJ; 3) RUCDR Infinite Biologicals NIMH Stem Cell Resource, Piscataway, NJ; 4) Department of Cell Biology and Neuroscience, Rutgers University, Piscataway, NJ.

Tourette Disorder (TD) is a heritable neurodevelopmental disorder characterized by the presence of both motor and vocal tics. Deleterious mutations in protein coding genes have been identified in subsets of TD individuals[GAH1]. However, no single gene with recurrent mutations has been reported. To better understand the genetic architecture of TD, we studied members of three multiplex families (families with multiply-affected individuals) that contain a family-specific variant co-segregating with TD. These variants are known to change the function of the gene product. Understanding how these variants alter gene expression of neurobiological pathways can help clarify the neuropathology of TD and suggest other genes involved in those pathways. To look for gene expression alterations during early neural development, we performed RNA sequencing on human induced pluripotent stem cells (iPSCs) and neural progenitors cells (NPCs) derived from individuals with and with TD from each family. By using a customized analytical pipeline comprised of publically-available bioinformatics tools, the gene expression profiles of TD and control iPSCs/NPCs were compared and differentially expressed genes (DEGs) were selected. Neurobiological pathways in which the DEGs are enriched in are identified using multiple pathway analysis and annotation approaches. 120 DEGs from NPCs of one family and 346 DEGs from iPSCs of another family were initially analyzed. We will study the functional changes of common pathways shared by the three families and further investigate cellular phenotypes in TD.

1209T

Gene expression profiling in LRRK2-G2019S purified iPSC-derived dopaminergic neuronal models of Parkinson's disease. *C. Webber[1], C. Sandor[1], P. Robertson[2], A. Heger[3], R. Wade-Martins[2].* 1) MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK; 2) Oxford Parkinson's Disease Centre, Department of Physiology, Anatomy and Genetics, Le Gros Clark Building, University of Oxford, Oxford, UK; 3) MRC Computational Genomics Analysis and Training Programme, MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, UK.

The *LRRK2* **G2019S** variant is significantly associated with both familial and sporadic forms of Parkinson's disease **(PD)**. However, the molecular pathways perturbed by this variant that predispose to the pathological loss of nigral dopaminergic neurons remain unknown. To identify these pathways, we compared the RNA-sequencing profiles of purified populations of induced pluripotent stem cell **(iPSC)-**derived dopaminergic midbrain neurons originating from three PD patients carrying *LRRK2* G2019S mutations to profiles similarly obtained from three controls. Comparison of the transcriptional profiles pre- and post-purification demonstrated that cellular heterogeneity accounts for more of the transcriptional variation than the genetic background. Amongst 53 postmortem tissue profiles reported by the Genotype-Tissue Expression **(GTEx)** project, our iPSC-derived dopaminergic midbrain neuron transcriptional profiles were most similar to the nigral tissue profile. Comparing the PD *LRRK2* G2019S line transcriptional profiles to controls identified 168 protein-coding genes that were differentially expressed (DE) between the cases and controls. Network analyses showed these 168 DE genes to be functionally-coherent with distinct clusters of functionally-related genes demonstrating concordant increases or decreases in expression. Notably, DE genes were enriched in genes whose orthologues' disruption in the mouse yields abnormal motor capabilities/coordination/movement phenotypes (**ab motor c/c/m**; FDR < 5%) and this enrichment concurringly segregates with those DE genes that demonstrate reduced expression. Of the 14 PD GWAS risk loci that included a gene expressed in these iPSC-derived neurons, 4 loci harboured a DE gene (p=0. 023) including *SYT4* (Synaptotagmin IV) which is also associated with ab motor c/c/m phenotypes. Analyses of the counteracting drug response profiles through Connectivity Map identifies several significant associations that include clioquinol, acompoundknown to rescue Parkinsonism and dementia phenotypes of the tau-knockout mouse. These results demonstrate the potential of iPSC-derived models to capture the molecular perturbations associated with a neurodegenerative disorder such PD.

1210F

The effects of perinatal Fluoxetine (FLX) exposure in the modulation of neurodevelopment using the *Shank3* $^{e4\text{-}9/+}$ mutant mouse model. *J. Murry[1], D. Engler[3], C. Ljungberg[2], I. Van den Veyver[1,2,4].* 1) Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Obstetrics & Gynecology, Baylor College of Medicine, Houston, TX; 3) Proteomics Core at Houston Methodist Research Institute, Houston, TX; 4) The Jan and Dan Neurological Research Institute at Texas Children's Hospital TX, 77030.

Autism spectrum disorder (ASD) manifests with repetitive behaviors, social and communication impairments. Early environmental exposures concurrent with genetic factors are implicated in ASD. Genes like *SHANK3,* affecting synaptic function, are causative in a subset of ASD patients and *Shank3* null mice exhibit ASD-like behavior. Perturbed serotonergic signaling has been reported in ASD and FLX, a Selective Serotonin Reuptake Inhibitor, is an increasingly common early life exposure. We hypothesized that ASD-associated risk alleles (*Shank3*) in conjunction with perinatal exposures (FLX) modulate neurodevelopment, increasing penetrance of offspring behavior disturbances. Using a *Shank3* mutant mouse model, HET dams (n=21-28) were exposed to FLX (15mg/kg) in drinking water during pregnancy (dpc0) and lactation (until PND7) and mated with HET males. FLX and its active metabolite, Norfluoxetine in plasma was within range by mass spectrometry. Adult male and female offspring (KO, HET, WT) (+/-FLX) were tested for anxiety-like, repetitive, and social behaviors. After behavior testing, brains were harvested for histological and molecular analysis. Two-way ANOVA was used to test gene-environment (GXE) interactions, with Tukey's posthoc test. Student's t-test was used to test preferences. Litter sizes of HET(+FLX) dams differed significantly from HET(-FLX) dams)(p<0. 05), with no difference in adverse pregnancy outcomes. There were minor weight differences at P28 in HET(+FLX) compared to HET(-FLX) males that equalized upon aging, but differences between KO(+FLX) and KO(-FLX) females (p<0. 05)persisted. In elevated plus maze, KO(-FLX) males had more open arm entries vs. WT and HET(-FLX) (p<0. 05) and traveled greater distance in open arm than WT(-FLX)(p<0. 001), with a significant GXE interaction (p<0. 05), but exposure resulted in a trend towards more anxiety in KO males. In open field assay, vertical locomotion was lower in KO(+FLX) males vs. HET(+FLX)(p<0. 01) with a significant GXE interaction (p<0. 01). In three-chamber test, both KO(+FLX) males and females lacked social preference (p=0. 8; 0. 4*)*. In direct social interaction test, WT and HET(+FLX) males have increased active social contact vs. WT and HET(-FLX) males (p<0. 001; 0. 05). No repetitive behavior was observed. These data suggest that chronic FLX exposure may inhibit social and anxiety-like behaviors. Ongoing experiments will address molecular and structural effects of developmental FLX exposure in these mice.

1211W

Multimodal twin research approach to identify genetic and environmental factors in neurodevelopmental disorders. *K. Tammimies[1], C. Willfors[1], M. Arora[2], B. Thiruvahindrapuram[3], V. Moricz[1], B-M. Anderlid[4], A. Nordgren[4], E. Cauvet[1], J. Neufeld[1], M. Linde[5], A. Vantwesteinde[1], J. Holmén-Larsson[6], D. Rajagopalan[3], S. Berggren[1,7], C. Coco[1], E. Kostrzewa[1], P. Fransson[8], C. Dalman[9], H. Anckarsäter[10], S. W. Scherer[3], A. Reichenberg[11], M. Landén[10,12], R. Diaz-Heijtz[5], A. Falk[5], P. Lichtenstein[12], S. Bölte[1,7].* 1) Women's and Children's Health, Center of Neurodeve, Karolinska Institutet, Stockholm, Sweden; 2) Department of Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, United States; 3) The Centre for Applied Genomics, Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Canada; 4) Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden; 5) Department of Neuroscience, Karolinska Institutet, Stockholm, Sweden; 6) Clinical Neurochemistry Laboratory, Institute of Neuroscience and Physiology, Department of Psychiatry and Neurochemistry, The Sahlgrenska Academy, University of Gothenburg, Mölndal, Sweden; 7) Child and Adolescent Psychiatry Stockholm, Center for Psychiatry Research, Stockholm County Council, Stockholm, Sweden; 8) Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden; 9) Public Health Epidemiology, Department of Public Health Sciences, Karolinska Institutet, Stockholm, Sweden; 10) Institute of Neuroscience and Physiology, the Sahlgrenska Academy at The Gothenburg University, Gothenburg, Sweden; 11) Departments of Psychiatry and Preventive Medicine, Mount Sinai School of Medicine, New York, NY, USA; 12) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden.

Neurodevelopmental disorders (NDDs), such as autism spectrum disorders (ASD) and attention deficit hyperactivity disorder (ADHD), are highly heterogeneous conditions with a strong genetic component. Also the contribution of environmental factors has been increasingly acknowledged. The Roots of Autism and ADHD Twin Study in Sweden (RATSS) is an ongoing project investigating monozygotic (MZ) and dizygotic twins carefully assessed for NDD phenotypes. In addition to the detailed neuropsychological profiles, we have collected medical records, brain imaging data and multilevel biological samples such as blood, cerebrospinal fluid, skin biopsies and baby teeth to enable multimodal approach to investigate the contribution of genetic and environmental factors in these disorders. To date more than 100 twin pairs have participated in RATSS. We are currently analyzing possible post-twinning somatic mutations and putative pathogenic mutations in the twins using whole exome sequencing (WES), whole genome sequencing and DNA microarrays. In parallel, the biological samples are analyzed for proteomics, metal concentrations and metabolomics to identify co-twin and global differences associated with NDDs and correlated with the genetic findings. In our pilot study, WES data was acquired for nine MZ twin pairs either primary or secondary discordant for ASD and ADHD. No early somatic mutations were found in these twin pairs, however putative pathogenic mutations were identified in three of the nine pairs. In all of these three pairs, medical records revealed that both twins have earlier exhibited symptoms in the NDD spectrum. For example, in male MZ twin pair showing differences in developmental trajectories for ASD and ADHD, a maternally inherited hemizygous splice site variant affecting *CASK* was identified. This variant resulted in two different mutant transcript (p. del389-415; p. A406fs*4). We are now analyzing neurons derived from immunopluripotent stem cells from this twin pair to identify cellular phenotypes associated with the *CASK* mutation and how the different environmental components identified in the other screens affect the splicing pattern caused by the mutation to understand the discordant phenotypes seen in the twins. Our multimodal approach and powerful co-twin design will be able to provide novel insights into penetrance and expressivity of genetic variants and explore the genetic x environment interactions to better understand the heterogeneous nature of NDDs.

1212T

An integrative disease-relevant multi-omics analysis to predict risk for stress-related psychiatric disorders. *J. Arloth[1,2], N. Møller[2], E. B. Binder[1,3].* 1) Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich,Germany; 2) Institute of Computational Biology - Helmholtz Zentrum Müchen, Neuherberg, Germany; 3) Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA.

Depression risk is exacerbated by genetic factors and stress exposure; however, the biological mechanisms through which these factors interact to confer depression risk are poorly understood. One putative biological mechanism implicates variability in the ability of cortisol, released in response to stress, to trigger a cascade of adaptive genomic and non-genomic processes through glucocorticoid receptor (GR) activation. We applied expression quantitative trait locus (eQTL) analysis using genome-wide gene expression data (Illumina HT12v3, v4) from GR-stimulated gene expression in peripheral blood cells of 300 individuals and genome-wide imputed SNP array data (Illumina 660W-Quad, OmniExpress) to identify genetic variants associated with GR-induced gene expression changes. Secondly, by adopting an expression quantitative trait DNA methylation (eQTM) approach using a linear regression model with an Elastic net penalty, we assessed the predictive ability of the transcriptional response to glucocorticoids and distinct methylation patterns (Illumina 450k). We show that GR-response eQTLs are more distant than SNPs that regulate the baseline gene expression as well as DNA methylation status. One possible explanation for this long-range regulation could be a chromatin interaction between the eSNP locus and the regulated transcript. Which we validated using both ChIA-PET as well as chromatin confirmation capture analysis. Moreover, we could show that such eQTL SNPs are highly enriched among SNPs associated with MDD relative to baseline eQTL SNPs as well as schizophrenia (FDRs<0. 001) and the PGC cross-disorder analysis (FDR<0. 066). The SNPs showing an association with both GR-mediated transcription and MDD cumulatively influence intermediate phenotypes for psychiatric risk such as amygdala activation to threat, as shown in an independent imaging cohort (n = 647). These cumulative GR-risk SNPs also interact with exposure to childhood adversity to predict mood and anxiety disorders in adulthood. Furthermore, network analysis suggests that the eQTL transcripts functionally interact to perform an orchestrated function. We currently explore whether epigenetic changes associated with these GR-response eQTLs could provide additional insights into disease risk mechanisms. Disease-relevant stimuli in eQTL and eQTM approaches can expand our understanding of the genetic basis of stress-related disorders, in which the GR-function plays an important pathophysiologic role.

1213F

Snus use is not associated with variation in the *CHRNA5-CHRNA3-CHRNB4* gene cluster in Sweden, Norway and Finland. *P. A. Lind[1], D. J. Benjamin[2], J. Buchwald[3], D. Cesarini[4], M. E. Gabrielsen[5], J. K. Hällfors[3], M. Johannesson[6], J. Kaprio[3], T. Korhonen[3], H. E. Krokan[7], A. M. Loukola[3], P. K. E. Magnusson[8], S. E. Medland[1], R. J. Rose[9], F. Skorpen[5], E. B. Stovner[10].* 1) Quantitative Genetics, QIMR Berghofer Medical Research Institute, Herston, Australia; 2) Center for Economic and Social Research, University of Southern California, Los Angeles, CA; 3) Department of Public Health, Faculty of Medicine, University of Helsinki, Helsinki, Finland; 4) Center for Experimental Social Science, Department of Economics, New York University, New York, NY; 5) Department of Laboratory Medicine, Children's and Women's Health, Norwegian University of Science and Technology, Trondheim, Norway; 6) Department of Economics, Stockholm School of Economics, Stockholm, Sweden; 7) Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway; 8) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 9) Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN; 10) Department of Computer and Information Science, Norwegian University of Science and Technology, Trondheim, Norway.

Meta-analyses of genome-wide association studies have established a robust association between the number of cigarettes smoked per day (CPD) and the *CHRNA5-CHRNA3-CHRNB4* gene cluster on chromosome 15q25 which encode nicotinic acetylcholine receptor subunits. Variation in this gene cluster is also associated with nicotine dependence and a wide range of cigarette smoking behaviours. This meta-analysis examines snus use in three Scandinavian cohorts to assess whether the gene cluster is also associated with other forms of tobacco use. Three phenotypes were analysed; age of onset of snus use (AOI), ever versus never snus use (EVRSNUS) and the number of boxes of snus consumed per month (BPM). Association analyses included cigarette use (ever or regular smoking), gender, age, birth year and Principal Components as covariates. This study included a maximum of 7619 participants spanning three countries (Sweden, Norway and Finland): 1474 EVRSNUS cases and 6145 controls, 1205 subjects with AOI and 1247 subjects with BPM data. Meta-analyses were conducted in METAL using summary statistics. There was little or no evidence for association between the snus traits and 15q25 variants (non adjusted P-values > 0. 001). However, three genome-wide significant associations ($p < 5E-08$) were observed in *CHRNB4* for the cigarette-related trait CPD in the Swedish sample. These results provide evidence that the *CHRNA5-CHRNA3-CHRNB4* gene cluster does not have a significant effect on snus use and that the influence of variants in this region on tobacco use may depend on the mode of ingestion.

1214W

Toxicogenetic approach reveals strong genetic control of environmental toxicant levels and surprising relationship with autism susceptibility. *M. Traglia[1], L. A. Croen[2], K. Lyall Dodge[3], A. R Torres[4], M. Kharrazi[5], G. N DeLorenze[2], O. Zerbo[2], G. C Windham[3], L. A Weiss[1].* 1) Department of Psychiatry and Institute for Human Genetics, University of California, San Francisco, 401 Parnassus Ave, LangPorter, San Francisco, CA 94143, USA; 2) Autism Research Program, Division of Research, Kaiser Permanente, 2000 Broadway, Oakland, CA 94612, USA; 3) Division of Environmental and Occupational Disease Control,California Department of Public Health, 850 Marina Bay Pkwy, Bldg. P, Richmond, CA 94804, USA; 4) Center for Persons with Disabilities, Utah State University, 6804 Old Main Hill, Logan, UT 84322-6804, USA; 5) Genetic Disease Screening Program, California Department of Public Health, 850 Marina Bay Pkwy, Richmond, CA 94804, USA.

Maternal exposure to environmental pollution could affect fetal brain development and increase Autism Spectrum Disorder (ASD) risk due to differential genetic susceptibility. To investigate this hypothesis, we analysed genome-wide SNP data along with a set of 7 polybrominated diphenyl ethers, a class of organohalogens used as flame retardants, which accumulate in humans and could act as neurodevelopmental toxicants. Levels of the flame retardants (PBDE 28, 47, 99, 100, 153, and BB153, and their sum)were measured in mid-pregnancy blood samples as previously described (Sjodin, Anal Chem 2004) in a genome-wide genotyped population of 790 women in California (Tsang, PLoS ONE 2013). The offspring from these pregnancies are 390 ASD affected and 400 control children studied in the EMA (Early Markers of Autism) study (Croen, Autism Res 2008). Log10-transformed PBDE levels showed high estimated SNP-based heritability (h2g) accounting for 44% to 87% of variance (P<0. 05, each) in a REML model (Yang, Nat Genet 2010) adjusted for genetic population stratification, maternal education level and country of birth, offspring gender and month of birth, as well as offspring affection status. In addition, 5 out of 7 PBDEs were associated with offspring affection status (P<0. 05, each). Genome-wide association via linear regression highlighted two associated loci for PBDE levels: one locus reached Bonferroni-adjusted significance and maps between ADAMTSL1 and SH3GL2 (PBDE28; P=3x10-8). ADAMTSL1 is involved in extracellular matrix functions, SH3GL2 in synaptic vesicle endocytosis and is associated with cognitive impairment. The second locus maps to CYP2B6 (PBDE99; P=1x10-7) that encodes a lipase oxidase, recently associated with a different class of organohalogens (PCBs). Our results suggest that genetic make-up is a major determinant of circulating levels of environmental toxicants, and imply that an observation of an association between a presumed toxicant and ASD could be due to pleiotropy of pathways involved in xenobiotic metabolism. The genes near loci implicated in controlling PBDE levels include xenobiotic metabolism genes, and genes with associations or functions in neurodevelopment. We thus conclude that future studies investigating associations between environmental exposures and neurodevelopmental outcomes should consider interactions with underlying genetic background.

1215T

The role of translational control in schizophrenia. *A. Dang Do, D. Avramopoulos, D. Valle.* McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

The complex but strongly familial expression, late onset, and variable phenotype of schizophrenia suggests a mechanism where environmental factors influence genetic expression. Genome-wide association studies (GWAS) and whole exome sequencing (WES) have provided many potential genetic targets. However, the neurobiological connection between sequence variants and disease is largely uncharacterized. Regulation of mRNA translation initiation is a major mechanism for selective protein expression in response to specific environmental stressors. This modulation works in parts through the presence of upstream open reading frames (uORF) or 5' terminal oligopyrimidine tract (5'-TOP). We hypothesize that transcripts of genes associated with schizophrenia contain 5' non-coding region (NCR) sequences that are targets for translational regulation. To test this we examined the intersections between published data sets from GWAS and WES studies of individuals with schizophrenia, and from *in silico* and experimental studies of transcripts containing uORF or 5'-TOP sequences. Of transcripts from genes highly expressed in nerve tissue 1. 7% overlap with the uORF transcript set and 4. 6% with the 5'-TOP transcript set. In comparison, of the transcripts from genes associated with schizophrenia by GWAS 1. 8% (Fisher's one-tailed p = 0. 52, compared to nerve transcript set) contain uORF, and 3. 4% (p = 0. 15) contain 5'-TOP sequences. Of the transcripts from genes associated with schizophrenia by WES 1. 4% (p = 0. 30) contain uORF, and 5. 6% (p = 0. 15) contain 5'-TOP sequences. When examined at the gene level, significantly more genes identified from the schizophrenia WES study have transcripts containing 5'-TOP sequences than those from the nerve gene set (14. 3% vs 8. 3 %, respectively; p = 0. 013). Factors explaining these results may include: 1) WES studies focus on exonic variants, thus variants in the NCR that may associate with schizophrenia may be missed; 2) The enrichment of transcripts with uORF or 5'-TOP sequences contributing to the pathology of schizophrenia may be specific to certain biological pathways or processes; 3) The differential regulation of transcript translation contributing to the pathology of schizophrenia may occur through regulation of proteins in the eukaryotic initiation factor 2 or mechanistic target of rapamycin pathways and not at the sequence level. We are exploring these possibilities.

1216F

Expression Profiles of miRNA in Peripheral Blood Associated with Post-Traumatic Stress Disorder in Military Service Members. *C. Martin[1], S. Yun[2], J. Gill[1], H. Rusch[1], H. Kim[1], A. Cashion[1].* 1) National Institute of Nursing Research, Bethesda, MD; 2) Yotta Biomed, Potomac, MD.

The etiology of Post-Traumatic Stress Disorder (PTSD) likely involves the interaction of numerous genetic and environmental factors via gene expression changes, which may be potential indicators of neural processes. MicroRNA (miRNA) is a potential regulator of gene expression, and miRNAs have been identified as biomarkers for a range of medical conditions from cancer to autoimmune disorders. However, it is not yet clear how expression of miRNA affects maintenance of and recovery from symptoms in PTSD. We tested the following hypotheses: (1) detectable levels of miRNA will be present in peripheral blood of combat-exposed military service members following deployment (2) post-trauma expression levels of miRNA in peripheral blood would differ between participants based on PTSD symptomology. Ion Torrent PGM was used to sequence miRNA in peripheral blood samples from 32 military personnel (11 PTSD-maintained subjects, 11 PTSD-improved subjects, and 10 non-PTSD control subjects) following return from deployment to war zones in Iraq or Afghanistan and again at a 6 month follow-up. Whole-genome miRNA profiles were generated by trimming and aligning sequences to both hg19 and miRBase20, and expression of annotated genes were compared by ANOVA run on an in-house PERL script. Preliminary results demonstrated presence of miRNA sequences in all peripheral blood samples. Mean length of pre-trimmed reads was 40bp. They were trimmed to 25bp and aligned to the hg19 reference genome. Of these reads, 55. 5% aligned to the miRBase20 reference set, suggesting that some of the miRNA associated with PTSD may be novel sequences. Of the miRNA sequences that were mapped, differences were noted both in variety and relative abundance of sequences between tested groups. This study is ongoing and current findings are being verified with quantitative PCR. Results will also be corroborated with microarray data of mRNA expression. These findings will lead to an increased understanding of how dynamic regulation of genes associated with PTSD affects presentation of clinical symptoms, as well as identify potential biomarkers that can be useful in a military context as a diagnostic tool to inform post-deployment medical screening, and identify appropriate treatment strategies.

1217W

Lower Extremity Skeletal Muscle DNA Methylation in Chronic Stroke. *H. Xu[1], AS. Ryan[2,3,4], TG. Forrester[5], YC. Cheng[1], A. Parihar[1], K. Tanner[1], JW. Cole[6], BD. Mitchell[1,7], CE. Hafer-Macko[6,8,9], RF. Macko[2,4,6,8,10].* 1) Division of Endocrinology, Diabetes, and Nutrition, University of Maryland School of Medicine, Baltimore, MD; 2) VA Maryland Exercise and Robotics Center of Excellence(MERCE), Baltimore, MD; 3) The Geriatric Research Education and Clinical Center (GRECC), Baltimroe Department of Veterans Affairs Medical Center,Baltimore, MD; 4) Department of Medicine, University of Maryland School of Medicine, Baltimore, MD; 5) UWI Solutions for Developing Countries, University of the West Indies, Mona Campus, Kingston 7, Jamaica; 6) Department of Neurology, University of Maryland School of Medicine, Baltimore, MD; 7) Department of Epidemiology and Preventive Medicine, University of Maryland School of Medicine, Baltimore, MD; 8) Rehabilitation Research and Development's Maryland Exercise and Robotics Center of Excellence, Baltimore VAMC, Baltimore, MD; 9) Myasthenia Gravis Center, University of Maryland School of Medicine, Baltimore, MD; 10) Department of Physical Therapy and Rehabilitation Science, University of Maryland School of Medicine, Baltimore, MD.

Stroke is the number one cause of serious long-term disability including extremity motor dysfunction in the United States. Skeletal muscle in paretic limbs switches toward fast twitch fiber-type with reduced oxidative capacity, increased intramuscular fat, inflammation, and atrophy, compared to non-paretic muscle. There are differences in muscle transcriptome profiles that relate to these specific phenotypic pathologies, which we postulate could result from post-stroke epigenetic changes. We hypothesize that hemiparetic stroke alters paretic leg skeletal muscle genomic DNA methylation profiles related to these transcriptome and phenotypic abnormalities, compared to non-paretic leg. Five males (mean age 56. 6 ±14. 5 years) with chronic (>6 months) hemiparetic gait post-stroke underwent bilateral vastus lateralis muscle biopsies. Methylation status of genomic DNA was assessed by Illumina 450K Methylation array. We identified 54 (P<0. 001) or 331 (p<0. 005) CpG loci showing differential DNA methylation status between paretic and non-paretic muscle (>1. 3 fold changes). Pathway analysis of the top 100 genes with most significant CpG loci revealed enrichment for those related to remodeling/plasticity, neuronal signaling, energy metabolism and inflammation. Methylation sites mapping to genes related to skeletal and muscle disorders (11 genes) and connective tissue disorders (10 genes) were over-represented in these top 100 genes. Consistent with our previous observation that myogenin (*MYOG*), a muscle-specific myogenesis transcription factor, has lower mRNA expression (-1. 57 fold) in paretic muscle, we show *MYOG*'s upstream transcription regulators, *SIX4* (p= 3. 0*10-3) and *HDAC4* (p= 4. 8*10-3), have reduced methylation at the 1st intron and 2nd intron in paretic vs. non-paretic muscle. Conclusion: We report that hemiparetic stroke produces a distinct profile of genomic DNA methylation alterations in paretic leg skeletal muscle that is consistent with phenotypic and transcriptome abnormalities related to energy metabolism, contractile activity, atrophy and structural remodeling. Further understanding of epigenomic mechanisms regulating skeletal muscle structure and function will facilitate the discovery of therapeutic targets for post-stroke rehabilitation and recovery.

1218T

Sp1 modulation of APP and BACE1 activity as a drug target in Alzheimer's disease. *B. L. Bayon[1], K. Nho[2], B. Maloney[3], N. Chopra[4], D. K. Lahiri[1,3,4].* 1) Department of Medical & Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN; 2) Department of Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN; 3) Department of Psychiatry, Institute of Psychiatric Research, Indiana University School of Medicine, Indianapolis, IN; 4) Department of Neuroscience, Indiana University School of Medicine, Indianapolis, IN.

β-site APP Cleaving Enzyme (BACE1) is the β-secretase responsible for the rate-limiting cleavage of amyloid-b precursor protein (APP) to amyloid-b (Ab), which oligomerizes and contributes to the plaque formation seen in Alzheimer's disease (AD) which generally occurs in the later stages of life. Expression levels of some transcription factors (TFs) such as Sp1 are perturbed in a latent fashion. Sp1 positively regulates APP and induces BACE1 via their respective promoters. We tested Sp1-mediated regulation of APP with an Sp1 inhibitor, Mithramycin A (MTM), and Tolfenamic acid (TA), an inducer of Sp1 degradation, and with siRNAs in mammalian cell lines (rat neuronal PC12 and human glioblastoma U373), a primary human fetal neuron culture (HFN) and mixed cultures derived from human fetal neurospheres (NSPc). Treatment of PC12 reveals minimal changes in confluence, cytotoxicity, neurite length, and neurite outgrowth after Sp1 knockdown via siRNA or treatment with Sp1 modulating drugs. Morphology and cell death tracking studies from U373 reveal increasing cytotoxicity in MTM concentrations above 10 μM after 36 hours. Treatment of HFN with TA did not affect cell viability with doses up to 5 μM. Western blotting shows a significant decrease in expression of BACE1 after MTM treatment in NSPc and U373. Treatment with TA does not significantly decrease APP or BACE1 in NSPc or U373. APP siRNA knocks down expression of APP in both of these cell types. APP expression is not altered by treatment with TA in NSPc, perhaps due to disparate mechanisms of these Sp1-inhibiting drugs. MTM reduces APP and BACE1 expression. Treatment with Sp1-inhibiting drugs or transfection with Sp1 siRNA does not affect cell viability of primary neurons or differentiated NSPc. We chose single nucleotide polymorphisms (SNPs) within the Sp1 gene from the AD Neuroimaging Initiative GWAS data, performed an association analysis with an AD-specific imaging biomarker (entorhinal cortex thickness), and identified a significant SNP (rs11170553) associated with entorhinal cortex thickness. rs11170553 was also associated with cerebral amyloid deposition. Compounds which modify Sp1 binding to sites on the BACE1 and APP promoters may provide a means to limit the production of Ab peptide and could slow the symptoms of AD. Our results show that appropriate modulation of a specific TF can be a novel drug target for AD.

1219F

Extracellular RNA sequencing to identify RNA biomarkers of head impact in college athletes. *R. Richholt[1], A. Yeri[1], R. Mccoy[2], M. Anastasi[2], S. Althoff[1], A. Allen[1], A. Siniard[1], M. DeBoth[1], I. Malenica[1], T. Beecroft[1], E. Carlson[1], L. Ghaffari[1], S. Allen[1], M. Shahbauder[1], K. Ryden[1], R. Bruhns[1], A. Janss[1], D. Vooletich[3], T. Ide[3], D. Arment[3], D. Leonard[4], J. Chu[4], A. Buck[4], T. Mcleod[5], J. Cardenas[5,6], R. Greenwald[4], T. Lee[3], J. Trent[1], K. Van Keuren-Jensen[1], M. Huentelman[1].* 1) Neurogenomics, Translational Genomics Research Institute, Phoenix, AZ; 2) Arizona State University (ASU), Sun Devil Athletics, Phoenix, Arizona; 3) Riddell Sports, Division of Easton Bell Sports, Rosemont, Illinois; 4) Simbex, Lebanon, New Hampshire; 5) Barrow Neurological Institute (BNI), Department of Child Neurology, Phoenix, Arizona; 6) A. T. Still University (ATSU), Department of Interdisciplinary Health Sciences, Mesa, Arizona.

Objective

When brain injuries occur, rest is essential for recovery. Whenthese injuries go undetected, prognosis worsens and there is increased likelihood ofrepeated injuryandlasting symptoms. Rapid objectivediagnosis in easily obtainable samples is critical to reduce the impact of traumatic brain injury. To work towards that goal, we studied college-aged football players using a combination of in-helmet monitoring andbiospecimenbased RNA biomarker profiling.

Methods

Measurements of head impacts were recorded using the Riddell SRS at every contact practice and game. Urine samples were collected from Arizona State University football players weekly within 24 hours following the completion of a game. Control urine samples were collected from a group of non-contact athletes of similar age as well as from every enrolled player prior to their first full contact practice. Extracellular RNA was isolated from urine with the NORGEN Total RNAPurification Maxi Kit. 2ng of total RNA werereverse transcribed with theClontechSMARTerUniversal Low Input RNA kit, andwhole RNAtranscriptomesequencing libraries were prepared with theClontechLow Input Library Prep kit. Libraries were sequenced on theIlluminaHiSeq2500 with paired-end 83bp runs. Gene counts were computed and normalized with DESeq2. Classification of samples into case and control groups was performed using aRandomForestapproach. 25 genes providing the highest information gain were selected andthentested by classifying cases and controls.

Results

From an average urine specimen, we generated 39 million reads and identified 41 thousand transcripts. Using a Random Forest algorithm, we classified cases and controls for highest magnitude hits and highest frequency hits. 25 genes with the highest information gain were selected for classification. In a preliminary validation of 6 cases and 16 controls, high magnitude hitters were classified with 100% sensitivity and 75% specificity. High frequency hitters were classified with 100% sensitivity and specificity.

Conclusions

This is an early attempt to assess brain injury by quantitativetranscriptomeanalysis. Accurate, unbiased diagnosis of TBI is critical to prognosis of individuals with brain injury. These data suggest that extracellular RNA expression changes following head impacts can be detected in peripheralbiofluidsand serve as biomarkers of TBI.

.

1220W

L-type voltage-sensitive calcium channel subunit (LVSCC)-A1C is associated with increased Parkinson disease risk only when plasma vitamin D concentration is deficient. *L. Wang[1,2], L. Maldonado[1], G. W. Beecham[1,2], E. R. Martin[1,2], M. L. Evatt[3], J. C. Ritchie[4], J. L Haines[5], C. P. Zabetian[6,7], H. Payami[8,9], M. A. Pericak-Vance[1,2], J. M. Vance[1,2], W. K. Scott[1,2].* 1) John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami Miller School of Medicine, Miami, FL; 3) Department of Neurology, Emory University, Atlanta, GA; 4) Department of Pathology, Emory University, Atlanta, GA; 5) Department of Epidemiology and Biostatistics, and Institute for Computational Biology, Case Western Reserve University, Cleveland, OH; 6) Veterans Affairs Puget Sound Health Care System, Seattle, WA; 7) Department of Neurology, University of Washington, Seattle, WA; 8) Departments of Neurology and Genetics, University of Alabama-Birmingham, Birmingham, AL; 9) HudsonAlpha Institute for Biotechnology, Huntsville, AL.

The hallmark of Parkinson disease (PD) is the loss of substantia nigra dopaminergic neurons (SNDNs), which have sustained opening of L-type voltage-sensitive calcium channels (LVSCC) for their pacemaker activity. Maintaining intracellular Ca2+ homeostasis in the face of extended Ca2+ influx puts SNDNs under substantial cellular stress making them vulnerable to cell death. *In vitro* studies show that vitamin D (vit D) promotes neuronal Ca2+ homeostasis by decreasing the expression of the A1C subunit of LVSCC, encoded by the *CACNA1C* gene. Recently, vit D deficiency has been associated with higher risk of PD. We hypothesize that the "missing heritability" in genome-wide association studies (GWAS) in PD includes gene-environment (GxE) interactions. To investigate *CACNA1C*-vit D interaction in modifying PD risk, we utilized two GWAS with vit D measurements: one with 477 cases and 430 controls genotyped on Illumina 610quad chips and one with 482 cases and 412 controls genotyped on Illumina Omni 1M chips. Vit D deficiency (plasma vit D <20 ng/ml) was associated with PD in both datasets (Odds Ratio (OR)=2. 7, *P*<0. 0001; OR=1. 9, *P*=0. 009). Genotypes in both datasets were imputed to 5. 3M SNPs (after quality control) using 1000Genomes reference panels. Joint meta-analysis (JMA) of GxE interaction was conducted by comparing a model containing vit D deficiency, SNP dosage, an interaction term, and covariates (age, sex and sampling season) to a restricted model with only vit D deficiency and covariates, in each dataset separately, and by then meta-analyzing across the two datasets. To adjust for testing multiple SNPs in *CACNA1C*, simpleM was used to estimate the effective number of independent tests (n=314) taking into account linkage disequilibrium among SNPs. The strongest evidence for interaction was found at rs34621387 (JMA *P*=7. 5x10-5, adjusted *P*=0. 02). The minor allele at rs34621387 is associated with increased risk of PD in vit D deficient individuals in both datasets (OR =2. 0, *P*=0. 002; OR=2. 1, *P*=0. 002). The SNP is not associated with PD risk in vit D non-deficient individuals (*P*>0. 8 in both datasets). Thus, risk of PD only increases in carriers of this allele who also have vit D deficiency. Our study demonstrates vit D-gene interaction in PD and suggests that GxE interactions might explain some of the "missing heritability". Importantly, our data suggest that people carrying the minor allele of rs34621387 might benefit more from vit D supplementation to lower PD risk.

1221T

Characterization of a missense variant in the fatty acid amide hydrolase gene (*FAAH*) in pharmacodynamic response to alcohol in nondependent social drinkers and alcohol-related traits in non-treatment-seeking drinkers and individuals seeking treatment for alcohol dependence. *J. Yan[1,2], B. L. Stangl[1], M. L. Schwandt[1], J. Westman[1], P. Spagnolo[1], C. Hodgkinson[3], D. Goldman[3], R. Momenan[1], V. A. Ramchandani[1].* 1) Laboratory of Clinical and Translational Studies, National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD; 2) Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA; 3) Laboratory of Neurogenetics, National Institute on Alcohol Abuse and Alcoholism, Rockville, MD.

The endocannabinoid system has been shown to modulate response to alcohol as well as display persistent adaptations to chronic use. A functional missense variant, rs324420 (C385A), in the gene encoding fatty acid amide hydrolase (*FAAH*) has been associated with amygdala habituation to threat, anxiety, fear extinction, stress reactivity and reward-related activity. The A allele of rs324420 confers decreased levels of the FAAH enzyme, which breaks down the endocannabinoid anandamide, and subsequent increased anandamide levels; thus, it is hypothesized to increase anandamide signaling in the brain and potentiate reward-related responses to alcohol. In an effort to investigate the pathways in which this variant plays in alcohol response, this study aimed to characterize its effect on intermediate alcohol-related phenotypes and its interaction with childhood adversity in a sample of individuals seeking treatment for alcohol dependence and non-treatment-seeking drinkers. The total sample consisted of 1170 individuals, a subset of whom participated in an intravenous alcohol self-administration (IV-ASA) study (n=82), wherein participants press a button to self-administer alcohol via IV-infusion, providing greater control over brain exposure to alcohol than oral intake. The association between rs324420 and its interaction with childhood trauma and alcohol traits was assessed using linear and logistic models with sex, age, and ancestry as covariates. Variation in rs324420 was assessed using a dominant model. Adversity was measured using the Childhood Trauma Questionnaire (CTQ), which assesses emotional, sexual, and physical abuse and emotional and physical neglect. Alcohol-related traits included alcohol consumption and problems, subjective response to alcohol, and peak and average breath alcohol concentration (BrAC) reached during the IV-ASA. The A allele was associated with increased average number of drinks consumed and number of heavy drinking days (p<0. 05). No main effects were found for alcohol self-administration in the IV-ASA sample. An interaction between rs324420 and CTQ total score was associated with subjective response following a priming dose of alcohol (p<0. 05). Individuals with the CC genotype and greater childhood trauma scores reported greater response to alcohol following priming. These results suggest differential response to alcohol and childhood trauma based on functional changes in anandamide signaling associated with this polymorphism.

1222F

Brain-derived Neurotrophic Factor genetic polymorphism Val66Met does not interact with stress in Colombian MDD individuals. *A. M. Gaviria-Manrique[1], M. Tovar-Parra[1], C. Cañizares-Escobar[2], Y. Gomez-Maquet[2], A. Arenas[3], M. C. Lattig-Matiz[1].* 1) Centro de Investigaciones Geneticas en Enfermedades humanas, Universidad de los Andes, Bogotá, Colombia; 2) Departamento de Psiquiatria, Universidad de los Andes, Bogotá, Colombia; 3) Clínica la Inmaculada, Bogotá, Colombia.

Major Depressive Disorder (MDD [MIM: 608516]) is a multifactorial disease that affects morbidity, mortality and quality of life. It has a reported heritability of approximately 35%, indicating an important genetic component. Several reports show a clear involvement of the Brain-Derived Neurotrophic Factor Gene (*BDNF*) Val66Met (rs6265) variant to MDD. Additionally, gene x environment interactions demonstrate that stressful life events and the Met variant play an important role in the development of MDD. However, most of the studies have been performed on Caucasian or Chinese populations while South American populations have not been fully studied in this regard. The interaction between early childhood abuse and Val66Met variant was related with negative affectivity in a Colombian non-clinical cohort by Perea et al. However, a meta-analysis conducted by Jiang's group (2013) including different ethnic groups found that the interaction is stronger between adulthood stress and Met genotype. Here we present a case-control study using a clinical Colombian population and match controls. A cognitive diagram of each individual was evaluated using the standard psychological tests M. I. N. I, ESV, STAI and IDER questionnaires after signing informed consent endorsed by the Ethics committee of the Universidad de los Andes; genotyping was performed using qPCR with taqman probes through Applied Biosystems 7500/7500 Fast Real time protocol, and data was analyzed using a multivariate logistic regression and interaction analysis in R software. The homozygous variant for the Val allele was the most common genotype observed in both cases and controls (59,5% and 61,8% respectively). A relevant association between abuse (physical or psychological) (OR=0. 15, p-value=0. 03) and sexual abuse (OR=0. 1, p-value=0. 007), and depression phenotype was found. Additionally, a strong correlation was noticed between the reported negative events number and MDD (OR=0. 04, p-value= 1. 39 e-5). Regarding genotype, no effect of the Val66Met variant and development of clinical depression was observed. When analysis for interaction was performed, no effect of gene x environment action was evidenced between studied variables. Even though it has been well established that *BDNF* plays an important role in the pathogenesis of MDD, this study concludes that we cannot generalize to all populations as our Colombian population may have a set of different genetic risk factors involved in the development of MDD.

1223W

The identification of novel genes in anxiety disorders: a gene x environment correlation and interaction study. *N. W. McGregor[1, 5], J. Dimatelis[2], S. M. J. Hemmings[1], K. Kinnear[3], D. Stein[4, 5], V. Russell[2], C. Lochner[1, 5].* 1) Psychiatry and Genetics, Stellenbosch University, Stellenbosch, Western Cape, South Africa; 2) Human Biology, University of Cape Town, Cape Town, Western Cape, South Africa; 3) Molecular Biology and Human Genetics, Stellenbosch University, Stellenbosch, Western Cape, South Africa; 4) Psychiatry and Mental Health, University of Cape Town, Cape Town, Western Cape, South Africa; 5) MRC, Anxiety and Stress Disorders Unit, Stellenbosch University, Tygerberg, Western Cape, South Africa.

There is clear evidence for a genetic component in anxiety disorders, and increasing focus has been placed on genetic and environmental interaction (GxE) in mediating disorder pathogenesis. Although a number of genetic studies have been conducted on anxiety disorders, no singular gene or genetic abnormality has been explicitly identified. The hypothesis is that a pre-existing genetic vulnerability (or genetic risk) interacts with the impact of adverse life events to result in the development of one or more anxiety disorder(s). Sprague Dawley rats exhibiting anxiety-like behaviours in the context of environmental stressors (maternal separation and restraint stress) were used as a model for the identification of novel susceptibility genes for anxiety disorders in humans. The striatum, previously implicated as a candidate in the brain architecture of anxiety pathogenicity, and the synaptic plasticity pathway were investigated in the rat brain using RT Profiler array assays. The human homologues of two susceptibility candidate genes (*MMP9* and *BDNF*) were screened in a human cohort of patients with obsessive-compulsive disorder (OCD), panic disorder (PD) or social anxiety disorder (SAD) (relative to controls) using probe-based genotyping arrays. Several genes were identified to be aberrantly expressed in "anxious" rats relative to controls (*Mmp9*, *Bdnf*, *Ntf4*, *Egr2*, *Egr4*, *Grm2* and *Arc*). Three single nucleotide polymorphisms (SNPs) were found to be significantly associated with these conditions (*MMP9*: rs3918242 and *BDNF*: rs6265 and rs10835210) in a case-control association fashion. Three SNPs were also found to significantly interact with the presence and severity of childhood trauma (*BDNF*: rs6265, rs10835210, rs11030107). This project yielded important findings pertaining to the aetiology of anxiety disorders. The use of a combined anxiety disorders cohort (OCD, PD and SAD) may suggest that the associations found here may hold true for anxiety disorders in general and not only for a particular clinically delineated condition. Several novel susceptibility genes, three significant SNP associations, and three significant SNP-environment interactions were contributed as candidates in the pathogenicity of anxiety disorders. Furthermore the severity of childhood trauma was confirmed as a risk factor for anxiety disorders.

1224T

The interplay between risky sexual behaviors and alcohol dependence: a genome-wide investigation. *R. Polimanti[1,2], H. Zhao[3,4], L. A. Farrer[5], H. R. Kranzler[6], J. Gelernter[1,2,4,7].* 1) Department of Psychiatry, Yale University School of Medicine, West Haven; 2) VA CT Healthcare Center, West Haven, CT; 3) Department of Biostatistics, Yale University School of Public Health, New Haven; 4) Department of Genetics, Yale University School of Medicine, New Haven, CT; 5) Departments of Medicine (Biomedical Genetics), Neurology, Ophthalmology, Biostatistics, and Epidemiology, Boston University Schools of Medicine and Public Health, Boston, MA; 6) Department of Psychiatry, University of Pennsylvania School of Medicine and VISN 4 MIRECC, Philadelphia VAMC, Philadelphia, PA; 7) Department of Neurobiology, Yale University School of Medicine, New Haven, CT.

Alcohol dependence (AD) is associated with risky sexual behaviors (RSB), but the basis for shared biological mechanisms leading to AD and RSB is poorly understood. We conducted AD-stratified genome-wide association studies (GWAS) and genome-wide gene-by-AD (GW-GxAD) analyses of RSB (modeled as a quantitative measure of unprotected sex and multiple sexual partners) in 3,924 African-American (AA) and European-American (EA) subjects (75% with AD, 55% AAs, and 32% women). We also used information about sexually transmitted disease (STD) status to evaluate the effect of RSB-AD loci on STD risk. Considering only consistent findings between the two genome-wide approaches, we identified five independent genome-wide significant (GWS) regions ($p<5x10-8$) and a suggestive GWS region ($p<6x10-8$). The AA-EA meta-analysis identified a GWS signal in *LHPP* (Phospholysine phosphohistidine inorganic pyrophosphate phosphatase; rs34997829, p=2. 51x10-8), a gene associated with major depression. In men, we observed a GWS signal in *FAM162A* (Family with sequence similarity 162, member A; rs2002594, p=4. 96x10-8), a suggestive locus for major depression. In women, there is a suggestive GWS locus in *PLGRKT* (Plasminogen receptor, C-terminal lysine transmembrane protein; rs3824435, p=5. 52x10-8), a previously identified GWS region for "Openness to Experience" in young women. In AAs, a GWS signal was observed within *GRK5* (G protein-coupled receptor kinase 5; rs1316543, p=1. 25x10-9), a gene involved in the regulation of G protein-coupled receptors. In AA men, we observed an intergenic GWS signal (rs12898370, p=4. 49x10-8) near *LINGO1* (Leucine rich repeat and Ig domain containing 1), a gene implicated in brain signaling. In EA men, a GWS signal was present in *CCSER1* (Coiled-coil serine-rich protein 1; rs62313897; p=7. 93x10-10), a gene with unknown function previously identified by multiple GWAS of brain phenotypes. The key loci identified offer molecular insights into the interplay between AD and RSB, highlighting mechanisms related to metal illness, personality features and other brain functions. We also investigated the effect of the loci identified in STD status, observing significant interaction between *LHPP* rs34997829 and AD (p=4. 97x10-3). To our knowledge, this is the first report of genome-wide investigation of RSB, and we provided the first genetic information about the interplay between AD and RSB, also indicating new insight about the relationship between AD and STDs.

1225F

Genetic vulnerability of the 5-HTTVNTR (STin2) variant of the SLC6A4 gene in major depressive disorder and childhood stressors in a Colombian population. *M. Tovar[1], C. Cañizares[1], Y. Gómez[1], M. C. Lattig[1], A. Arenas[2].* 1) biological science, Universidad de los Andes, Bogotá, Colombia; 2) Hermanas Hospitalarias: Clinica la Inmaculada.

Major depressive disorder- MDD (MIM: 608516) affects 25% of women and 17% of men around the world. It is a heterogeneous condition resulting from the interaction among genetic, environmental and social factors. Childhood stressors are defined as negative experiences occurred in childhood or adolescence, and are known to be risk factors for MDD. Serotonin is known to play an essential role in central nervous system development; thereby genetic variants related to serotonin uptake could cause alterations in maturation of neuron circuits involved in stress responses. The serotonin transporter protein (SLC6A4) is involved in serotonin reuptake at the presynaptic cell. A variant involving variable number of tandem repeats (VNTR) located in intron 2 (5-HTTVNTR (STin2) variant, Entrez: 6532) has two common alleles (10 and 12). The 5-HTTVNTR (STin2) variant has not been widely studied; however the genotype 10/10 was previously linked with depression and suicide. We performed a case-control study evaluating a clinical MDD population from the city of Bogotá-Colombia. All individuals signed informed consent approved by the university ethics committee. The aim of the study was to evaluate if stressors (childhood sexual, psychological and physical abuse) and 5-HTTVNTR(STin2) variant act as risk factors for MDD in our population. We also evaluated interaction among stressxSTin2 and the development of MDD, as this interaction has been reported for other genetic variants regarding the SLC6A4 gene. Psychological risk was assessed using the M. I. N. I and IDER standard tests; stress response was measured using the ESV test. STin2 VNTR variant was genotyped using standard PCR protocol. Statistical analysis included a multivariate logistic regression and interaction analysis using the statistical R software. Childhood sexual abuse was found as a risk factor for MDD (OR=19. 83 p=1. 0293e-03) as well as psychological and physical childhood abuse (OR=14. 63 p=2. 6031e-06). When analyzing the 5-HTTVNTR (STin2) variant and the development of MDD no significant association was observed. Interaction among childhood abuse and the Stin2 variant was not observed. Our results demonstrate that childhood stressors are high risk factors for the development of MDD as has been demonstrated in almost all populations studied. To our knowledge this is the first report of the STin2 variant in a clinical study in Colombian population and does not seem to have a crucial role in the development of MDD.

1226W

RNA-Seq uncovers specific genetic mechanisms in the hippocampus that are associated with behavioural and cognitive amelioration after environmental enrichment in a mouse model of prenatal alcohol exposure. *A. Chokroborty Hoque.* Biology, University of Western Ontario, London, Canada.

Maternal alcohol consumption during pregnancy gives rise to a spectrum of behavioural and cognitive deficits (Fetal Alcohol Spectrum Disorders (FASDs)) and their severity is significantly affected by the postnatal environment. We are using C57BL/6 mice to study the poorly understood interaction between prenatal alcohol exposure and the postnatal environment. Alcohol exposure during synaptogenesis produces high levels of anxiety-like traits and decreased memory performance in C57BL/6 mice. Ethanol- exposed mice (and matched controls) were put in 'environmentally enriched' conditions of voluntary exercise, physical activities and cognitive stimulation to ascertain the effects of a positive postnatal environment. While environmental enrichment ameliorates anxiety-like behaviour and memory performance of ethanol-exposed mice relative to ethanol-exposed non-enriched mice, the recovery is incomplete, indicative of the permanent damage of prenatal alcohol exposure on the developing brain. RNA was obtained from dissected mouse hippocampi for RNA-Seq (Illumina pair-end sequencing with V4 chemistry) and gene-expression microarray (GeneChip® Mouse Gene 2. 0 ST Array) analysis to determine the underlying genetic mechanisms associated with behavioural and cognitive recovery. DNA was obtained from dissected mouse hippocampi for global methylation analysis. For this conference, I will present (1) behavioural data (anxiety-like behaviour, recognition memory, visuo-spatial memory) (2) molecular data obtained from RNA-Seq analysis (Tuxedo Pipeline and edgeR to determine differential gene expression and GeneMania and WebGestalt to determine gene ontologies and significantly affected pathways) and global methylation analysis (3) elucidate upon the molecular biology of behaviour and cognition in a mouse model of prenatal alcohol exposure and (4) identify key genes and pathways associated with behavioural and cognitive recovery. Initial analysis implicates an upregulation of neuroactive ligand receptor interactions and cell-adhesion processes that may be targeted when formulating rehabilitative therapies to treat behavioural and cognitive deficits characteristic of FASDs. This work provides us with a detailed picture of gene regulation in the hippocampus associated with (1) alcohol exposure during neurodevelopment of mice and (2) cognitive and behavioural amelioration exhibited by affected mice in the presence of a positive, nurturing environment.

1227T

Large causal network analysis of transcriptional specialization in the brains of humans. *M. Yang, P. Wang, M. Xiong.* University of Texas School of Public Health, Houston, TX.

Develop statistical methods for construction of brain connectivity networks that connect brain regions is a key issue in neuroscience. In the past decade, a variety of methodological approaches have developed for the estimation of brain connectivity. However, most methods for construction of brain networks explore correlation or association between imaging measurements of the neural activity of the brain subregions. The correlation analysis is difficult to reveal the causal relationships among information flow in the brain regions. The statistical methods for causal networks of brain regions have less been developed. To overcome these limitation, we develop an analytic platform and novel statistical methods for inferring brain causal connectivity network that model information flow of molecular process. An essential issue for using causal graphs to study brain neuron activities is how to accurately and efficiently estimate the structure of causal graph. Early methods for structure learning mainly focused on approximation algorithms, but such methods are unable to ensure the generation of the true causal graph. To obtain the causal graph from observation data as close to the biological causal graph as possible, "score and search"-based methods for exact learning causal graphs of brain to find the best-scoring structures for a given dataset are developed. We use Allen Human Brain Atlas dataset to evaluate the methods for the construction of the brain causal networks and whole genome- gene expression causal networks where the whole genome-wide gene expressions of each of 231 brain subregions for six samples were profiled by microarray. A causal brain network with 98 nodes and 138 edges from gene expression data is inferred. The whole-genome gene expression causal network consists of two steps: In the first step, we constructed 100 clusters of the gene expressions. For each cluster, we calculate the principal components. These principal components were used as summary statistics of the cluster and were used to construct causal networks with 76 nodes and 101 edges. In the second step, for each cluster we infer gene expression causal networks. For example, for the 64th cluster, we obtained a causal gene expression network with 62 nodes and 72 directed edges. The whole genome gene expression causal networks in this were inferred. To our knowledge, this is the first attempt to infer whole genome gene expression causal networks.

1228F

Integrative Causal Network Analysis of Imaging and Genetic Data in Schizophrenia Studies. *N. Lin[1], P. Wang[4], Y. Zhu[5], L. Luo[2], V. Calhoun[3], M. Xiong[1].* 1) University of Texas School of Public Health, Houston, TX; 2) University of New Mexico; 3) Mind Research Network; 4) Fudan University; 5) Tulane University.

Rapid advances in medical imaging and genomic technologies dramatically change the research in genetic studies of schizophrenia and other psychiatric disorders. Integrative analysis of genetic and imaging will facilitate discovery of the mechanism of the complex diseases. The popular methods for integrative image-genetic analysis are mainly based on correlation and association analysis. These methods cannot efficiently detect, distinguish and characterize the true biological, mediated and spurious pleiotropic effects. Therefore, these approaches may not provide clear biologically or clinical relevant information that allows the mechanisms of genetic effects to be discovered and understood. To overcome these limitations, we develop a new framework for inferring causal networks of genotype-images and detecting, distinguishing and characterizing the true biological pleiotropic, mediated pleiotropic and spurious pleiotropic effects of genetic variants on imaging measurements as endophenotype. We also develop a new sufficient dimension reduction methods for network classification. The proposed method for causal network analysis was applied to the MIND clinical imaging consortium's schizophrenia image-genetic study with 142 series of diffusion tensor MRI images (DTI) and 14,412 genes typed in 64 schizophrenia patients and 78 healthy controls. DTI series were segmented into 41 regions by super pixel method and the mean intensity of each region were taken as the image measurement of the brain region. The causal image-genotype networks were separately constructed for schizophrenia patient population and normal population. In cases, the brain image network consisted of 41 nodes and 54 directed edges, and in controls, the brain image network consisted of 41 nodes and 57 directed edges. There were 29 common directed edges in both the case and control brain image networks. The causa image-genotype networks were also constructed. We identified 840 genes and 238 genes that were significantly connected to the brain image region at the genome-wide significance level, respectively, in cases and controls. The developed network classification algorithms were also applied to the dataset to predict schizophrenia. Using 5 fold cross validation for evaluation, we can achieve 100% prediction accuracy in all the five training data set and the average prediction accuracy, sensitivity and specificity over 5 test data sets were 95. 1%, 96. 2% and 93. 9% respectively.

1229W

The use of genetic risk factors to assess prodromal brain changes in Alzheimer's disease. *M. K. Lupton[1], L. Strike[1,2,3], W. Wen[4], K. A. Mather[4], N. J. Armstrong[5], A. Thalamuthu[4], K. L. McMahon[2], G. I. de Zubicaray[3], A. A. Assareh[4], A. Simmons[6], P. Proitsi[6], J. F. Powell[6], G. W. Montgomery[1], D. P. Hibar[7], E. Westman[8], M. Tsolaki[9], I. Kloszewska[10], H. Soininen[11], P. Mecocci[12], B. Velas[13], S. Lovestone[6,14], H. Brodaty[4], D. Ames[15], J. N. Trollor[4], N. G. Martin[1], P. M. Thompson[7], P. S. Sachdev[4], M. J. Wright[1], the Alzheimer's Disease Neuroimaging Initiative.* 1) QIMR Berghofer Medical Research Institute, Brisbane, Australia; 2) Centre for Advanced Imaging, University of Queensland, Brisbane, Australia; 3) School of Psychology, University of Queensland, Brisbane, Australia; 4) Centre for Healthy Brain Ageing, University of New South Wales, Sydney, Australia; 5) Mathematics and Statistics, Murdoch University, Perth, Australia; 6) King's College London, Institute of Psychiatry, Psychology and Neuroscience, London, United Kingdom; 7) Imaging Genetics Center, Keck School of Medicine, University of Southern California, Marina del Ray, California, USA; 8) Department of Neurobiology, Care Sciences and Society, Division of Clinical Geriatrics, Alzheimer's Disease Research Centre, Karolinska Institute, Stockholm, Sweden; 9) Memory and Dementia Centre, Aristotle University of Thessaloniki, Thessaloniki, Greece; 10) Department of Old Age Psychiatry & Psychotic Disorders, Medical University of Lodz, Lodz, Poland; 11) Department of Neurology, Kuopio University and University Hospital, Kuopio, Finland; 12) Section of Gerontology and Geriatrics, Department of Clinical and Experimental Medicine, University of Perugia, Perugia, Italy; 13) Department of Internal and Geriatrics Medicine, Gerontopole, Hôpitaux de Toulouse, Toulouse, France; 14) University of Oxford, Department of Psychiatry, Warneford Hospital, Oxford, United Kingdom; 15) National Ageing Research Institute, Parkville and University of Melbourne Academic Unit for Psychiatry of Old Age, Kew, Australia.

Alzheimer's disease (AD) is the most common form of dementia with a global prevalence of 24 million, predicted to quadruple by 2050. The major genetic risk factor is the Apolipoprotein E (*APOE*) locus, with large-scale GWAS meta-analyses having identified a further 19 common risk SNPs. A low frequency missense variant in *TREM2* also substantially increases AD risk. A reduction in subcortical brain volumes, observed using structural magnetic resonance images (MRI), are specific indicators of AD pathology. This is especially apparent in the hippocampus and amygdala, detectable in prodromal and early stage AD. We examined whether increased genetic risk for AD, based on *APOE* genotype, *TREM2* genotype and polygenic risk scores (PRS) combining common risk variants influences hippocampal and amygdala volumes. This was tested in AD case-control (ADNI and AddNeuroMed) and population based cohorts including elderly (the Older Australian Twin Study, the Sydney Memory and Aging Study) and young people (the Queensland Twin Imaging cohort) with a total sample size of 2141. Incidence of AD and Mild Cognitive Impairment (MCI) were taken into account in the elderly group. The PRS (excluding *APOE*) and a SNP in *TREM2* both associate with reduced hippocampal volume in the elderly, with a significant association still apparent in the healthy older group without AD or MCI. *APOE* ϵ4 associates with hippocampal and amygdala volume in those with AD and MCI, but not in the healthy older group. No associations were found in young adults (16-30 years N=467). We have shown that genetic risk for AD affects the hippocampus before the clinical symptoms of AD, reflecting a neurodegenerative effect prior to clinical manifestations in older adults. This correlation with early MRI markers of AD shows evidence for a genetic modulation of neurodegeneration and the potential for a combination of PRS and brain biomarkers to aid in the prediction of future cognitive decline and the development of AD. Future work is investigating the effect of genetic risk variants on hippocampal microstructure and subcortical thinning, which are potentially more sensitive to early changes.

1230T

Investigating the non-linear association of anxiety-depression score with right lingual surface area. *B. Couvy-Duchesne[1,2,3], L. Strike[1,2,3], P. M. Thompson[5], K. L. McMahon[3], G. I. de Zubicaray[2], N. G. Martin[6], I. B. Hickie[4], M. J. Wright[1].* 1) NeuroImaging Genetics, QIMR Berghofer Medical Research Institute, Herston, QLD, Australia; 2) School of Psychology, University of Queensland, Brisbane, QLD, Australia; 3) Centre for Advanced Imaging, University of Queensland, Brisbane, QLD, Australia; 4) Brain & Mind Research Institute, University of Sydney, Sydney, NSW, Australia; 5) Imaging Genetics Center, Keck School of Medicine, University of Southern California, Marina del Rey, CA, USA; 6) Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Herston, QLD, Australia.

There is little consensus in the literature about brain changes associated with depression and anxiety. Structural markers of depression-anxiety are of importance as they could provide insight into the aetiology of depression and anxiety, as well as quantitative assessment of the depressed/anxious brain. In a population sample of young adults comprising 865 twins and siblings (63% female) we tested the association of 170 standard structural brain phenotypes (subcortical volumes, cortical surface area and thickness) with anxiety-depression from the Somatic and Psychological HEalth REport (SPHERE) (Hansell, et al. , 2012; Hickie, et al. , 2001). The only association to survive multiple testing correction was between anxiety-depression and surface area of the lingual gyrus (cubic relationship, p-value=4. 9E-5). Higher anxiety-depression score was associated with a reduction (up to~15%) in surface area, with a genetic correlation of -0. 31 [-0. 51, -0. 13], p-value=7. 0E-4. The SPHERE provides a self-reported measure of anxiety and depression computed from 18 items. Boundaries of the lingual gyrus were defined using parcellation available in Freesurfer 5. 3. Models were corrected for age at MRI, sex, time difference between scoring and scanning, mean surface area or cortical thickness. The association was observed despite the anxiety depression score being acquired on average 4 years prior to scanning, (mean age at scan was 22, sd=3. 1). In a nested case-control sample, a 3. 3% reduction of right lingual surface area was observed in recurrent MDD cases (DSM-IV definition, p-value=0. 015). We sought replication in an independent young adult clinical sample and investigated the robustness of the result using a genetically derived cortical atlas and a voxel based approach. We aimed to confirm the association of right lingual gyrus surface area with our continuous measure of anxiety-depression.

1231F

A practical grading definition for the diagnosis of autosomal recessive spastic ataxia of Charlevoix-Saguenay. *I. Coupry[1,8], J. Pilliod[1,8], S. Moutton[1,2], J. Lavie[1], E. Maurat[1], C. Hubert[1,3], N. Bellance[1], N. Houcinat[2], J. Van-Gils[2], C. M. Durand[1], C. Rooryck[1,2], D. Lacombe[1,2], R. Rossignol[1], G. Stevanin[4,5,6,7], G. Benard[1], A. Durr[4,5,6], C. Goizet[1,2].* 1) Université de Bordeaux, Laboratoire Maladies Rares: Génétique et Métabolisme, EA4576, Bordeaux, France; 2) CHU Pellegrin, Service de Génétique Médicale, Bordeaux, France; 3) Université de Bordeaux, Centre de Génomique Fonctionnelle, Bordeaux, France; 4) APHP, Hôpital de la Pitié-Salpêtrière, Service de Génétique, Paris, France; 5) Institut du Cerveau et de la Moelle, INSERM U1127, Paris, France; 6) Sorbonne Universités, UPMC Université Paris VI, Paris, France; 7) Ecole Pratique des Hautes Etudes, Laboratoire de NeuroGénétique, Paris, France; 8) These authors contributed equally to this work.

Objective: Autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS, [MIM 270550]) is caused by mutations in the *SACS* gene. *SACS* encodes sacsin, a protein whose function remains unknown, despite the description of numerous protein domains and the recent focus on its potential role in the regulation of mitochondrial physiology. This study aimed to identify new mutations in a large population of ataxic patients and to functionally analyze their cellular effects in the mitochondrial compartment. **Methods:** 321 index patients with spastic ataxia, selected from the SPATAX network, were analysed by direct sequencing of the *SACS* gene, and 156 patients from the ATAXIC project presenting with congenital ataxia were investigated either by targeted or whole-exome sequencing. For functional analyses, primary cultures of fibroblasts were obtained from 11 patients, carrying either mono- or bi-allelic variants, including one case harboring a large deletion encompassing the entire *SACS* gene**Results:** We identified biallelic *SACS* variants in 33 patients from SPATAX, and in 5 non-progressive ataxia patients from ATAXIC. Moreover, a drastic and recurrent alteration of the mitochondrial network was observed in 10 of the 11 patients tested. **Interpretation:** Our results permitted to extend the description of clinical and mutational spectrum of ARSACS patients. Moreover, we suggest that the observed mitochondrial network anomalies could be used as a trait biomarker for the diagnosis of ARSACS when *SACS* molecular results are difficult to interpret (i. e. , missense variants and heterozygous truncating variant). Based on our findings, we propose a grading diagnostic definition using clinical, genetic, and cellular criteria for ARSACS.

1232W

Animal model module of AutDB catalogues the effects of rescue agents on ASD phenotypes. *I. Das, M. A. Estevez, S. Banerjee-Basu.* Mindspec, Inc. , McLean, VA.

The Autism Database (AutDB) is a publicly available, manually annotated, modular database that serves as an ongoing collection of genes linked to Autism Spectrum Disorders (ASD). Here, we describe further development of the animal model module of AutDBwhich catalogues ASD-relatedrodent models. A new release of the database also includes rat models in addition to mouse models, since both of these species are widely used in current research. In addition to ASD models, the animal model module also contains rescue models based on existing parent ASD models. All data is extracted from published, peer-reviewed primary reports. The metadata is standardized in a phenotypic database (PhenoBase), which is a routinely updated comprehensive list of phenotypic terms (pheno-terms) and experimental paradigms. These pheno-terms reflect the actual research and are divided into categories that align with human ASD phenotypic features. The recently updated PhenoBase has been standardized to about 380 pheno-terms divided into 16 categories. Given the complex manifestation of ASD phenotypes, described as core and auxiliary in the human population, a wide variety of drugs and genetic manipulations are being used in the mouse models to assess their ameliorative effects on some or all of these phenotypes. We categorize these 'rescue' paradigms based on the type of agent: biological, behavioral, genetic or pharmaceutical, and have developed annotation methods to clearly represent the treatment effects on ASD phenotypes paralleled in mouse models. The scientific standardization of phenotype, their presentation in animal models and the effects of rescue, allow for data mining and bioinformatics analysis. Coupled to our large curated dataset this can be used to elucidate ASD research trends and etiology.

1233T

Phenotype-genotype associations in concordant and discordant monozygotic and dizygotic twins based on quantitative trait and case-control association analyses. *V. Hu, C. Devlin.* Biochemistry & Molecular Medicine, The George Washington University School of Medicine & Health Sciences, Washington, DC.

Autism is a highly heterogeneous and heritable neurodevelopmental disorder characterized by impaired social communication and interactions, often coupled with stereotyped behaviors and restricted interests. Although there has been much discussion regarding the magnitude of the genetic contribution to autism based on concordance rates of the diagnosed disorder in monozygotic vs. dizygotic twins, there is little information on genetic contributions to specific behaviors on which the diagnosis is based. Here, we compared the phenotypic profiles of both concordant and discordant mono- and dizygotic twins across five behavioral traits, using cumulative trait severity scores from the Autism Diagnostic Interview-Revised (ADI-R) instrument for each individual, as previously described [1]. These traits included spoken language, nonverbal communication, play skills, social behaviors, and insistence on sameness or ritualistic behaviors. The five trait profiles across the different twin populations were then used for quantitative trait association analyses using genotype data for the respective individuals. Trait-associated single nucleotide polymorphisms with a nominal p-value of 10-5 were then used for case-control association analyses. Cases included individuals in the lowest (first) and uppermost (third) quartiles of the quantitative trait distribution curves, and controls consisted of 2438 previously genotyped nonautistic individuals [2]. Results: The deep phenotypic analyses revealed that there are significant differences in trait profiles between the discordant nonautistic monozygotic and discordant nonautistic dizygotic twins, with the dizygotic twins exhibiting significantly lower severity across virtually all behaviors analyzed. By comparison, all of the twins diagnosed with autism exhibit similar behavioral severity profiles across all ADI-R items, regardless of zygosity. The combination of quantitative trait and case-control association analyses revealed Bonferroni significant genetic variants that clearly distinguish individuals in the uppermost quartile of trait severity from those in the lowest quartile. In summary, while the comparison of discordant nonautistic dizygotic vs. monozygotic twins suggests that genotype is a significant determinant of autistic traits, the similarity of trait profiles in the concordant dizygotic twins also suggests a role for environment. 1. Hu & Steinberg, Autism Res. 2:67-77, 20092. Wang *et al*. Nature 459:528-533, 2009.

1234F

Genetic involvement in progression of MS disease course. *M. F. Davis[1], B. Peaden[1], S. Sriram[2], J. L. Haines[3], J. C. Denny[4].* 1) Department of Microbiology and Molecular Biology, Brigham Young University, Provo, UT; 2) Department of Neurology, Vanderbilt University Medical Center, Nashville, TN; 3) Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, OH; 4) Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN.

Multiple sclerosis (MS) is a neurodegenerative autoimmune disease of the central nervous system, with over 100 known genetic loci associated with risk of development of the disease. However, disease course is incredibly varied among individuals with MS and it is unknown if disease course is also associated with genetic variation. The timed 25-foot walk (T25FW) is a quantitative clinical measure of disability progression that MS patients at the Vanderbilt University Multiple Sclerosis Clinic complete each visit. We used a validated natural language processing approach to extract 4,792 T25FW measurements over 17 years of clinic visit for 717 patients, with an average walk time of 9. 7 seconds. Patients were genotyped on the Illumina ImmunoChip and are a part of Vanderbilt's BioVU DNA database. We analyzed the data using mixed model linear regression in R, with the null model consisting of amount of time taken to walk 25 feet as a response variable, the amount of time since the first recorded T25FW and three principle components for ethnicity as fixed effects, and the amount of time since last walk and the patient as random effects. This model was tested against models that added a SNP as a fixed effect. The p-value was calculated based on the difference between the two models. Using a Bonferroni corrected p-value of 5. 7 x 10-7, three SNPs in *ABHD6* (p=8. 4 x 10-8), *ILKAP* (p=5. 7 x 10-9), and *FGF20* (p-1. 9 x 10-8) were significant. No known MS risk alleles achieved Bonferroni significance; the strongest association with known MS risk alleles was rs8070345 with p=0. 02. Of special interest is the region around *ABHD6* on chromosome 3, which has been associated with risk of development of systemic lupus erythematosus, another autoimmune disease.

1235W

Novel mutations in SPTAN1: expanding the neurological phenotype.
V. Gartner[1,2], T. Markello[2], C. J. Tifft[1,2], W. A. Gahl[1,2], P. R. Lee[1,2]. 1) Office of the Clinical Director, NHGRI/NIH, Bethesda, MD; 2) NIH Undiagnosed Diseases Program, NIH Office of Rare Diseases Research and NHGRI, Bethesda, MD.

De novo mutations in Spectrin Alpha Non-Erythrocytic 1 (SPTAN1) have previously been linked to Early Infantile Epileptic Encephalopathy 5 (EIEE5), a rare and severe neurological disorder. Defining features of EIEE5 include refractory epilepsy, hypoplasia and/or atrophy of the cerebellum and brain stem, and cerebral hypomyelination. Patients with EIEE5 additionally have hypsarrhythmia on EEG, acquired microcephaly, intellectual disability, and spastic quadriplegia. Previous reports indicate that mutations in SPTAN1 that disrupt the C-terminus domains of the alpha-II spectrin protein have a dominant negative effect that may correlate with phenotypic severity. Only one case of a *de novo* mutation in SPTAN1 causing neurological symptoms without seizures has previously been reported. The aim of our study was to provide careful phenotypic descriptions of two additional cases with observed *de novo* mutations in SPTAN1 to further expand the spectrum of this rare disease. In both cases we describe novel *de novo* mutations that were identified by whole-exome sequencing completed as part of comprehensive examinations of the patients. One patient is heterozygous for a missense mutation leading to a premature stop codon in exon 21. This patient has cerebellar and brain stem hypoplasia, cataracts, intellectual disability, and autism spectrum disorder. However, this patient does not have seizures and therefore represents only the second known case with a SPTAN1 mutation to not have a diagnosis of epilepsy. The second patient we describe is heterozygous for an intronic missense mutation leading to the insertion of 15 nucleotides before exon 50, and 5 residues added in the C-terminus of the protein. This patient suffers from severe ataxia and cerebellar and brain stem hypoplasia, and has ADHD, intellectual disability, and autism spectrum disorder. While this patient has a more severe phenotype than our first case, this patient would be the third described without seizures. The mutation loci and relative phenotypic severity of these two cases expand on prior descriptions of the disorder. Additionally, we conclude that mutations in SPTAN1 do not invariably lead to seizures or severe epileptic encephalopathy. Notably, we provide a unique example in which a mutation in the C-terminus region of the protein is correlated with greater phenotypic severity without seizures, whereas prior C-terminus mutations all had a significant epileptic component to their SPTAN1-related disease.

1236T

The Neurodevelopmental Spectrum Associated with SHANK-3 Mutations.
P. Lawson[1], O. Goker-Alpan[2], S. Lipinski[2], C. Samango-Sprouse[1,3]. 1) The Focus Foundation, Crofton, MD; 2) O and O Alpan LLC, Genetics Division, Fairfax, VA; 3) The Neurodevelopmental Diagnostic Center for Young Children, Davidsonville, MD.

Background: The *SHANK3* gene encodes a scaffolding protein in the cortex and cerebellum that has a role in synaptic signal transmission. *SHANK3* is disrupted in Phelan-McDermid Syndrome (PMS, or 22q13.3 deletion syndrome), which is characterized by variable central nervous dysfunction including autism spectrum disorder (ASD), speech, motor, and intellectual deficits with hypotonia and ataxic gait. Typically, receptive language is less impaired than expressive communication. Dysmorphic features include periorbital fullness, long eyelashes, dysplastic toenails, bushy eyebrows, and a bulbous nose. Larger deletions correlate with greater developmental dysfunction. **Case Description and Results:** We present a 4 ½ year-old girl with a history of regression, limb asymmetry, advanced bone age, and coordination difficulties. Early development was normal, however feeding disturbances and speech delay with intermittent regression of motor and speech-language skills were noted at 18 months. She had truncal hypotonia with motor difficulties, balance problems, and oculomotor apraxia (OMA), but had normal gait. She did not have ASD with GARS-2 and GADS scores below the first percentile, however, she had atypical peer social interactions with Childhood Apraxia of Speech (CAS) and misarticulations, and visual perceptual deficiencies (SS=74 on VMI). The SRS-2 revealed mildly atypical scores in social cognition and social awareness, with the four remaining scores well within normal limits. Intellectual function was normal with a Fluid Reasoning (FR) of 93, Working Memory Index (WMI) of 103, Verbal Comprehension Index (VCI) of 81, and FSIQ of 83 on the Wechsler Intelligence Tests. The expressive and receptive vocabulary scores were 97 and 103 on EOWPVT-R and ROWPVT-R, respectively. Whole Exome Sequencing identified a de novo c. 2265+3 A>T variant in intron 19 in the *SHANK3* gene. **Discussion:**This report highlights the first case where a single-point mutation in the *SHANK3* gene has resulted in a constellation of findings associated with PMS, supporting that the *SHANK3* gene is critical in PMS. While *SHANK3* is implicated in ASD, this patient shows no autistic features. This case is also unique, in that FSIQ, FR, WMI and VCI are within normal limits. This presentation thus supports higher intellectual capability than has been previously reported in patients with *SHANK3* mutations.

1237F

Phenotypic spectrum of *GNAO1* variants: epileptic encephalopathy to involuntary movements with severe developmental delay. *H. Saitsu[1], R. Fukai[1,2], B. Ben-Zeev[3,4], Y. Sakai[5], M. Mimaki[6], N. Okamoto[7], Y. Suzuki[8], Y. Monden[9], H. Saito[9], B. Tziperman[3], M. Torio[5], S. Akamine[5], N. Takahashi[6], H. Osaka[9], T. Yamagata[9], K. Nakamura[10], Y. Tsurusaki[10], M. Nakashima[10], N. Miyake[10], M. Shiina[11], K. Ogata[11], N. Matsumoto[1].* 1) Human Genetics, Yokohama City University, Yokohama, Japan; 2) Neurology and Stroke Medicine, Yokohama City University, Yokohama, Japan; 3) The Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Ramat Gan, Israel; 4) Sackler school of medicine, Tel aviv University, Tel aviv, Israel; 5) Pediatrics, Kyushu University, Fukuoka, Japan; 6) Pediatrics, University of Tokyo, Tokyo, Japan; 7) Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 8) Pediatric Neurology, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 9) Pediatrics, Jichi Medical University, Tochigi, Japan; 10) Pediatrics, Yamagata University Faculty, Yamagata, Japan; 11) Biochemistry, Yokohama City University, Yokohama, Japan.

*De novoGNAO1*variants have been found in four patients including three patients with Ohtahara syndrome and one patient with childhood epilepsy. In addition, two patients showed involuntary movements, suggesting that*GNAO1*variants can cause various neurological phenotypes. Here we report additional four patients with*de novo*missense*GNAO1*variants, one of which was identical to that of previously reported. All the three novel variants were predicted to impair Gαofunction by structural evaluation. Two patients showed early-onset epileptic encephalopathy, presenting with migrating or multifocal partial seizures in their clinical course, but remaining two patients showed no or few seizures. Together with four previously reported patients, all the eight patients showed severe intellectual disability, and motor developmental delay. In addition, involuntary movements were observed in six of eight patients, indicating that involuntary movements may be one of key features suspecting *GNAO1* variants. Progressive cerebral atrophy and thin corpus callosum were common features in brain images. Our study demonstrated that*GNAO1*variants can cause involuntary movements and severe developmental delay without/with seizures including various types of early-onset epileptic encephalopathy.

1238W

VarDB: a catalog of genetic variability in neurodegenerative disease. *D. M. Evans, M. J. Farrer.* Centre for Applied Neurogenetics, CBH, University of British Columbia, Vancouver, British Columbia, Canada.

There are several databases of whole exome or whole genome data available online based off of hundreds or thousands of individuals. However, there are very few public resources that capture genetic variability in conjunction with any available phenotype information. The Djavad Mowafaghian Centre for Brain Health has developed a publicly available database of short nucleotide variants from more than 1000 individuals with associated biographic and clinical information. Variants can be returned by searching on any subset of subject data including neurodegenerative disorder affection status, call quality, ethnic background and sequencing platform. Results include single nucleotide and short indel variant annotation with externally generated population frequencies as well as variant identifiers and resulting cDNA and amino acid changes. A subset of samples have also undergone high-density SNP genotyping, allowing for the interrogation of copy number variation data by locus. The database is available at http://vardb. can. ubc. ca.

1239T

Genome-wide Association Study of Clinical Features in the SchizophreniaPsychiatric Genomics Consortium: Confirmation of Polygenic Effect on Negative Symptoms. *A. Fanous[1,2,3], T. B. Bigdeli[3], S. Ripke[4], R. L. Amdur[1,2], P. A. Holmans[5], A. Corvin[6], P. G. C. SCZ working group[1-7], K. S. Kendler[3,7].* 1) Washington VA Medical Center, Washington, DC; 2) Department of Psychiatry, Georgetown University School of Medicine; 3) Department of Psychiatry, Virginia Commonwealth University School of Medicine; 4) Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts; 5) Department of Psychological Medicine and Neurology, School of Medicine, Cardiff University, Cardiff, Wales, UK; 6) Neuropsychiatric Genetics Research Group, Trinity College Dublin, Dublin, Ireland; 7) Department of Human Genetics, Virginia Commonwealth University School of Medicine.

Objective: Schizophrenia is a clinically heterogeneous disorder. Proposed revisions in DSM-5 included dimensional measurement of different symptom domains. We sought to find common genetic variants affecting these dimensions, and confirm our previous finding that polygenic risk of schizophrenia influences Negative symptom severity. **Methods:** The Schizophrenia Psychiatric Genomics Consortium comprised 8,388 cases of European ancestry with available clinical phenotype data. Symptoms averaged over the course of illness were assessed using the OPCRIT, PANSS, LDPS, SCAN, SCID, and CASH. Factor analyses were performed to identify the Positive, Negative, Manic, and Depressive symptom dimensions. Genome-wide association studies (GWAS) of the resulting quantitative traits were performed using linear regression adjusting for ancestry, and meta-analyzed across study sites. We considered also whether each symptom factor was significantly associated with polygenic scores indexing schizophrenia risk. **Results:**We observed the strongest evidence of association for the Positive symptom factor, at rs74263251 (*P*=6. 27x10-8) in an intron of*RFX8*(2q11. 2). Polygenic scores for schizophrenia risk were significantly correlated with the Negative factor (*P*=6. 39x10-9), explaining approximately 0. 4% of the variance. These scores were not strongly associated with Positive or Manic symptoms, and were found to significantly predict Depressive symptoms only among female patients (*P*=0. 00922). **Conclusions:**While we found no definitive evidence of modifier loci for schizophrenia, associations between SNPs in*RFX8*and Positive symptoms approached genome-wide significance (*P*=5x10-8). As this may have been due to insufficient statistical power, follow-up in additional samples is warranted. Critically, we confirmed our previous finding that polygenic risk of schizophrenia explains at least some of the variance in Negative symptoms, a core illness dimension.

1240F

***C9ORF72* repeat expansions that cause frontotemporal dementia are detectable among patients with psychosis without dementia.** *V. Nimgaonkar[1], G. Coppola[2], A. Watson[3], M. Pribadi[2], K. Chowdari[3], S. Clifton[3], J. Wood[3], B. Miller[4].* 1) University of Pittsburgh, Pittsburgh, PA; 2) University of California, Los Angeles, CA; 3) Department of Psychiatry, University of Pittsburgh, School of Medicine, Pittsburgh, PA; 4) University of California, San Francisco, CA.

Background: A pathologic hexanucleotide repeat expansion in the regulatory region of *C9ORF72* causes frontotemporal dementia (FTD) or amyotrophic lateral sclerosis (ALS). In addition to features of dementia, behavioral abnormalities can be present among mutation carriers. It is uncertain whether a proportion of individuals clinically diagnosed with psychoses bear pathologic *C9ORF72* expansions in the absence of dementia. Aim: To screen for *C9ORF72* repeat expansions among individuals diagnosed with functional psychoses such as schizophrenia (SZ) / schizoaffective disorder (SZA), and to describe the clinical features of individuals with the mutation. Results: Pathogenic *C9ORF72* repeat expansions were detected in two pairs of related individuals following a survey of 740 participants in a psychiatric genetic research study. The mutation carriers included two siblings with schizophrenia, another unrelated individual with schizoaffective disorder and her non-psychotic mother. All the mutation-bearing patients with SZ/SZA had severe, florid illness, but did not provide a history suggestive of dementia or ALS. Conclusions: A small proportion of patients with SZ/SZA bear *C9ORF72* repeat expansions. The patients show atypical psychotic features without the typical cognitive features of dementia.

1241W

An open-label trial in Friedreich ataxia suggests clinical benefit with high-dose resveratrol, without effect on frataxin levels. *M. Delatycki[1, 2, 3], E. Yiu[1, 3, 4], R. Peverill[5], K. Lee[3, 6], K. Croft[7], T. Mori[7], B. Scheiber-Mojdehkar[8], B. Sturm[8], M. Praschberger[8], A. Vogel[1, 9], G. Rance[9], S. Stephenson[1, 3], J. Sarsero[10], C. Stockley[11], C. Lee[12], A. Churchyard[13], M. Evans-Galea[1, 3], M. Ryan[3, 4, 14], P. Lockhart[1, 3], L. Corben[1, 3].* 1) Bruce Lefroy Ctr, MCRI, Royal Children's Hosp, Parkville, Australia; 2) Department of Clinical Genetics, Austin Health, Heidelberg, Victoria, Australia; 3) Department of Paediatrics, The University of Melbourne, Parkville, Victoria, Australia; 4) Department of Neurology, Royal Children's Hospital Melbourne, Parkville, Victoria, Australia; 5) Monash Cardiovascular Research Centre, Monash HEART and Monash University Department of Medicine, Monash Medical Centre, Clayton, Australia; 6) Clinical Epidemiology and Biostatistics Unit, Murdoch Childrens Research Institute, Parkville, Victoria, Australia; 7) School of Medicine and Pharmacology, University of Western Australia, Perth, Western Australia, Australia; 8) Department of Medical Chemistry, Medical University of Vienna, Vienna, Austria; 9) Department of Audiology and Speech Pathology, The University of Melbourne, Parkville, Victoria, Australia; 10) Cell and Gene Therapy, Murdoch Childrens Research Institute, Parkville, Victoria, Australia; 11) Australian Wine Research Institute, Adelaide, South Australia, Australia; 12) School of Biological Sciences, The University of Hong Kong, Hong Kong, China; 13) Department of Neurology, Monash Health, Clayton, Victoria, Australia; 14) Neurosciences Research, Murdoch Childrens Research Institute, Parkville, Victoria, Australia.

Friedreich ataxia (FRDA) is due to a triplet repeat expansion in *FXN*, resulting in deficiency of the mitochondrial protein frataxin. Resveratrol is a naturally occurring polyphenol, identified to increase frataxin expression in cellular and mouse models of FRDA, and has anti-oxidant properties. This open-label, non-randomized trial evaluated the effect of two different doses of resveratrol on peripheral blood mononuclear cell (PBMC) frataxin levels over a 12-week period in individuals with FRDA. Secondary outcome measures included PMBC *FXN* mRNA, oxidative stress markers, and clinical measures of disease severity. Safety and tolerability were studied. Twenty-four participants completed the study; 12 received low-dose resveratrol (1 g daily) and 12 high-dose resveratrol (5 g daily). PBMC frataxin levels did not change in either dosage group [low dose group change: 0. 08 pg/μg protein (95% CI -0. 05, 0. 21, p=0. 21); high dose group change: 0. 03 pg/μg protein (95% CI -0. 10, 0. 15, p=0. 62)]. Improvement in neurologic function was evident in the high-dose group [change in Friedreich Ataxia Rating Scale -3. 4 points, 95% CI (-6. 6, -0. 3), p=0. 036] but not the low-dose group. Significant improvements in audiologic and speech measures, and in the oxidative stress marker plasma F2-isoprostane were demonstrated in the high-dose group only. There were no improvements in cardiac measures or patient-reported outcome measures. No serious adverse events were recorded. Gastrointestinal side-effects were a common, dose-related adverse event. This open-label study shows no effect of resveratrol on frataxin levels in FRDA, but suggests that independent positive clinical and biologic effects of high-dose resveratrol may exist. Further assessment of efficacy is warranted in a randomized placebo-controlled trial.

1242T

Cell-free microRNA (miRNA) deregulation in the plasma of Friedreich's Ataxia patients. *D. Subrahamanyam[1], A. Srivastava[2], MR. Rajeswari[1].* 1) Biochemistry, All India Institute of Medical Sciences, New Delhi, India; 2) Neurology, All India Institute of medical Sciences, New Delhi, India.

Friedreich ataxia (FRDA) is a multisystem disease affecting predominately nervous system, heart and pancreas. Autosomal recessive inheritance of trinucleotide repeat (GAA) expansion in the *FXN* gene causes the abnormal phenotype in FRDA patients. Current research focused on therapeutic intervention using drugs aimed at molecular amelioration, but there are no reliable biomarkers available to monitor the efficacy of the drug. The present study investigates the alterations of plasma cell-free miRNAs (micro RNAs) in FRDA patients and attempt to find the significance in relevance with molecular pathogenesis. Clinically suspected patients for FRDA were genetically diagnosed for GAA repeat expansion using Long Range PCR (polymerase chain reaction). Total RNA was isolated from the plasma and subjected to miRNA microarray analysis using the Agilent microarray platform. Deferentially regulated miRNAs were validated by Sybr green Real Time PCR. We identified twenty-one (21) deregulated miRNAs (FDR p<0. 01, fold change ≥1. 5≤) when compared with healthy controls. Out of which 17 miRNAs were upregulated and 4 miRNAs were downregulated. The majority of the identified miRNAs reported to be highly expressed in the brain. Further validation (n= 18) through Quantitative RT-PCR found significant (p< 0. 05) upregulation in 3 miRNAs, hsa-miR-103a-3p, hsa-miR-15b-5p, hsa-miR-21-5p in comparison with healthy controls. Target and pathway analysis of these miRNAs have shown association with neurodegenerative and cardiomyopathy features. However, there was no association identified with clinicopathological features of FRDA patients. Considering progressive decline in the function of nervous and cardiac systems in FRDA individuals, we expect the present study may aid in the monitoring of effective therapeutics.

1243F

A neuronal/synaptic carnitine deficiency hypothesis for prevention of 10-20% of autism. *A. Beaudet.* Molecular & Human Genetics, Baylor Col Med, Houston, TX.

The discovery that X-linked *TMLHE* deficiency prevents endogenous biosynthesis of carnitine and is a likely risk factor for autism has led us to hypothesize that carnitine deficiency in the brain in the first three years of life may cause autism. *TMLHE* deficiency itself is found in less than 1% of autism, but we hypothesize that diet, minor illnesses, and the many known common genetic variants affecting carnitine metabolism may contribute to carnitine deficiency causing as much as 10-20% of autism. Boys with *TMLHE* deficiency typically have a normal physical exam except perhaps acquired macrocephaly and normal structural imaging of the brain. We have reported regression in association with very low plasma carnitine in one *TMLHE* deficient male with autism. The male:female ratio is particularly high in milder autism, in non-dysmorphic autism, and in Asperger syndrome. One study reported a 23 to 1 male to female ratio in autism probands with a normal physical exam and normal structural imaging of the brain. This extreme sex bias in a significant fraction of autism is the basis for proposing that this hypothesis might involve 10-20% of all autism. We also speculate that lack of X-inactivation for the *SLC6A14* putative blood-brain barrier carnitine transporter might explain the sex bias. We hypothesize that brain carnitine deficiency may be a common cause of autism in the subgroup of males with normal physical exam and brain imaging. We hypothesize that the risk of dietary deficiency would start at the time that solid foods, usually very low in carnitine, are introduced into the diet and might continue until meat and especially beef was added into the diet. There are some reports of low plasma carnitine in autism, but these are not from children in the 6 to 36 month age group that would be of greatest relevance to this hypothesis. We hypothesize that low brain carnitine levels might explain some cases of regression even in the presence of normal plasma levels of carnitine. An extensive review of the literature identifies findings for and against the neuronal/synaptic carnitine deficiency hypothesis. The effect of carnitine deficiency on synaptic development is being studied in mouse models of abnormal carnitine metabolism, and human genetic data are also being analyzed. We urge implementation of a clinical trial of carnitine supplementation in infant siblings of autism probands, where a very high recurrence risk is reported.

1244W

A Mendelian randomization investigation of the causal relationship between atopic dermatitis and psychobehavioral outcomes. *L. Paternoster, H. Sallis.* MRC Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom.

Atopic dermatitis (AD, or eczema) is a chronic inflammatory skin condition, common in early childhood (affecting ~20% of children). AD sufferers are at increased risk of some physical, psychological and psycho-social problems, including depression and conduct problems, suggesting successful early treatment of AD may have a wider public health impact than management of the immediate symptoms alone. However, these observational associations could be influenced by confounding and might not represent causal effects. We attempted to address this issue by using Mendelian randomization (MR) analysis to determine the causal relationship between AD and pyschobehavioral outcomes. AD was classified according to data collected throughout childhood for participants enrolled in the Avon Longitudinal Study of Parents and Children (ALSPAC). Parent-reported adolescent mental health was measured using the Strengths and Difficulties Questionnaire (SDQ) completed when the child was 12 and 16 years old. Observational analyses identified associations between AD and increased hyperactivity scores at age 12 and increased conduct problems at age 16. MR uses genetic instruments to proxy a modifiable risk factor and is largely free from issues such as confounding and reverse causality, thus strengthening causal inference in the use of observational data. However, potential pleiotropic effects of the variants should be considered. We constructed two instruments for AD; 1) using mutations in the *FLG* gene (known to impact eczema through skin barrier dysfunction) (F=38. 2), and 2) using SNPs identified in a recent genome-wide association study and believed to effect AD through immune mechanisms (F=30. 4). By using multiple instruments representing different etiological pathways of AD we can investigate if pleiotropic mechanisms are at play. We found no evidence of a causal association between AD and hyperactivity scores using either instrument. However, for conduct disorder, the two genetic instruments showed contrasting effects. The immune SNPs were associated with decreased conduct problems and the *FLG* mutations were associated with increased conduct problems. We also investigated directional pleiotropy using Egger regression. Our results suggest that the causal relationship between AD and pyschobehavioral outcomes is complex. Pleiotropic mechanisms are likely to exist and it is not clear what effect early successful treatment and/or prevention of AD may have on these outcomes.

1245T

Using genetic information to predicting response to treatment in patients with epilepsy. *I. Lopes-Cendes[1], R. Secolin[1], M. S. Silva[1], B. S. Carvalho[1], E. Bilevicius[2], C. V. Maurer-Morelli[1], F. Cendes[2].* 1) Dept Med Gen, Univ Campinas - UNICAMP, Campinas, SP, Sao Paulo, Brazil; 2) Dept Neurology, Univ Campinas - UNICAMP, Campinas, SP, Sao Paulo, Brazil.

Background: Mesial temporal lobe epilepsy (MTLE) is the most common form of adult epilepsy. Currently, the only characteristic, which seems to predict poor response to clinical treatment in these patients, is the presence of hippocampal sclerosis (HS). Studies showed that single nucleotide polymorphisms (SNPs) in genes encoding drug transporter and metabolism proteins could influence response to drug therapy. **Objective:** To evaluate whether combining information from clinical variables and SNPs in candidate genes could improve the discriminatory accuracy of predicting response to drug therapy in patients with MTLE. **Methods:** We evaluated 73 patients with MTLE who were responsive to antiepileptic drug (AED) therapy and 163 patients refractory to drug therapy. We genotyped 119 SNPs within the *ABCB1*, *ABCC2*, *CYP1A1*, *CYP1A2*, *CYP1B1*, *CYP2C9*, *CYP2C19*, *CYP2D6*, *CYP2E1*, *CYP3A4*, and *CYP3A5* genes. We assessed a first scenario using only clinical variables and a second one including SNP information. We used random forests combined with leave-one-out cross-validation to identify the best predictive model in each scenario and compared their discriminatory accuracies using the area under the curve statistic, by R software. Additionally, we built a variable importance plot to present the set of most relevant predictors on the best model. **Results:** The selected best model included the presence of HS and 53 SNPs among all genes evaluated. Additionally, including SNPs in the model improved the discriminatory accuracy from 0. 5865 to 0. 8244. **Conclusion:** To the best of our knowledge, this study demonstrated for the first time that genetic information improves the prediction of response to drug treatment in patients with MTLE. This may allow patients more likely to develop resistance to AED treatment to have a timely referral for epilepsy surgery. Supported by CEPID-BRAINN FAPESP, São Paulo, Brazil.

1246F

Characterization of APOE and TOMM40 Allele Frequencies in a Japanese Population. *H. Nonomura[1], A. Nishimura[1], S. Tanaka[1], M. Yoshida[1], Y. Maruyama[2], Y. Aritomi[1], A. Saunders[3], D. Burns[2], M. Lutz[3], G. Runyan[4], E. Lai[4], K. Budur[4], A. Roses[2].* 1) Takeda Pharmaceutical Company Limited, Osaka, Japan; 2) Zinfandel Pharmaceuticals, Inc. , Durham, NC, USA; 3) Duke University School of Medicine, Durham, NC, USA; 4) Takeda Development Center Americas, Inc. , Deerfield, IL, USA.

Identification of cognitively normal individuals at high-risk for developing Alzheimer's disease (AD) symptoms may enable early medical interventions or preventive therapies. Using longitudinal cohorts of elderly Caucasian subjects, a Biomarker Risk Assignment Algorithm (BRAA), comprised of Apoliprotein E (*APOE)* genotype, Translocase of outer mitochondrial membrane 40 homolog rs10524523 (*TOMM40* '523) genotype, and age of an individual was developed to predict the risk of developing mild cognitive impairment (MCI) due to AD within the next 5 years in the 65-83 year old age group. A clinical study (TOMMORROW; registered with ClinicalTrials. gov, NCT01931566) is being conducted to qualify the BRAA and evaluate the efficacy of pioglitazone 0. 8 mg slow release (SR) to delay the onset of MCI due to AD in high-risk subjects as determined by the BRAA. However, it is unknown if the BRAA developed in Caucasians is applicable to a Japanese population. In order to develop/validate the algorithm to identify Japanese high-risk individuals for a potential delay of onset study, it is important to characterize the allele and haplotype architecture of the *TOMM40* '523 *APOE* variations in this population. We determined frequencies of *APOE* and *TOMM40* '523 genetic polymorphisms and haplotypes in a Japanese population cohort (ages 20-68, N=300, 150 males and 150 females) from whole blood samples and compared them with samples from previous findings in Caucasian, African, and African-American subjects. Point estimates show that Japanese subjects had a lower frequency of *APOE*ϵ4 than Caucasian populations and a much lower frequency of *APOE*ϵ4 than African and African-American populations. For the *TOMM40* '523 polyT allele frequency, based on point estimates, Japanese subjects had higher frequencies of the Long (L20-29 Ts) alleles and Very Long (VL>29 Ts), and lower frequencies of the Short (S<20 Ts) alleles than other ethnicities compared. For the Japanese subjects, we did not find any extra-long alleles (T40-T57), which were occasionally observed in African and African-American subjects. In addition, we observed *APOE*ϵ4/*TOMM40* '523 S haplotype, which is rarely observed in Caucasian cohorts. These findings highlight the importance of determining the genetic architecture of this region of distinct ethnic groups for assessing its contribution to AD risk.

1247W

Progression rate from intermediate to advanced age-related macular degeneration varies with the number of risk alleles at the *CFH* locus. *P. J. Persad[1], R. J. Sardell[1], S. G. Schwartz[2], J. L. Kovach[2], J. A. Fortun[2], M. A. Brantley[3], A. Agarwal[3], J. L. Haines[4], W. K. Scott[1], M. A. Pericak-Vance[1].* 1) Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL; 2) Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL; 3) Ophthalmology and Visual Sciences, Vanderbilt University School of Medicine, Nashville, TN; 4) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH.

As a complex disease, risk of advanced age-related macular degeneration (AMD) is associated with both demographic and environmental factors, such as age, sex, smoking history, and diet, along with common and rare genetic variants at over 19 loci. However, risk is just one facet of AMD disease architecture; disease progression rate is another crucial aspect of variability among individuals. Whether genetic variants also affect progression rate from early/intermediate stages to advanced AMD is largely unknown. We tested whether genetic risk at 4 major common AMD risk-associated loci (*CFH*, *ARMS2*, *C2*, and *C3*) also predicts disease progression. We recruited AMD cases of European ancestry and monitored disease progression via regular clinical exams. Subjects were graded using the modified Age-Related Eye Disease Study (AREDS) scale. We tested whether progression rate from intermediate (grade 3) to advanced AMD (grade 4, geographic atrophy, or grade 5, choroidal neovascularization) was associated with the number of risk alleles (0, 1, or 2 risk alleles) at *CFH* (rs10737680), *ARMS2* (rs10490924), *C2* (rs429608), and *C3* (rs2230199), accounting for age at first grade 3 exam and sex. Cox proportional-hazards models were fit to allow for right censoring (some eyes had not progressed by the most recent exam), and for repeated measures (both left and right eyes may be included). Over the study period (max. 13 years), 172 (of 550) eyes from 376 individuals progressed from grade 3 to grade 4 or 5. Progression rate differed with the number of risk alleles at the *CFH* locus; individuals with 2 risk alleles progressed faster than those with 1 risk allele [hazard ratio (HR) = 1. 60, 95% CI 1. 07-2. 39, p=0. 02]. Genotype at the other 3 risk loci did not predict progression, but older subjects had higher hazard ratios in all models [HR= 1. 05, 95% CI=1. 03-1. 08, p<0. 01]. Sex was not significant in any model. These results support other studies' findings that progression rate to advanced AMD varies with the number of risk alleles at the *CFH* locus. However, since the effect of *CFH* was not strong and effects for the other 3 risk loci were not significant, additional genetic or environmental factors are likely to explain variation in AMD progression. These results warrant the need to test for association with other known AMD risk variants, novel loci, and potential environmental predictors.

1248T

Identification of IKAP responsive-genes as biomarkers for therapy of Familial Dysautonomia. *E. Morini, A. Ragavendran, M. Salani, S. Erdin, A. Stortchevoi, A. Brenner, M. Nilbratt, M. Talkowski, S. Slaugenhaupt.* Center for Human Genetic Research, Massachusetts General Hospital Research Institute/Harvard Medical School, Boston, MA.

Familial dysautonomia (FD) is a fatal, congenital sensory and autonomic neuropathy. We previously identified the genetic cause of FD, namely a "leaky" mRNA splicing defect in *IKBKAP* gene that results in reduced levels of IKAP protein. IKAP, or ELP1, is a scaffolding protein of the Elongator complex which is required for efficient transcriptional elongation. We found that kinetin can correct the *IKBKAP* splicing defect and increase the amount of normal mRNA and protein in FD cell lines and in mice following oral dosing. Our ongoing analyses are seeking to produce a new class of splicing modulators with significantly increased potency and efficacy over kinetin. If successful, such efforts could derive a therapy for FD by enhancing exon 20 inclusion and consequently increasing production of functional IKAP protein. Therefore, the identification of early disease-relevant pathways will permit the identification of biomarkers that can be used to assess the *in vivo* efficacy of therapies that alter IKAP levels. To identify IKAP-regulated pathways necessary for proper development of the sensory and autonomic nervous system, transcriptome sequencing analysis was performed in embryos expressing increasing amounts of IKAP: *Ikbkap*-/-, TgFD1/*Ikbkap*-/-, TgFD9/*Ikbkap*-/- and WT embryos. Consistent with the role of IKAP in transcriptional elongation, approximately 60% of all differentially expressed genes were down-regulated, and overall, down-regulated genes were longer than up-regulated genes. Importantly, we identified 262 genes whose expression increased strictly as a monotonic function of IKAP levels. These genes highlight pathways involved in nervous system development, including synapse formation, neuron differentiation, axon guidance, axon growth and neuronal cell adhesion. Remarkably, the IKAP dose-responsive genes show strong enrichment of genes involved in neurotrophin signaling pathway ($p < 10\text{-}4$). Many genes known to be associated with clinically similar types of hereditary sensory and autonomic neuropathy are involved in this pathway, including *NTRK1* for HSAN IV and *NGF* for HSAN V. Our findings emphasize the vital role of IKAP in embryo development and indicate that IKAP levels tightly regulate the expression of many genes involved in neurodevelopmental pathways. Further characterization of these pathways could yield robust biomarkers that will allow us to determine efficacy of promising therapeutics.

1249F

Cancer frequency among Machado-Joseph disease/Spinocerebellar Ataxia Type 3. *L. B. Jardim[1, 2, 5, 6, 7], G. N. Souza[1], N. Kersting[1], D. L. O. Pacheco[3], A. C. Krum-Santos[2], M. L. Saraiva-Pereira[4, 6], J. A. M. Saute[1, 6].* 1) Post-Graduation Program in Medical Sciences, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; 2) Medicine School, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; 3) Social Service School, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; 4) Department of Biochemistry, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; 5) Department of Internal Medicine, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; 6) Medical Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil; 7) Instituto Nacional de Genética Médica Populacional (INAGEMP), Porto Alegre, Brazil Porto Alegre, Brazil.

Background: Former studies reported a reduced risk of cancer among polyglutamine (polyQ) disorders, suggesting that poliQ protects against cancer. This was not tested yet in Machado-Joseph disease/spinocerebellar ataxia type 3 (MJD/SCA3). **Aims**: to compare the cumulative incidence (CI) of cancer, and the proportion of cancer among causes of death of MJD/SCA3 symptomatic individuals with those found in unrelated controls, in the last 15 years. **Methods**: after consent, structured interviews were applied in individuals older than 18 years-old: unrelated controls and symptomatic SCA3/MJD cases registered at our institution in Rio Grande do Sul, Brazil, where 180 families with molecular diagnosis and 700 symptomatic individuals were living by May, 2015. Cancer occurrences were obtained. The CIs were also compared to the expected number of cancers for this population, as published by "Instituto Nacional do Cancer" (INCA), Brazil, 2014. Interviews also asked for dead affected relatives of SCA3/MJD individuals in the last 15 years. Causes of death were obtained, as well as those from two death controls of similar age and same gender, that occurred imediatelly after or before, according to death registry "Sistema de Informacao sobre Mortalidade" (SIM). **Results**: 154 SCA3/MJD cases (45±12 years-old; 44. 2% males) (41 families), and 80 unrelated controls (45±13 years-old, 32. 5% males) were interviewed between May 2014 and April 2015. The CI of cancer in the last 15 years was of 7/154 (4. 5%) in SCA3/MJD and of 5/80 (6. 25%) in unrelated controls (χ^2=0,314; p=0. 575; RR 1. 375, CI 95% 0. 451 to 4. 195). The expected CI for this population in this period, according to INCA data, was 6%. With the absolute reductions of risk (ARR) of 1. 8%, 1,938 individuals per group were needed for a study with a 80% power to detect significant differences (p<0. 05). The SCA3/MJD individuals reported 55 deaths among their affected relatives in the last 15 years. Death causes were compared to 108 deaths of controls. Cancer was a cause of death in 4/55 (7. 3%) SCA3/MJD and in 27/108 (24. 5%) non-SCA3/MJD individuals (χ^2=7. 17; p=0. 007): an ARR of 17. 2% and a relative risk of 0. 296 (CI 95%: 0. 109-0. 805) of dying from cancer were associated to SCA3/MJD. **Discussion**: the study of CI of cancer is unattainable because sample sizes needed far exceed the overall population of SCA3/MJD in our state. Notwithstanding, SCA3/MJD was associated to a significant reduction of risk of death related to cancer.

1250W

Genetic aspects of sporadic spinocerebellar degeneration in the Japanese population. *Y. Takahashi, M. Kanai, S. Watanabe, M. Murata.* Dept Neurology, National Center of Neurology and Psychiatry, Kodaira, Tokyo, Japan.

[Background]Spinocerebellar degeneration (SCD) is a group of disorders characterized by progressive ataxia caused by degeneration of cerebellum or its afferent or efferent connections. By excluding multiple system atrophy, approximately 60% of SCD are familial SCD (fSCD) and the remaining sporadic SCD (sSCD). Genetic backgrounds of fSCD are highly heterogeneous: the modes of inheritance include autosomal-dominant (AD-SCD), autosomal-recessive (AR-SCD) or X-linked (X-SCD), with numerous causative genes. Meanwhile, the clinical entity of late-onset sSCD contains wide variety of diseases with diverse etiology, whereas that of early-onset sSCD presumably consists mostly of AR-SCD. [Purpose]To elucidate genetic aspects of sSCD in the Japanese population. [Subjects and Methods] Sixty patients with clinical diagnosis of sSCD excluding MSA were enrolled in this study. The ages of onset of the patients ranged from 0 to 83 years. DNA samples obtained with informed consent were subjected initially to the fragment analysis for triplet repeat expansion diseases including SCA1, 2, 3/MJD, 6, 8, 12, 17 and dentate-rubro-pallido-luysian atrophy (DRPLA), followed by fragment analysis and repeat-primed PCR for SCA31. DNA samples from 29 patients with the ages at onset below 40 were further subjected to the mutational analysis of *APTX* and *SACS* employing direct nucleotide sequencing method. This study was approved by the Institutional Review Board of National Center of Neurology and Psychiatry. [Results] Causative mutations were identified in 5 patients (8. 5%) with sSCD, consisting of 2 patients with SCA6, 2 with DRPLA, and one with SCA31. The ages at onset were 37 and 65 years old with SCA6, 19 and 57 years old with DRPLA, and 60 years old with SCA31. The expanded CAG repeat numbers were 24 and 22 in SCA6 (normal; 4~19) and 60 and 59 in DRPLA (normal; 6~36), respectively. In contrast, no homozygous or compound heterozygous pathogenic mutations were identified in *APTX* or *SACS*. [Discussion and Conclusion] Substantial proportion of patients with sSCD harbored causative mutations of AD-SCD, demonstrating that screening of AD-SCD in early-onset as well as late-onset sSCD proved to be fruitful. In contrast, screening of causative genes for AR-SCD did not establish genetic diagnosis, indicating that exome analysis would be efficient for elucidating genetic aspects of early-onset sSCD. Optimization of diagnostic algorithms is necessary for accurate clinical diagnosis of sSCD.

1251T

RAN Translation in Huntington disease. *M. Banez-Coronel[1,2], F. Ayhan[1,2], A. D. Tarabochia[1,2], T. Zu[1,2], B. A. Perez[1,2], S. K. Tusi[1,2,6], O. Pletnikova[7], D. R. Borchelt[1,4], C. A. Ross[8,9,10,11,12,13,14], R. L. Margolis[8,9,10,12,14], A. T. Yachnis[5], J. C. Troncoso[7], L. P. W. Ranum[1,2,3,6].* 1) Center for NeuroGenetics. University of Florida, Gainesville, FL; 2) Dept. of Molecular Genetics and Microbiology. University of Florida, Gainesville, FL; 3) Dept. of Neurology. University of Florida, Gainesville, FL; 4) Dept. of Neuroscience. University of Florida, Gainesville, FL; 5) Dept. of Pathology, Immunology and Laboratory Medicine. University of Florida, Gainesville, FL; 6) Genetics Institute. University of Florida, Gainesville, FL; 7) Dept. of Pathology. The Johns Hopkins University School of Medicine. Baltimore, MD; 8) Division of Neurobiology. The Johns Hopkins University School of Medicine. Baltimore, MD; 9) Dept. of Psychiatry. The Johns Hopkins University School of Medicine. Baltimore, MD; 10) Dept. of Neurology. The Johns Hopkins University School of Medicine. Baltimore, MD; 11) Dept. of Pharmacology. The Johns Hopkins University School of Medicine. Baltimore, MD; 12) Dept. of Neuroscience. The Johns Hopkins University School of Medicine. Baltimore, MD; 13) Program in Cellular and Molecular Medicine. The Johns Hopkins University School of Medicine. Baltimore, MD; 14) Baltimore Huntington's Disease Center. The Johns Hopkins University School of Medicine. Baltimore, MD.

Huntington disease (HD) is caused by a CAG•CTG expansion in the *huntingtin (HTT)* gene. Although most HD research has focused on the downstream effects of the HTT polyGln-expansion protein, the discovery of repeat associated non-ATG (RAN) translation raises the possibility that additional HD mutant RAN proteins are expressed across the CAG•CTG expansion mutation and contribute to the disease. We now report that four novel, homopolymeric expansion proteins (polyAla, polySer, polyLeu & polyCys) accumulate in HD human brains. These sense and antisense HD-RAN proteins accumulate most abundantly in brain regions with substantial neuronal loss, microglial activation and apoptosis, including the caudate/putamen and in juvenile-onset cases, also the cerebellum. RAN protein accumulation and aggregation patterns are length-dependent. Codon substitution experiments show that HD-RAN proteins are toxic to neural cells independent of RNA effects. Taken together, these data support a pathogenic role of RAN proteins in HD and suggest that therapeutic strategies targeting both sense and antisense genes may be required for efficacy. This is the first demonstration that RAN proteins are expressed across an expansion located in an open-reading frame and suggest RAN translation may also contribute to other polyglutamine diseases.

1252F

C9orf72 repeat expansion detection using short-read whole-genome sequencing data. *M. A. Eberle[1], J. F. A. van Vugt[2], R. J. Shaw[3], W. van Rheenen[2], A. M. Dekker[2], M. A. Bekritsky[3], S. S. Ajay[1], S. Chowdhury[1], T. Hambuch[1], R. Taft[1], D. Bentley[3], L. H. van den Berg[2], J. H. Veldink[2].* 1) Department of Population and Medical Genomics, Illumina, Inc. , San Diego, CA; 2) Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands; 3) Illumina Cambridge Ltd. , Chesterford Research Park, Saffron Walden, Essex, CB10 1XL, UK.

The hexanucleotide repeat expansion of CCGGGG in the *C9orf72* gene has been recognized as a major cause of ALS and Frontotemporal Dementia (FTD) and been implicated in other neurodegenerative and psychiatric disorders. Approximately 30% of familial ALS cases, 20% of familial FTD cases, 6% of sporadic ALS cases and 5% of sporadic FTD cases have been attributed to the *C9orf72* repeat expansion. While the exact pathological cut-off has not yet been established, the *C9orf72* repeat is considered expanded when it counts more than 30 hexamers (>180bp). Since it is so prevalent, all genetic studies in ALS and FTD must be stratified or adjusted for *C9orf72* repeat status, requiring a continuous and separate RP-PCR analysis. Moreover, the exact length of the repeat expansion is difficult, if not impossible, to determine and currently involves a combination of RP-PCR, Southern blotting and targeted sequencing. Here, we present a sequence-based method to accurately determine the length of any 3-6 nucleotide repeat, even when they are extremely expanded, in paired-end whole-genome sequence data. Sequencing was performed on 349 individuals (including 77 with the C9orf72 expansion) on an Illumina HiSeq 2000 using 2x100bp paired reads to ~40x depth. Of the 77 *C9orf72* samples that RP-PCR identified as expanded, only 1 sample had a repeat count of less than 20 according to the sequence analysis, whereas 6 samples had a repeat count between 20 and 30 and the rest of the samples had a repeat count higher than 30. Of the 272 *C9orf72* samples that were not expanded according to RP-PCR, only 1 sample had a repeat count of 260 according to the sequence analysis, whereas 2 samples had a repeat count between 20-30 and the rest of the samples had a repeat count lower than 20. If a count of 30 is the pathological cutoff, sensitivity of the tool is 90. 9%, and specificity 99. 6%. If the true cutoff is 20, these numbers are 98,7% and 98. 9% respectively. Further validation of this tool is ongoing in independent samples and for other repeat expansions, which will be presented. Since microsatellite expansions are a common genetic risk factor in neurodegenerative diseases, accurately determining the length of a repeat at known genomic locations using NGS data and also discovering unknown repeat expansions will allow studying the impact of repeat expansion on neurodegenerative disorders.

1253W

Large-scale single-molecule sequencing of tandem repeats on the human X chromosome. *A. Zablotskaya[1,2], G. Peeters[1,2], W. I. M. Meert[2,3], H. Van Esch[2,4], K. J. Verstrepen[5], G. Froyen[2,6,7], J. R. Vermeesch[1,2].* 1) Laboratory for Cytogenetics and Genome Research, Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium; 2) Center for Human Genetics, University Hospitals Leuven, KU Leuven, 3000 Leuven, Belgium; 3) Genomics Core, University Hospitals Leuven, 3000 Leuven, Belgium; 4) Laboratory for Genetics of Cognition, Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium; 5) VIB Laboratory for Systems Biology & CMPG Laboratory for Genetics and Genomics, KU Leuven, 3001 Leuven, Belgium; 6) Human Genome Laboratory, Department of Human Genetics, KU Leuven, 3000 Leuven, Belgium; 7) Jessa Hospital, Laboratory for Molecular Diagnostics, Clinical Biology, 3500 Hasselt, Belgium.

Tandem repeats are short DNA sequences that are repeated head-to-tail with a propensity to be variable. They constitute a significant proportion of the human genome, also occurring within coding and regulatory regions. Variation in these repeats can alter the function and/or expression of genes allowing organisms to swiftly adapt to novel environments. Importantly, some repeat expansions have also been linked to certain diseases. Unfortunately, due to the nature of short read sequencing technologies, tandem repeats are not analyzed during whole genome or exome sequencing studies. We developed a novel capture assay for large-scale genotyping of tandem repeats (Duitama J. , Zablotskaya A. et al. , Nucl. Acids Research, 2014) and extended the assay for the identification of X linked disease-related repeats using long read (averaging 12 kb) PacBio RS II technology. For 837 (83% of all) potentially functional repeats, unique capture baits were designed, as well as for 1000 intronic and intergenic repeats. Of these, a full tandem repeat length sequence was obtained for 88-90% of the targets in male DNA samples. Sequencing read length and analysis pipeline allows to detect cases of tandem repeat expansion. We are currently implementing this assay to screen for potentially causal variation underlying X-linked disorders that are not explained following array and exome sequencing. Grant references: Fonds voor Wetenschappelijk Onderzoek (FWO)-Vlaanderen [G. 0795. 11 to K. J. V. , J. R. V. , G. F.]; Marguérite-Marie Delacroix [GV/B-155 to A. Z. , G. F].

1254W

Can rare variants account for signals from common variants? *F. Begum[1], M. A. Taub[2], I. Ruczinski[2], E. J. Leslie[3], D. C. Koboldt[4], T. H. Beaty[1], J. C. Murray[5], M. L. Marazita[3,6], Cleft Seq 1 Consortium.* 1) Department of Epidemiology, Johns Hopkins University, Baltimore, MD; 2) Department of Biostatistics, Johns Hopkins University, Baltimore, MD; 3) Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA; 4) The Genome Institute,Washington University School of Medicine, St. Louis, MO; 5) Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, IA; 6) Department of Human Genetics, Graduate School of Public Health, and Clinical and Translational Science Institute, School of Medicine, University of Pittsburgh, Pittsburgh, PA.

While genome-wide association studies (GWAS) have successfully identified polymorphic markers associated with complex diseases, few are directly causal. High-throughput sequencing holds the promise of identifying directly causal mutations. Several statistical methods for sequencing data under case-control study designs are available, but fewer methods for family-based studies exist. Also, rare variants (RV) capable of producing deleterious gene products cannot be identified in conventional association study designs due to lack of power. We are proposing a new empirical method, which is primarily motivated by the concept of 'synthetic association' with a common marker and disease being a reflection of multiple causal RVs. Our method looks for an excess of putatively functional RVs on transmitted risk haplotypes tagged by a common marker showing significant association in the allelic transmission disequilibrium test (aTDT), compared to corresponding untransmitted risk and non-risk haplotypes. We illustrate our approach using 1,409 case-parent trios ascertained through a child with a non-syndromic oral cleft from 3 ethnic groups (Chinese, Filipino and European) sequenced on 6. 3 Mb in 13 candidate genes/regions. We used counts of putatively deleterious RVs transmitted with the common risk allele and compared them to counts of RVs on both untransmitted risk and non-risk haplotypes, and found some significant differences for selected genes. We designed permutation-based tests to assess evidence that non-random sets of RVs are transmitted to the affected child. We observed significant over-transmission of putatively deleterious RVs in the regions around *VAX1*, *IRF6*, *MAFB* and *FOXE1* (all candidate genes for oral clefts) and borderline empirical significance in other regions. We are following up our findings to identify which among these deleterious RVs could be actually casual. We will expand this approach to consider local linkage disequilibrium structure and further characterize functionality of these RVs, which may offer a viable method for resolving causal variants that could explain associations found with polymorphic markers in GWAS.

1255T

A Comparison Study of Fixed and Mixed Effect Models for Gene Level Association Studies of Complex Traits. *C. Chiu[1], J. Jung[2], D. Weeks[3], A. Wilson[4], J. Bailey-Wilson[4], C. Amos[5], M. Xiong[6], R. Fan[1].* 1) Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH); 2) National Institute on Alcohol Abuse and Alcoholism, NIH; 3) University of Pittsburgh; 4) National Human Genome Research Institute, NIH; 5) Dartmouth Medical College; 6) University of Texas - Houston.

In association studies of complex traits, fixed effect regression models are usually used to test for association between phenotypic traits and major gene loci. In recent years, variance-component tests based on mixed models were developed for region-based genetic variant association tests. In the mixed models, the association is tested by a null hypothesis of zero variance via a sequence kernel association test (SKAT) and its optimal unified test (SKAT-O). Although there are some comparison studies to evaluate the performance of mixed and fixed models, there is no systematic analysis to determine when the mixed models perform better and when the fixed models perform better. Here we evaluated, based on extensive simulations, the performance of the fixed effect and mixed model statistics, using genetic variants located in 6, 9, 12, and 15 kb simulated regions. We compared the performance of three models: (1) mixed models which lead to SKAT and SKAT-O, (2) traditional fixed effect additive effect models, and (3) fixed effect functional regression models. We performed simulation analyses for two scenarios: (1) all causal variants are rare; (2) some causal variants are rare and some are common. We found that the fixed effect tests have accurately controlled false positive rates. In most cases, either one or both of the fixed effect models performed better than or similar to the mixed models, when some causal variants are rare and some are common, or when all causal variants are rare except for the 12 and 15 kb region cases. We argue that the fixed effect models are useful in most cases. In practice, it makes sense to perform analysis by both the fixed and mixed effect models and make a comparison, and this can be readily done using our R codes and the SKAT and SKAT-O packages.

1256F

A Statistical Approach for Testing Pleiotropic Effects of Rare Variants. *M. P. Epstein[1], K. A. Broadaway[1], D. J. Cutler[1], R. Duncan[1], J. L. Moore[2], E. B. Ware[3], M. A. Jhun[3], L. F. Bielak[3], W. Zhao[3], J. A. Smith[3], P. A. Peyser[3], S. LR. Kardia[3].* 1) Department of Human Genetics, Emory University, Atlanta, GA; 2) Department of Evolution and Ecology, University of California, Davis, Davis, CA; 3) Department of Epidemiology, University of Michigan, Ann Arbor, MI.

Empirical findings suggest many trait-influencing genetic factors are associated with multiple distinct phenotypes; these associations are defined as pleiotropic or cross-phenotype associations. When cross-phenotype effects exist, association methods of multiple phenotypes that leverage pleiotropy are more powerful than univariate methods that consider each phenotype separately. While several statistical approaches exist for testing pleiotropy in genome-wide association studies, there is a lack of statistical tools for pleiotropic analysis of rare variation that aggregate information across a gene or region of interest. In order to fill this important gap, we introduce a new statistical method for gene-based pleiotropic analysis of rare variants using a nonparametric distance-covariance approach that constructs sample pairwise similarity (or dis-similarity) matrices for phenotypes and genotypes and then evaluates whether the individual elements in the two matrices are independent. The approach possesses many valuable features. The method can accommodate an arbitrary number of categorical and/or continuous phenotypes, thereby scaling easily to handle high-dimensional data, and further can adjust for covariates. Further, the resulting test is computationally efficient and scalable to genome-wide application since the test has a closed form (circumventing complicated optimization) and follows an asymptotic distribution thereby allowing analytic derivation of P-values usingDavies' exact method. We use simulated data to demonstrate that our method, which we refer to as the Gene Association with Multiple Traits (GAMuT) test, provides increased power over univariate SKAT testing of individual traits when pleiotropy exists. We also illustrate our approach using exome-chip data from the Genetic Epidemiology Network of Arteriopathy.

1257W

Functional Regression Models for Gene-based Association Studies of Complex Traits. *R. Fan[1], C. Chiu[1], Y. Wang[2], J. Jung[3], Y. Jiang[4], W. Chen[4], D. Weeks[4], H. Ren[5], C. Amos[6], A. Wilson[7], J. Bailey-Wilson[7], M. Xiong[8].* 1) Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH), Bethesda, MD; 2) Food and Drug Administration (FDA), Silver Spring, MD; 3) National Institute on Alcohol Abuse and Alcoholism, NIH, Bethesda, MD; 4) University of Pittsburgh, Pittsburgh, PA; 5) Regeneron Pharmaceuticals, Inc. , Basking Ridge, NJ; 6) Dartmouth Medical College , Lebanon, NH; 7) National Human Genome Research Institute, NIH, Bethesda, MD; 8) University of Texas - Houston, TX.

By using functional data analysis techniques, fixed effect functional regression models are developed to test associations between complex traits and genetic variants, which can be rare or common variants or a combination of the two, adjusting for covariates. By treating multiple genetic variants of an individual in a human population as a realization of an underlying stochastic process, the genome of an individual in a chromosome region is a continuum of sequence data rather than discrete observations. The genome of an individual is viewed as a stochastic function which contains both genetic position and linkage disequilibrium information of the genetic markers. Then, the genetic effect of genetic variants is assumed to be a smooth function, i. e. , a continuous genetic effect function, which can be estimated by B-spline or Fourier basis functions. The association between phenotypic traits and genetic variants is tested by testing a null hypothesis that the genetic effect function is equal to 0. In our published papers [Genetic Epidemiology 37:726-742 (2013), 38:622-637 (2015), and 39:259-275 (2015)], test statistics of the proposed functional regression models are built to test association between phenotypic traits (which can be quantitative or dichotomous) and genetic variants. Extensive simulation analysis and real data analysis were performed to demonstrate that the proposed models and tests perform better in most cases than existing popular procedure of sequence kernel association test (SKAT) and its optimal unified test (SKAT-O). In our unpublished work, we extend the functional regression approach to pedigree data to analyze quantitative or dichotomous traits. One motivation is the superior performance of the functional regression models in analyzing population data, and this merit should be useful for analyzing pedigree data. We build functional linear mixed models to model the phenotypic traits. Again, the genetic effect of genetic variants is modeled by a genetic effect function, and the relation among the family members is modeled by a variance-covariance matrix as variance component models of traditional population genetics. The performance of the model is being evaluated and the methods will be used to analyze complex traits with exome chip genotyping of pedigree data. The proposed models can be useful in whole genome and whole exome association studies of complex traits.

1258T

A heterogeneity sequence association test identified novel genes of alcohol addiction in African Americans. *W. Ouyang[1,2], HZ. Qin[1,2].* 1) Center for Bioinformatics and Genomics; 2) Department of Biostatistics and Bioinformatics, Tulane University, New Orleans, LA.

Phenotypic variance heterogeneity is ubiquitous in systems epidemiology data from the multi-ethnic populations. It occurs via various means, such as genetic admixture, ancestral background linkage disequilibrium, genetic pleiotropy, as well as gene-by-gene and gene-by-environment interactions. Existing gene-based association methods essentially focus on valid analyses of mean effects and neglect variance effects. In this report, we proposed a heterogeneity sequence association test (HSAT) to integrate valuable marker-specific heterogeneity signals with group-wise mean effects. Under a wide range of simulation designs in major journals, the proposed HSAT appropriately controlled Type I error rate and appeared notably more powerful than several existing prominent gene-based sequence association tests, e. g. , the SKAT, SKAT-O and GEMMA. The proposed HSAT demonstrated even more noteworthy power gains over the established tests under simulated scenarios of admixture genetics. To illustrate its practical utility, we applied the HSAT to the DNA sequence and alcohol dependence data of 1265 African Americans from the SAGE. The HSAT identified *RIC3*, *IL36RN*, *RAM23, SPTBN1, PARP8* as novel significant genes that were missed by aforesaid prominent tests and were not reported in the literature. 41 out of 114 top significant genes identified by the HSAT have been reported to be associated with alcohol dependence.

1259F

Human Population Classification Based on Common and Rare Variants Using Random Forests, Artificial Neural Networks, and Support Vector Machines. *M. Rao[1], T. Mersha[2], M. Wathen[3], D. Tadesse[4].* 1) Environmental Health, Univ Cincinnati, Cincinnati, OH 45267; 2) Division of Asthma Research, Cincinnati Children's Hospital, 3333 Burnet Avenue, Cincinnati, OH 45229; 3) Division of Biostatistics and Bioinformatics, Department of Environmental Health, University of Cincinnati, OH 45267; 4) Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH 45267.

In genetic studies of common variants population structure is well-known and well-understood. Population structure in the realm of rare variants is unclear. The objective of the presentation is to analyze relationship between common and rare variants in terms of population structure by finding answers to the following questions. Is the population structure (or inferred ancestry) similar in common and rare variants? From a population classification persprctive, identify SNPs with the most inpact on classification. We perform extensive population structure analysis using common and rare variants data available from the 1000 Genomes Project for 11 populations. The data are divided into two groups including common and rare variants. Both variants are summarized across the study populations. For SNP data, each respective SNP is considered as a categorical variable with three genotypes. Three analytical approaches are compared including random forests (RF), neural network (NNET), and support vector machines (SVM) in terms of classification accuracy with respect to both rare and common variants. Performances are compared based on the models' predictive ability.

1260W

A New Method for Detecting Associations with Rare Copy-Number Variants. *J. Szatkieiwcz[1], J. Tzeng[2,3], P. Magnusson[4], P. Sullivan[1,4], The Swedish Schizophrenia Consortium.* 1) University of North Carolina, USA; 2) North Carolina State University, USA; 3) National Cheng-Kung University, Taiwan; 4) Karolinska Institutet, Stockholm, Sweden.

Copy number variants (CNVs) play an important role in the etiology of many diseases such as cancers and psychiatric disorders. Due to a modest marginal effect size or the rarity of the CNVs, collapsing rare CNVs together and collectively evaluating their effect serves as a key approach to evaluating the collective effect of rare CNVs on disease risk. While a plethora of powerful collapsing methods are available for sequence variants (e. g. , SNPs) in association analysis, these methods cannot be directly applied to rare CNVs due to the CNV-specific challenges, i. e. , the multi-faceted nature of CNV polymorphisms (e. g. , CNVs vary in size, type, dosage, and details of gene disruption), and etiological heterogeneity (e. g. , heterogeneous effects of duplications and deletions that occur within a locus or in different loci). Existing CNV collapsing analysis methods (a. k. a. the burden test) tend to have suboptimal performance due to the fact that these methods often ignore heterogeneity and evaluate only the marginal effects of a CNV feature. We introduce CCRET, a random effects test for collapsing rare CNVs when searching for disease associations. CCRET is applicable to variants measured on a multi-categorical scale, collectively modeling the effects of multiple CNV features, and is robust to etiological heterogeneity. Multiple confounders can be simultaneously corrected. To evaluate the performance of CCRET, we conducted extensive simulations and analyzed large-scale schizophrenia datasets. We show that CCRET has powerful and robust performance under multiple types of etiological heterogeneity, and has performance comparable to or better than existing methods when there is no heterogeneity.

1261T

An extended variance-component score test for association of multivariate phenotypes and gene-level rare variants. *I. Patanam[1], W. Qi[2,3], M. E. Garrett[1], B. Anderson[1], C. H. Yu[3], X. J. Qin[3], A. E. Ashley-Koch[1,2,4], Y. J. Li[2,3].* 1) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC; 2) Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC; 3) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC; 4) Department of Medicine, Duke University Medical Center, Durham, NC.

In genetic studies of complex diseases, investigators often collect multiple disease-related measurements in addition to the primary dichotomous disease outcome. Multivariate analysis can increase the power for genetic association tests and detect pleiotropy, where a single gene may affect multiple traits (Williams 1957). Methods for multivariate analysis such as multivariate analysis of variance (MANOVA) and generalized estimating equation (GEE) have been well documented in classical statistics literature. There is also a robust literature on multivariate genetic association tests for common variants, including application to genome-wide association study (GWAS). Here, motivated by the increasing availability of next generation sequencing data, we extended the concept outlined in the GEE-SPU method of Zhang et al (2014) for common variants and the SKAT score test (Wu et al. 2011) for a set of rare variants to a multivariate variance-component score test for multiple phenotypes at the gene level. Assuming *t* traits of interest, we utilized GEE to estimate the parameters of covariates and rare variants in order to compute the marginal mean of *t* traits for each individual. We derived a variance component score statistic (*Q*) with form similar to that of SKAT but in *t*t* dimension for *t* traits. We, therefore, propose a multivariate variance component score test in a form of the sum of eigenvalues of the *Q* matrix, which is equivalent to the sum of *Qt* where *Qt* is the score statistic of the *tth* trait. Multiple weighting schemes for rare variants and traits were evaluated. Simulation studies are in progress to evaluate type I error and statistical power, and to compare the performance of our method to the TATES method (van der Sluis et al. 2013), which uses p-values derived from SKAT for each trait to generate a combined p-value for each gene. The new method and TATES were also applied to cranial morphologic data (as described in Markunas et al. , 2014) with targeted rare-variant data on a set of candidate genes. This new method provides a flexible solution to the analysis of rare variants as they contribute to multivariate phenotypes.

1262F

A unified powerful set-based test for sequencing dataanalysis of GxE interactions of rare variants. *Y. Su[1], U. Peters[1], S. Berndt[2], S. Bézieau[3], H. Brenner[4], P. Campbell[5], A. Chan[6], J. Chang-Claude[7], M. Lemire[8], P. Newcomb[1,9], J. Potter[1,9,10], M. Slattery[11], M. Woods[12], C. Di[1], L. Hsu[1].* 1) Public Health Science Division, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 3) Service de Génétique Médicale, CHU Nantes, Nantes, France; 4) Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany; German Cancer Consortium (DKTK), Heidelberg, Germany; 5) Epidemiology Research Program, American Cancer Society, Atlanta, GA; 6) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 7) Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany; 8) Ontario Institute for Cancer Research, Toronto, Canada; 9) School of Public Health, University of Washington, Seattle, WA; 10) Centre for Public Health Research, Massey University, Wellington, New Zealand; 11) Department of Internal Medicine, University of Utah Health Sciences Center, Salt Lake City, UT; 12) Discipline of Genetics, Memorial University of Newfoundland, St. John's, NL, Canada.

Next generation sequencing allows researchers to study comprehensively the contributions of genetic variations, in particular rare variants, and their interaction with environmental factors in relation to risk of common complex disease such as cancer. However, analysis of gene-environment interactions (GxE) for rare variants poses considerable challenges for making statistical inferences because very few subjects carry the variants and even fewer subjects who carry the variants are exposed to risk factors. To tackle this challenge, we propose a set-based statistical modeling and testing procedure that assesses the effects of a set of rare variants (e. g. , genes or regulatory regions). Under this framework, GxE is modeled by two components: an interaction (BxE) between a burden risk score (the total number of rare-variant minor alleles carried) and the environmental factor, and a random component of the interaction effects, which follows a distribution with mean 0 and variance tau2, to account for the residual GxE that has not been accounted for by BxE. We derive score tests for both components and combine the two score tests to gain power on GxE signal detection. The concept of the proposed combination here is different from that in meta-analysis, in that the two combined components are from the same study but represent different parameters, whereas for meta-analysis the combined components are from different studies, but estimate the same parameter. We propose two new data-adaptive combination approaches and establish the asymptotic distributions, which makes the genome-wide search for GxE computationally feasible. The extensive simulation results show that the proposed approaches maintain the correct type I error and possess power comparable to existing methods such as the sequence kernel association test for interaction effects (iSKAT, Lin *et al*. 2013). We applied the proposed methods and iSKAT to an exome chip dataset from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) to identify potential GxE interaction effects between 7,600 genes and NSAID use. We found a novel GxE interaction effect for *TELO2* that reaches exome-wide significance using the proposed methods, but failed to identify this finding using the iSKAT approach. The novel finding may provide new insights into the role of NSAID use and its interplay with genetic variation on colorectal cancer risk.

1263W

Generalized linear Mixed Model Association Tests (GMMAT) for rare variants to control for population stratification and relatedness in sequencing studies with continuous and binary phenotypes. *H. Chen[1], C. Wang[2], S. Lee[3], S. Redline[4], X. Lin[1].* 1) Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA; 2) Computational and Systems Biology, Genome Institute of Singapore, Singapore; 3) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 4) Division of Sleep and Circadian Disorders, Departments of Medicine and Neurology, Brigham and Women's Hospital, Boston, MA.

SNP-set based tests have been widely used for rare variant association analysis in sequencing studies, where a SNP set consists of a set of genetic variants in a gene, region or a pathway. Existing SNP-set based tests for binary traits such as burden tests, the sequence kernel association test (SKAT) and SKAT-O are applicable to independent samples. To control for population stratification, family and cryptic relatedness, and complex sampling designs in sequencing association studies, we propose a general framework of SNP-set based tests and meta-analysis using Generalized linear Mixed Model Association Tests (GMMAT). We show that applying linear mixed models to binary traits may lead to incorrect type I error rates in the presence of population stratification. We develop a computational efficient score based association tests by fitting generalized linear mixed models for rare variant analysis. Fitting generalized linear mixed models using GMMAT is almost two orders of magnitude faster than existing implementation in SAS PROC GLIMMIX. Simulation studies show that our tests have correct type I error rates for binary traits in the presence of population stratification and relatedness, while naïve burden tests, SKAT, SKAT-O and linear mixed model based SKAT and SKAT-O fail to control type I error rates. We compare the power of our tests in various scenarios and illustrate how they could be used to test a broad class of different scientific hypotheses. We also apply our methods to a case-control rare variant study to perform a trans-ethnic meta-analysis of both population-based and family-based studies.

1264T

Robust association testing for quantitative traits and rare variants. *P. Wei[1], Y. Cao[1], Z. Xu[2], Y. Zhang[2], J. Crosby[1], E. Boerwinkle[1], W. Pan[2].* 1) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 2) Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, MN.

With the advance of sequencing technologies, it has become a routine practice to test for association between a quantitative trait and a set of rare variants (RVs). However, to our knowledge, there is no study yet on the robustness of association testing to the non-Normality of the observed traits, e. g. due to skewness, which is expected to be ubiquitous for quantitative traits. By extensive simulations, we demonstrate that SKAT and SKAT-O are not robust to heavily-tailed or right-skewed error (or trait) distributions with inflated Type I error rates; in contrast, some other tests, including our proposed SPU tests and their adaptive version called aSPU test, are much more robust. We further propose a robust version of the SPU and aSPU tests, denoted as SPUr and aSPUr. We conduct extensive simulations to evaluate the power of the tests, finding that in most situations the aSPU test achieves similar power to that of SKAT or SKAT-O when testing on a smaller number of RVs; on the other hand, for a larger number of RVs, the aSPU test is often more powerful than others, owing to its high data-adaptivity. As expected, the aSPUr test loses power in the absence of outliers. We also compare different tests by conducting association testing of (1) triglyceride in 1731 individuals using the NHLBI ESP whole exome sequencing data, and (2) brain imaging phenotypes in 812 individuals using the Alzheimer's Disease Neuroimaging Initiative (ADNI) whole genome sequencing data. In the former, the QQ plots for SKAT, SKAT-O and T1 tests of the untransformed triglyceride were severely inflated (lambda = 1. 89, 1. 78 and 1. 13, respectively), while those for aSPU and aSPUr behaved normally (lambda < 1. 04). The positive control gene *APOC3* [MIM: 614028] (Crosby et al, NEJM 2014) was ranked 37th by aSPUr, but was not among the top 200 genes by all other methods. When applied to Log and inverse-normal transformed triglyceride, all tests had lambdas < 1. 05, but only T1, SKAT-O, aSPU and aSPUr ranked APOC3 among the top two. In the ADNI example, SKAT could have severely inflated QQ plots even with Log or inverse-normal transformed imaging phenotypes, for which aSPU and aSPUr controlled the lambda well. Due to its relatively high robustness to outliers and high power of the aSPU test, we recommend its use complementary to SKAT and SKAT-O. If there is evidence of inflated Type I error rate from the aSPU test, we would recommend the use of the more robust (but less powerful) aSPUr test.

1265F

High-dimensional genotypes and phenotypes in rare variant association studies. *M. Pirinen[1], C. Spencer[2], M. J. Daly[3,4], M. I. McCarthy[2,5], M. A. Rivas[2,3].* 1) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 3) Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 4) Analytical and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; 5) Oxford Centre for Diabetes, Endocrinology and Metabolism, Oxford, UK.

We consider assessing association between multiple rare variants and multiple phenotypes. This topic is particularly timely in the context of biobank studies where detailed genetic data can be combined with several correlated phenotypes. We propose the multiple rare variants and phenotypes (MRP) approach, a general statistical framework that is able to exploit knowledge about the correlation of genetic effects 1) across a group of genetic variants, 2) across a group of phenotypes, and 3) across a group of studies. Our implementation is efficient, making the analysis of large study designs practical; flexible and extensible making the analysis of gene-sets, pathways, and networks feasible; and includes standard univariate and multivariate gene-based tests as special cases. We work with summary statistics data permitting the joint analysis of multiple independent studies. Simulations show that on a standard laptop an exome-wide analysis of 10,000 to a million individuals can be achieved in 17 seconds to less than 3 days for the number of quantitative traits in the range 3-100. We study the method's ability to detect signal by simulations mimicking the effect of known protein truncating variants in *APOC3* on lipid levels. Results show that under many scenarios rare variant analysis across multiple variants and multiple phenotypes has a considerable advantage, measured by ROC area under the curve, over analyses that consider only a single variant or a single phenotype at a time. To support the analysis of rare variants across multiple phenotypes using summary statistics alone, as in RAREMETAL for univariate phenotypes, we extend our framework to allow for possible differences in genetic effects across ethnicities or populations. As expected, high concordance is observed ($r2 \approx 1.0$) between a joint-analysis of full phenotype and genotype data and the analysis based on the summary statistics only. Finally, we apply the framework to coding sequence data from 25 GWAS risk genes in UK residents comprising 24,892 subjects with six autoimmune disease phenotypes and 17,019 controls. We identify a potential new association of protein truncating variants at *TNFAIP3*. We envision that our method will be a valuable tool for future studies of complex traits in settings where high-dimensional genetic and phenotypic data are available.

1266W

Rare Variants Association Analysis in Large-Scale Sequencing Studies at the Single Locus Level. *J. Y. Tzeng[1,2], X. J. Jeng[1], W. Lu[1], Z. J. Daye[3].* 1) Department of Statistics, Campus Box 7566, North Carolina State University, Raleigh, NC 27695, USA; 2) Bioinformatics Research Center, Campus Box 7566, North Carolina State University, Raleigh, NC 27695, USA; 3) Department of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ 85724, USA.

Genetic association analyses of rare variants in next-generation sequencing (NGS) studies are fundamentally challenging due to the presence of a very large number of candidate variants at extremely low minor allele frequencies. Recent developments often focus on pooling multiple variants to provide association analysis at the gene instead of the locus level. Nonetheless, pinpointing individual variants is a critical goal for genomic researches as such information can facilitate the precise delineation of molecular mechanisms and functions of genetic factors on diseases. Due to the extreme rarity of mutations and high-dimensionality, significances of causal variants cannot easily stand out from those of noncausal ones. Consequently, standard false positive control procedures, such as the Bonferroni and false discovery rate (FDR), are often underpowered, as a majority of the causal variants can only be identified along with a few noncausal ones. To provide informative analysis of individual variants in large-scale sequencing studies, we propose the FAlse Negative control Screening (FANS) procedure in this paper. The proposed procedure is computationally efficient and can adapt to the underlying proportion of causal variants. Extensive simulation studies across a plethora of scenarios demonstrate that the FANS is advantageous for identifying individual rare variants, whereas the Bonferroni and FDR are exceedingly over-conservative for rare variants association studies. In the analyses of the CoLaus dataset, FANS has identified individual variants most responsible for gene-level significances. Moreover, single-variant results using the FANS have been successfully applied to infer related genes with annotation information.

1267T

Evaluating Power and Sample Size Considerations for Gene Discovery in Rare Diseases. *M. H. Guo[1,2,3], J. N Hirschhorn[1,2,3].* 1) Division of Endocrinology, Boston Children's Hospital, Boston, MA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 3) Department of Genetics, Harvard Medical School, Boston, MA.

Over the past few years, next generation sequencing has greatly facilitated gene discovery for rare genetic diseases. However, for many rare diseases, the genetic causes have not yet been found or have only been found for a subset of cases. The sample sizes needed to discover novel genetic causes from next generation sequencing are often unknown. Key considerations include genetic architecture (penetrance, locus heterogeneity) and the background rate of rare neutral variants that are difficult to distinguish from pathogenic mutations. Here, we have created an empirical simulation framework to evaluate power and sample size considerations in a typical scenario for rare disease gene discovery efforts, where unrelated probands are sequenced and a gene-based burden test is performed against suitable controls. Our simulation framework utilizes exome sequences in approximately 3000 individuals from the NHLBI Exome Sequencing Project and allele frequency data from the Exome Aggregation Consortium. Using our framework, we demonstrate how locus heterogeneity, penetrance, ability to distinguish neutral from pathogenic variants, minor allele frequency thresholds, mode of inheritance, and control population sizes each affect power. The simulation results demonstrate that, across genes, there is a wide range in the sample sizes needed to achieve a given power. We also found that increasing locus heterogeneity results in rapid increases in sample sizes needed. For example, under modest locus heterogeneity (each gene contributing to 5% of cases) and complete penetrance, approximately 300 cases are needed to achieve 90% power to detect a gene. We demonstrate the validity of our framework using existing results from several exome sequencing projects for rare diseases. Our studies provide a wide-ranging examination of power in rare disease gene discovery, with the added advantage of incorporating real sequencing data. Our studies provide useful guidance for study design for exome sequencing under a variety of genetic architectures. Broadly, our studies demonstrate that rare disease gene discovery will often require large sample sizes and necessitate international collaborations.

1268F

Rare variant association tests for pleiotropic traits in family studies. *Y. Chiu[1], L. Chien[1], F. Hsu[2], D. Brown[3].* 1) Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Miaoli County 35053, Taiwan, ROC; 2) Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA; 3) Personalized Medicine Research, Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA.

Given the functional relevance of many rare variants, their identification is frequently critical for dissecting disease etiology. Functional variants are likely to be aggregated in family studies enriched with affected members, and this aggregation increases the statistical power for rare variant detection. However, methods for the analysis of rare variants for families with pleiotropic traits remain fairly limited. These methods should be capable of accounting for different sources of correlations and handling large amounts of sequencing data efficiently. To identify rare variants for pleiotropic traits in family studies, we extended the powerful pedigree-based burden and kernel association tests to genetic longitudinal studies. Generalized estimating equation (GEE) approaches were used to generalize the pedigree-based burden and kernel tests to multiple correlated phenotypes under the generalized linear model framework. Adjustments were made for the fixed effects of confounding factors. These tests accounted for the complex correlations between multiple phenotypes and between individuals within the same family. Comprehensive simulation studies were conducted to compare the proposed tests with mixed-effects models and marginal models using GEEs under various configurations. The proposed tests were applied to the Diabetes Heart Study.

1269W

Conditioning Adaptive Combination of *P*-values to Analyze Case-Parent Trios with or without Population Controls. *W. Lin, Y. Liang.* Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan.

Rare causal variants (minor allele frequency (MAF) < 1%) detection can help to uncover the etiology of complex diseases. Recruiting case-parent trios is a simple and popular study design in family-based studies. If researchers can also obtain data from population controls, utilizing unrelated controls in trios can improve the power of a statistical method. The transmission disequilibrium test (TDT) is a well-known method to analyze case-parent trio data. It has been extended to rare-variant association testing (referred to as 'rvTDT' hereafter), with the flexibility to incorporate population controls. The rvTDT method is robust to population stratification because of its conditioning approach. However, power loss may occur in the conditioning process. We here propose a 'conditioning adaptive combination of *P*-values method' (abbreviated as 'conADA'), to analyze trios with/without unrelated controls. By first truncating the variants with larger P-values, it is less vulnerable to the inclusion of neutral variants compared with other statistical approaches. Moreover, because the test statistic is developed by conditioning on parental genotypes, conADA generates valid statistical inference in the presence of population stratification. We evaluate the validity and power performance of conADA and some other methods through simulations. Furthermore, a real data analysis about hypertension is presented. The validity and power of next-generation sequencing data analyses can be affected by population stratification and the inclusion of neutral variants, respectively. We recommend the conADA method for its robustness to these two factors.

1270T

Kernel-Based Association Mapping on the X Chromosome in Samples with Complex Structure. *C. McHugh, T. Thornton.* Biostatistics, Univ Washington, Seattle, WA.

As the cost of whole-genome sequencing continues to decrease, researchers are able to generate rich datasets that include both rare and common variants. Large-scale genomic studies often include individuals who are related through documented pedigrees or are cryptically related. Furthermore, genetic studies include individuals from diverse populations with complex ancestry, such as admixed populations. In recent years, kernel-based statistical association methods have been proposed for association testing with rare (and common) variants within a gene region of interest on the autosomes. Some methods allow for family structure, but only if this structure is known. A limitation of existing methods is that they are not directly applicable to association testing of regions on the X chromosome. Here, we propose a method for **KE**rnel-based **A**ssociation **T**esting in **S**tructured samples on the **X** chromosome (KEATS-X) that is valid for samples with known and/or unknown complex structure. KEATS-X uses empirical genetic relatedness matrices and variance componentsestimated fromthe autosomes and X chromosome separately to account for sample correlation. Our approach is flexible and can include other sources of correlation, such as shared environmental effects. Fixed covariates can be included to account for population stratification or other common effects such as sex or age. In simulation studies, we demonstrate that the proposed KEATS-X method has the correct type I error for gene-based testing on the X chromosome, with both rare and common variants. We also show that existing methods applied to X chromosome genetic variants are not properly calibrated and have inflated type I error for gene-based tests on the X chromosome. We illustrate the KEATS-X method by analyzing genetic variants on the X chromosome using hematological traits measured in individuals from the Hispanic Community Health Study/Study of Latinos (HCHS/SOL).

.

1271F

Family-based rare variant association analysis: a fast and efficient method of multivariate phenotype association analysis. *L. Wang[1], D. Qiao[2,5], M. Cho[2,3], E. Silverman[2,3], S. Won[1,4].* 1) Interdisciplinary Program in bioinformatics, Seoul National University, Seoul, Seoul, South Korea; 2) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA; 3) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA; 4) Graduate School of Public Health, Seoul National University, Seoul, 151-742, Korea; 5) Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA.

Motivation: Family-based designs have been repeatedly shown to be powerful in detecting the significant rare variants associated with human disease. Furthermore, human diseases are often outcomes of multiple phenotypes, and thus we expect multivariate, family-based analyses to be efficient in detecting associations with rare variants. However, few statistical methods implementing this strategy have been developed. In this report, we describe one such implementation: the multivariate family-based rare variant association tool (*mFARVAT*). **Results:** *mFARVAT* is a quasi-likelihood-based score test for the rare variant association analysis with multiple phenotypes, and tests homogeneous and heterogeneous effects of each variant on multiple phenotypes. Simulation results show that the method is generally robust and efficient for various disease models, and we identify some promising candidate genes associated with chronic obstructive pulmonary disease. **Availability and Implementation:** The software is freely available at http://healthstat. snu. ac. kr/software/mfarvat/, implemented in C++ and supported on Linux and MS Windows.

1272W

Prevalence of Allelic Heterogeneity in Genotype-Tissue Expression (GETx) data. *F. Hormozdiari[1], A. Zhu[1], G. Kichaev[2], B. Pasaniuc[2,4], E. Eskin[1,3].* 1) Department of Computer Science, UCLA, Los Angeles, CA; 2) Inter-Departmental Program in Bioinformatics,UCLA, Los Angeles; 3) Department of Human Genetics, UCLA, Los Angeles; 4) Department of Pathology and Laboratory Medicine, UCLA, Los Angeles.

Allelic heterogeneity (AH) is a phenomena where more than one variant in the same gene affects a phenotype. The Genotype-Tissue Expression project data provides an unprecedented opportunity to quantify the amount of allelic heterogeneity across tissues. The current approach for detecting theallelic heterogeneity is based on conditional analysis used in the context of fine-mapping. The conditional approach relies on an iterative selection of the most associated variants followed by re-computation of the statistical score for the remaining variants conditional on the already selected variants. The conditional process continues until statistical score for all the variants falls below a pre-defined significant threshold. As we show in our experiments, the conditional approach is sensitive to complex patterns of linkage disequilibrium and can often result in inaccurate predictions. In our work, we propose a novel probabilistic method to quantify allelic heterogeneity. Our method is an extension of our fine-mapping method, CAVIAR. The benefit for our method is that it provides a probability for different number of variants that affect the phenotype. Thus, we can assign probability to each gene which it harbors allelic heterogeneity and as a result we can have a confident level for our prediction ofgenes which harbors allelic heterogeneity. We assess the performance of our method using simulated and real datasets. Our simulation results show that our method has high confidence in detecting the allelic heterogeneity loci. In addition, we apply our method to GTEx datasets that consist of nine tissues, weobserve multiple allelic heterogeneity loci in different tissues. We considered genes which are detected as allelic heterogeneity for both methods. We observe both tissue-specific and cross-tissue allelic heterogeneity. The software for our method is freely available for download at http://genetics. cs. ucla. edu/caviar/.

1273T

EIGEN: A spectral approach for the integration of functional genomics annotations for both coding and noncoding sequence variants. *I. Ionita-Laza[1], K. McCallum[1], B. Xu[2], J. Buxbaum[3].* 1) Department of Biostatistics, Columbia University, New York, NY; 2) Department of Psychiatry, Columbia University, New York, NY; 3) Mount Sinai School of Medicine, New York, NY.

Over the past few years, substantial effort has been put into the functional annotation of variation in human genome sequence. Such annotations can play a critical role in identifying putatively causal variants among the abundant natural variation that occurs at a locus of interest. The main challenges in using these various annotations include their large numbers, and their diversity. We discuss an unsupervised approach (EIGEN) to derive an integrative score of these diverse annotations. The proposed method produces estimates of predictive accuracy for each functional annotation score, and subsequently uses these estimates of accuracy to derive the aggregate functional score for variants of interest as an optimal weighted linear combination of individual annotations. We show that the resulting meta-score has good discriminatory ability using disease associated and putatively benign variants from published studies (for both Mendelian and complex diseases), and is more strongly associated with the disease association status of such variants compared with the recently proposed CADD score. Furthermore, we show how the meta-score is particularly useful in prioritizing likely causal variants in a region of interest when it is combined with sequencing data in the framework of a hierarchical model.

1274F

denovolyzeR : an R package for the interpretation of *de novo* variation in human disease. *J. S. Ware[1,2,3,4], K. E. Samocha[2,3,4], J. Homsy[2,5], M. J. Daly[2,3,4].* 1) NIHR Cardiovascular Biomedical Research Unit at Royal Brompton Hospital and Imperial College London, London UK; 2) Department of Genetics, Harvard Medical School, Boston MA; 3) Broad Institute of MIT and Harvard, Cambridge MA; 4) Analytical and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston MA; 5) Cardiovascular Research Center, Massachusetts General Hospital, Boston MA.

Spontaneously arising (*de novo*) genetic variants are important in human disease, yet every individual carries many such variants, with a median of 1 *de novo* variant affecting the protein-coding portion of the genome. We recently described a mutational model that provides a powerful framework for the robust statistical evaluation of such coding variants. Here we present a new open-source software package, **denovolyzeR**, that implements this model and provides tools for the analysis of *de novo* coding sequence variants. Exome sequencing and analysis of *de novo* variants has identified genes underlying rare and genetically homogeneous Mendelian diseases. In Kabuki syndrome, for example, *de novo* missense variants were identified in *KMT2D* (*MLL2*) in 9 out of 10 unrelated individuals: an accumulation that is extremely improbable in the absence of a causal role in the disease. It is more challenging to dissect the role of *de novo* variants in conditions with high levels of locus heterogeneity, including heritable complex traits and some Mendelian conditions, where *de novo* variants may be spread across many genes, and may make a smaller overall contribution to pathogenesis. The interpretation of *de novo* variants is complicated by the background mutation rate, which varies greatly between genes. Additionally, as more individuals are sequenced, it is inevitable that multiple *de novo* variants will be observed in some genes by chance. Our statistical framework addresses these challenges. Briefly, the mutability of each gene is determined based on local sequence context, and the probability of a *de novo* event in a single generation is calculated. The consequence of each possible *de novo* SNV is computed, and *de novo* probabilities are tabulated by variant class. The probability of a small frameshifting indel is also estimated. For a given study population, observed variant numbers are compared against model-derived expectation using a Poisson framework. This permits robust significance estimates for the pileup of *de novo* variation in individual genes and gene sets, and increases the power of genome-wide analyses. **denovolyzeR** makes this statistical framework widely available. We demonstrate the application of this software to analyses in autism spectrum disorder and congenital heart disease, confirming the important role of *de novo* variation in the pathogenesis of these conditions, and identifying new genes and pathways associated with disease.

1275W

DISSECT: A new tool for analyzing extremely large genomic datasets. *O. Canela-Xandri[1], A. Law[1], A. Gray[2], J. Woolliams[1], A. Tenesa[1,3].* 1) The Roslin Institute. University of Edinburgh, Edinburgh, United Kingdom; 2) EPCC. University of Edinburgh, Edinburgh, United Kingdom; 3) MRC HGU at the MRC IGMM. University of Edinburgh, Edinburgh, United Kingdom.

Computational tools are quickly becoming the main bottleneck to analyze large-scale genomic and genetic data. This big-data problem, affecting a wide range of fields, is becoming more acute with the fast increase of data available. To address this, we developed DISSECT, a new, easy to use, and freely available software able to exploit the parallel computer architectures of supercomputers to perform a wide range of genomic and epidemiologic analyses which currently can only be carried out on reduced sample sizes or in restricted conditions. We showcased our scalable tool by addressing the challenge of predicting phenotypes from genotype data in human populations using Mixed Linear Model analysis. We analyzed simulated traits from half a million individuals genotyped for 590,004 SNPs using the combined computational power of 8,400 processor cores. We found that prediction accuracies in excess of 80% of the theoretical maximum could be achieved with large numbers of training individuals.

1276T

Cost-effective omnibus meta-analysis of massive studies. *H. Qin[1,2], W. Ouyang[1,2], SL. Cao[3], Y. Zhu[4].* 1) Center for Bioinformatics and Genomics; 2) Department of Biostatistics and Bioinformatics; 3) Department of Biomedical Engineering; 4) Department of Epidemiology,Tulane University, New Orleans, LA.

In this report, we propose a general framework for integrating massive independent pieces of evidence. Infimum omnibus statistics are constructed by integrating the *p*-values from separate studies. The proposed framework allows for both continuous and categorical traits as well as heterogeneity, population structure, and cryptic relatedness. As mathematically proven, the exact null distributions of the integrated statistics are invariant to the number of studies. Thus, the proposed framework suffers no penalty in degrees of freedom due to the increase in the number of studies involved. We have developed a high-performance computational package to accurately compute the *p*-value of the infimum omnibus statistics. Extensive numerical results demonstrated that the proposed framework has improved statistical power over conventional meta-analysis methods, especially when a large number of separate studies are involved. Our analytical results and computational tool would benefit researchers to maximize statistical power by integrating valuable pieces of information and minimize the harms caused by incautiously combining pure noises in large-scale meta-analyses.

1277F

Phasing using rare variants and large haplotype reference panels. *K. Sharp[1,3], O. Delaneau[2], J. Marchini[1,3].* 1) Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK; 2) Département de Génétique et Développement (GEDEV), University of Geneva, Switzerland; 3) The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK.

Estimation of haplotypes from genotypes, known as phasing, is a central part of the pipeline of many modern genetic analyses. Such haplotypes are often estimated as an imperfect mosaic of high quality haplotypes (copying states) from a reference panel. Increasingly, both reference panels and unphased genotypes are derived from sequencing data and include large numbers of low frequency variants that are much harder to phase. Consequently, selection of the most informative set of copying states in a computationally tractable way is becoming a fundamental problem for the field. Despite their challenges, rare variants can also be phase informative: sharing of such rare variants between two individuals is more likely to arise from a recent common ancestor and, hence, also more likely to indicate similar shared haplotypes. Our method exploits this idea to select a small set of highly informative copying states. When combined with SHAPEIT2, we obtain significant gains in phasing accuracy over the current selection approach, as well as an improvement in speed. We tested our method using 7,510 UK10K haplotypes. We phased two regions of chromosome 20, comprising 48MB, for 45 individuals from this dataset using the remaining haplotypes as a reference panel. Our method of choosing states consistently improved accuracy compared to the current version of SHAPEIT2 and showed less variation over different runs. Averaged over individuals, regions phased and 20 different runs, our method achieved a 33% improvement in switch error rate. In addition, we phased the same regions for two high coverage (130x) trio parents of European ancestry. Our new version of SHAPEIT permits the rare variant haplotype selection method to be used in combination with phase-informative paired end reads. Combining these sources of information led to a total reduction in switch error rates of 61. 4% when averaged over both trio parents and 20 runs.

1278W

Minimac4: A next generation Imputation Tool for Mega Reference Panels. *S. Das, G. Abecasis, C. Fuchsberger.* Department of Biostatistics, University of Michigan, Ann Arbor, MI.

Genotype Imputation is a key component in genetic association studies increasing power, facilitating meta-analytic and fine-mapping efforts, and aiding in interpretation of signals. Statistical methods for imputation have developed immensely over the last few years enabling researchers to draw benefits from increasingly large sequencing-based reference panels. Modern reference panels like the Haplotype Reference Consortium (HRC) with over 32,000 whole genome sequenced samples can boost the imputation accuracy of causal variants in the lower end of the allele frequency spectrum. However, computational resources needed to impute study samples from these panels are still quite demanding and not feasibly available to all researchers. In 2015, we implemented a web imputation server that allows users to impute study samples using the HRC reference panel without having direct access to the data. In this study, we devise a substantially faster computation tool for imputation on the server that will allow an imputation throughput of 10,000s of genomes a day using the HRC reference panel. To attain that, we developed a novel method for reducing the state space by collapsing haplotypes that are identical only at positions that were genotyped in the study sample. Involving a complex implementation of hidden markov models which tracks “fuzzy” alleles at intervening sites, this additional round of collapsing produces significant cost savings. We are also exploring approximate versions of this model that may slightly decrease imputation accuracy (compared to current approaches) but that will scale to reference panels with 100,000s of samples. The ability of these new methods to handle much larger reference panels should compensate for any minor decrease in imputation accuracy compared to current methods that cannot feasibly scale to very large samples.

1279T

Beyond Homozygosity Mapping: Family-Control analysis based on Hamming distance for prioritizing variants in exome sequencing. *A. Imai[1,2], A. Nakaya[2], S. Fahiminiya[3], M. Tétreault[3], J. Majewski[3], Y. Sakata[1], S. Takashima[1,4], M. Lathrop[3], J. Ott[5,6].* 1) Department of Cardiovascular Medicine, Osaka University Graduate School of Medicine, Suita, Japan; 2) Department of Genome Informatics, Osaka University Graduate School of Medicine, Osaka, Japan; 3) McGill University and Genome Québec Innovation Centre, Montréal, Canada; 4) Department of Medical Biochemistry, Osaka University Graduate School of Medicine, Osaka, Japan; 5) Institute of Psychology, Chinese Academy of Sciences, Beijing, China; 6) Laboratory of Statistical Genetics, Rockefeller University, New York, USA.

A major challenge in current exome sequencing in autosomal recessive (AR) families is the lack of an effective method to prioritize single-nucleotide variants (SNVs). AR families are generally too small for linkage analysis, and length of homozygous regions is unreliable for identification of causative variants. Various common filtering steps usually result in a list of candidate variants that cannot be narrowed down further or ranked. To prioritize shortlisted SNVs we consider each homozygous candidate variant together with a set of SNVs flanking it. We compare the resulting array of genotypes between an affected family member and a number of control individuals and argue that, in a family, differences between family member and controls should be larger for a pathogenic variant and SNVs flanking it than for a random variant. We assess differences between arrays in two individuals by the Hamming distance and develop a suitable test statistic, which is expected to be large for a causative variant and flanking SNVs. As we are unsure of the optimal distance from a candidate variant within which to consider flanking SNVs, we maximize the primary test statistic over a wide range of such distances and prioritize candidate variants based on this maximum statistic. We applied our approach to six patients with known pathogenic variants and found these to be in the top 2 to 10 percentiles of ranks. While homozygosity mapping requires various parameters to be fixed at the outset, our approach works with a single test statistic estimated by the data.

1280F

Web application for statistical power calculations in genotype-based recall randomized controlled trials. *N. Atabaki Pasdar[1], A. Ali[1], M. Ohlsson[2], D. Shungin[1], T. V. Varga[1], A. Kurbasic[1], E. Ingelsson[3], E. R Pearson[4], P. W. Franks[1,5,6].* 1) Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmö, Sweden; 2) Department of Astronomy and Theoretical Physics, Computational Biology and Biological Physics Unit, Lund University, Lund, Sweden; 3) Department of Medical Sciences, Molecular Epidemiology, Uppsala University, Uppsala, Sweden; 4) Division of Cardiovascular & Diabetes Medicine, Medical Research Institute, University of Dundee, Dundee, UK; 5) Department of Public Health & Clinical Medicine, Umeå University, Umeå, Sweden; 6) Department of Nutrition, Harvard School of Public Health, Boston, MA, USA.

Validation of gene ×environment interaction results from epidemiological studies in randomized clinical trials (RCTs) is usually hampered by insufficient statistical power. In this framework we test if a genotype-based recall (GBR) approach of recruiting two groups of participants with distinct genetic profiles can yield sufficiently higher statistical power with reduced sample sizes, leading to a reduction of experimental costs, compared with conventional RCTs of comparable sample size. To this end, we modelled GBR sampling, with participants chosen from the two extremes of a genetic risk score (GRS) distribution, and compared this with conventional sampling. We performed power calculations using simulations in the R program using assumptions from the Diabetes Prevention Program. We have calculated the required sample sizes to reach sufficient statistical power when analyzing the interaction between a genetic factor and intensive lifestyle intervention in a linear regression model, with 1-year small low-density lipoprotein (LDL) particles level as the outcome. Similarly, the statistical power for the interaction effect between a genetic factor and metformin treatment was simulated in a Cox proportional hazards regression model, with time to developing type 2 diabetes as the outcome. Statistical power in various scenarios including different effect sizes, allele frequencies, initial sampling frames and error rates were also examined. Results from almost all simulations confirm that GBR is more powerful than conventional sampling to detect gene ×environment interactions in both linear and Cox regression models. For instance, in the linear regression with a GRS comprised of common SNPs conveying large effects, the sample sizes required to achieve 80% power in GBR and conventional sampling are ~100 and ~800, respectively. With a GRS comprised of common frequency SNPs that convey small effects, these sample sizes are ~1000 and ~3900, respectively. With a GRS comprised of low-moderate frequency SNPs that convey large effects, these sample sizes are ~300 and ~1600, respectively. With a GRS comprised of low-moderate frequency SNPs with small effects, these sample sizes are ~2700 and ~7000, respectively. We additionally have developed an online application for statistical power calculations for gene-treatment interactions for both linear and Cox regression models and made available to the community to facilitate the design of genotype-based recall clinical trials.

1281W

Structural equation modeling of tuberculosis immune response using genetic data and pedigree structure. *N. B. Hall[1], R. P. Igo[1], W. H. Boom[2,3], M. Joloba[4,5], E. Wampande[5,6], N. J. Morris[1], C. M. Stein[1,7].* 1) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Division of Infectious Diseases, Case Western Reserve University, Cleveland, OH; 3) Department of Medicine, Case Western Reserve University, Cleveland, OH; 4) Uganda-CWRU Research Collaboration, Kampala, Uganda; 5) College of Health Sciences Makerere University and Mulago Hospital, Kampala, Uganda; 6) College of Veterinary Medicine, Animal resource and Bio-security, Kampala, Uganda; 7) Tuberculosis Research Unit, Case Western Reserve University, Cleveland, OH.

Structuralequation modeling (SEM) is a hypothesis-driven approach to the analysis of multifactorial phenotypes. SEM involves the estimation of multiple simultaneous equations using the data's covariance structure. Models represent biological and epidemiological relationships among latent and observed variables, using structural and measurement equations to estimate them. This multivariate approach models complex relationships between variables with multiple correlated variables. Previous methods using SEM to model data have not been suited to the use of family data with both measurement and structural equations. Our objective was to use a novel approach to SEM to include both genetic and pedigree data to identify factors associated with tuberculosis (TB) immune response. Strum is a user-friendly R package which implements a SEM framework that incorporates pedigree information, facilitating the use of family data, as developed by Morris and colleagues. This framework allows for latent measures and covariates within a structural model. It can perform genetic association and linkage analysis, and estimate polygenic effects. We demonstrate the utility of this method using data gathered through a household contact study of tuberculosis (TB) in Kampala, Uganda. In this analysis, 1407 individuals from this study were used to model interferon-γ (IFNg) response to Mycobacterium tuberculosis (Mtb) as a latent variable,. SNPs from 8 candidate genes, the 3 main lineages of Mtb found in Uganda, and production levels of IFNg in response to various Mtb-specific antigens were included in the model. Human genotype data came from a custom Illumina panel as well as an Omni5 chip, and Mtb lineage was determined using phylogenetic analysis using 8 SNPs genotyped. Strum was used to comparatively fit models to identify SNPs and Mtb lineages independently associated with IFNg response. The most parsimonious model had a theoretically corrected chi-square parameter with p-value =0. 69, indicating a good fit. A significant association was seen between IFNg response and a SNP in*TICAM2*(p-value = 0. 007), which we previously identified as associated with TB, but none of the Mtb lineages were associated with IFNg response. Antigens ESAT6 (0. 012) and AG85B (0. 006) significantly captured IFNg responsiveness. By revealing the independent effects of a human gene and Mtb-specific antigens, this demonstrates the value of SEM to examine multiple biological risk factors simultaneously.

1282T

FastPop: a rapid principle component derived method to infer intercontinental ancestry using genetic data. *Y. Li[1], J. Byun[1], G. Cai[2], X. Xiao[1], J. Dennis[3], D. Easton[3], I. Gorlov[1], M. Seldin[4], C. Amos[1].* 1) Department of Biomedical Data Science, Dartmouth College, Hanover, NH; 2) Department of Genetics, Dartmouth College, Hanover, NH; 3) Centre for Cancer Genetic Epidemiology, Cambridge University, Cambridge, UK; 4) Rowe Program in Genetics, U. C. Davis, Davis, CA.

First three and last two authors contributed equally**Summary:**PCA has become a standard procedure in population genetics study for substructure analysis. The eigenvectors from PCA are easy to use for population adjustment in GWA studies. However it lacks the ability to provide clear information for ancestral origin, and usually does not yield an approach that can be generalized from study to study. To fill this gap, we developed FastPop, an efficient R packagefor inference of ancestry with PCA scores as the input. FastPop will first compute scores for individuals based on eigenvectors from PCA analysis, and then estimate the proportional ancestry of each individual based on the scores. We demonstrate the use of our software using markers that are present on the Core Content of Illumina products. We selected 2318 SNPs across the whole genome based on having a large fixation index value among European, African and Asian populations for PCA analysis. We conducted PCA analysis of 505 Hapmap samples with European, African or Asian ancestry along with a collection of 19661 additional samples of unknown ancestry. We also ran Structure program on the same dataset to benchmark the performance of FastPop. The results from FastPop are highly consistent with that from structure across the 19661 samples. The correlations of the results between FastPop and structure are 0. 99, 0. 97 and 0. 99 for European, African and Asian, respectively. Compared with Structure, FastPop is more efficient as it can finish ancestry inference for 19661 samples in 16 minutes compared with 21-24 hours required by structure. FastPop can also provide PCA scores based on SNP weights so the scores of reference population such as Hapmap samples can be applied to other studies providing same set of markers are used.

Availability and Implementation: FastPop is an R package that can be easily installed and run. It has three functions:function “PCAScore” enables the user to compute scores based on eigenvectors from PCA analysis given by 2318 SNP weights;function “InterContinentalDistanceMetrics” calculates the proportion of each continental ethnicity from scores of individuals and this function also works for inference using eigenvectors as input directly; function “PCAScore_plot” helps to visualize and compare new scores with 505 Hapmap samples chosen to indicate each centroid on three Continental definitionsThe package is freely available at https://sourceforge. net/projects/fastpop/files/.

1283F

Individualized Coherent Absolute Risk Estimator (iCARE): A Flexible Tool for Absolute Risk Estimation Using Genetic and Environmental Risk Factors. *P. Maas, N. Chatterjee.* National Cancer Institute, Division of Cancer Epidemiology and Genetics, Biostatistics Branch, Rockville, MD.

Background

Large genome-wide association studies have led to the discovery of many low penetrant, common single nucleotide polymorphisms (SNPs) associated with risk of complex chronic diseases such as breast cancer, type-2 diabetes and coronary heart disease. The utility of these SNPs, in combination with other risk factors, for public health strategies of disease prevention needs investigation. Evaluation of absolute risk, as opposed to relative risk which is typically used for summarizing associations, is fundamental for these public health applications.

Methods

We present a new tool for building absolute risk models and estimating absolute risk based on SNP and risk factor profiles for individual subjects. Application of the tool requires users to specify the disease incidence rate in the underlying population, a model for relative risk and the associated parameters specifying the joint effects of the risk factors, and a reference sample of individual-level data for the joint distribution of the risk factors in an underlying population. As output, the tool produces absolute risk estimates over specified age intervals. If SNP or risk factor information is missing for an individual subject, the tool produces an appropriate model-averaged risk with respect to the reference distribution. We discuss the methodology underlying the absolute risk calculation and the tool’s handling of missing data.

Applications

We illustrate the utility of the tool by applying it to build a model for breast cancer risk prediction, incorporating information on a total of 91 SNPs and 10 risk factors. This application highlights the value of evaluating absolute risk, and particularly risk stratification, in a public health context.

Conclusion

The iCARE tool provides researchers and clinicians a flexible way to rapidly translate etiologic knowledge, including genetic susceptibility, into an understanding of risk and a guide for risk reduction interventions.

1284W

Doing better than best in linear mixed model association testing. *J. A. Mefford, J. Witte, N. Zaitlen.* University of California, San Francisco, San Francisco, CA.

Linear mixed models (LMM) are used in GWAS to improve power for detecting associations and to control for confounding by population structure. The power gains arise by accounting for the ancillary contributions of other genetic features to the phenotype of interest. We maximize such gains with a novel leave-one-out cross validation approach to generate “better”, cross-validated best linear unbiased predictions (cvBLUPs). Existing LMM-GWAS approaches estimate variance components and test associations for individual SNPs using a generalized least squares formulation. Another use of a fitted LMM is the generation of BLUPs--estimates of the realized values of random effects for individuals. Existing LMM approaches over-fit to the available data, causing the BLUPs to be correlated with un-modeled contributions to the phenotype, including the effects of test-SNPs. To address this, we generate cvBLUPs, which we show are independent of un-modeled factors. Using these cross-validated genetic predictions in association tests allows unbiased estimates of the test-SNP effect size, even in analyses with multiple variance components. In addition to providing the first method for unbiased estimation with multiple variance components, we leverage these predictions to improve power. We use cross-validated estimates of the random genetic effects to weight the contributions of features to the relatedness matrix, update the covariance structure, and calculate re-weighted cross-validated predictors (wcv-BLUPs). This improves power by allowing the model to leverage the improved effect size estimates from cvBLUP. To evaluate our method we examine performance on simulated and real data sets. We first simulate GWAS under different settings of number of subjects, genetic markers, fraction of causal markers, and population structure. The wcv-BLUP based tests have greater power than standard LMM approaches (EMMAX, GCTA, GEMMA), obtaining up to a 30% increase an effective sample size. We then apply our methods to gene expression data from the GEUVADIS cohort. We use two variance components to capture the genetic and expression covariances with two sets of cvBLUPs, one for each variance component. cvBLUP cuts spurious correlation of naïve BLUPs with the phenotype, from 96 to 11, for cvBLUPs based on trans-SNPs in an analysis of gene expression. Additional features, such as non-infinitesimal models (BOLT-LMM) and leverage of data from external cohorts (LDPred) are trivially included.

1285T

Analyzing case-parent trio data with the R package trio. *H. Schwender[1], Q. Li[2], C. Neumann[3], M. A. Taub[4], S. G. Younkin[5], P. Berger[1], R. B. Scharpf[6], T. H. Beaty[7], I. Ruczinski[4].* 1) Mathematical Institute, Heinrich Heine University, Duesseldorf, Germany; 2) Inherited Disease Research Branch, National Human Genome Research Institute, Baltimore, MD, USA; 3) Faculty of Statistics, TU Dortmund University, Dortmund, Germany; 4) Department of Biostatistics, Johns Hopkins University, Baltimore, MD, USA; 5) Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, USA; 6) Department of Oncology, Johns Hopkins University, Baltimore, MD, USA; 7) Department of Epidemiology, Johns Hopkins University, Baltimore, MD, USA.

Case-parent trio designs are frequently employed in genome-wide association studies to detect SNPs associated with disease. The most popular statistical tests in this study design are transmission/disequilbrium tests (TDTs), considering either the alleles or the genotypes as units in the analysis. While the genotypic TDT can be extended to the test for both gene-environment interactions and interactions between pairs of SNPs, higher-order interactions require procedures such as trio logic regression and trioFS that more cleverly search for association with disease without evaluating all possible interactions. All these procedures are implemented in an R package called trio as user-friendly functions enabling efficient and fast association analysis. Besides these functions, trio also contains functionalities for, e. g. , simulating trio data, estimating sample size and power, dealing with ped and vcf files, and estimating measures of linkage disequilibrium as well as LD-blocks. In this presentation, we exemplify the usage of the functions in trio, which is freely available at http://www. bioconductor. org, by reanalyzing data from the International Cleft Consortium comprising genotypes from about 2000 children with different types of clefts.

1286F

A new method for ancestry specific association mapping in admixed populations (asaMap). *L. Skotte[1], T. S. Korneliussen[2], H. O. H. Sirelkhatim[3], I. Moltke[3], A. Albrechtsen[3].* 1) Department of Epidemiology Research, Statens Serum Institut, 2300 Copenhagen, Denmark; 2) Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark; 3) The Bioinformatics Centre, Department of Biology, University of Copenhagen, 2200 Copenhagen, Denmark.

We introduce a new statistical method (asamap) for association mapping in recently admixed populations. Unlike most current methods for association mapping, asamap does not assume that the effect of the tested variant is independent of the local ancestry of the allele, but instead asamap estimates ancestry specific effect sizes. The new method does not require exact knowledge of the local ancestry of each allele, which can be difficult to infer without uncertainties, yet simulations show a dramatic increase in statistical power to detect associations in some scenarios (e. g. an increase in statistical power from 15% to 79% for a SNP with an effect size of 0. 3 standard deviation in one ancestral population and no effect in the other population, based on 2500 samples with admixture proportions uniformly sampled in {0, 0. 25, 0. 5, 0. 75, 1} and a p-value cutoff of 10e-8). In addition, asamap allows testing for a difference in effect size between ancestral populations which may indicate that a SNP is non-causal. We demonstrate the usefulness of asamap on data from the Greenlandic population.

1287W

Assessing mitochondrial DNA variation and copy number in lymphocytes of ~2,000 Sardinians using tailored sequencing analysis tools. *J. Ding[1], C. Sidore[2,3], T. J. Butler[1], M. K. Wing[2], Y. Qian[1], O. Meirelles[1], F. Busonero[2,3], L. C. Tsoi[2], A. Maschio[2,3], A. Angius[3], H. M. Kang[2], R. Nagaraja[1], F. Cucca[3], G. R. Abecasis[2], D. Schlessinger[1].* 1) Laboratory of Genetics, National Institute on Aging, Baltimore, MD; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 3) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Monserrato, Cagliari, Italy.

DNA sequencing identifies common and rare genetic variants for association studies, but studies typically focus on variants in nuclear DNA and ignore the mitochondrial genome. In fact, analyzing variants in mitochondrial DNA (mtDNA) sequences presents special problems, which we resolve here with a general solution for the analysis of mtDNA in next-generation sequencing studies. The new program package comprises 1) an algorithm designed to identify mtDNA variants (i. e. , homoplasmies and heteroplasmies), incorporating sequencing error rates at each base in a likelihood calculation and allowing allele fractions at a variant site to differ across individuals; and 2) an estimation of mtDNA copy number in a cell directly from whole-genome sequencing data. We also apply the methods to DNA sequence from lymphocytes of ~2,000 SardiNIA Project participants. As expected, mothers and offspring share all homoplasmies but a lesser proportion of heteroplasmies. Both homoplasmies and heteroplasmies show 5-fold higher transition/transversion ratios than variants in nuclear DNA. Also, heteroplasmy increases with age, though on average only ~1 heteroplasmy reaches the 4% level between ages 20 and 90. In addition, we find that mtDNA copy number averages ~110 copies/lymphocyte and is ~54% heritable, implying substantial genetic regulation of the level of mtDNA. Copy numbers also decrease modestly but significantly with age, and females on average have significantly more copies than males. The mtDNA copy numbers are significantly associated with waist circumference (p-value = 0. 0031) and waist-hip ratio (p-value = 2. 4×10-5), but not with body mass index, indicating an association with central fat distribution. To our knowledge, this is the largest population analysis to date of mtDNA dynamics, revealing the age-imposed increase in heteroplasmy, the relatively high heritability of copy number, and the association of copy number with metabolic traits.

1288T

PODKAT: a software package implementing the position-dependent kernel association test. *U. Bodenhofer, S. Hochreiter.* Institute of Bioinformatics, Johannes Kepler University, Linz, Austria.

High-throughput sequencing technologies have facilitated the identification of large numbers of single-nucleotide variations (SNVs), many of which have already been proven to be associated with diseases or other complex traits. Several large sequencing studies, such as, the 1000 Genomes Project, the UK10K project, or the NHLBI-Exome Sequencing Project, have consistently reported a large proportion of private SNVs, that is, variants that are unique to a family or even a single individual. The role of private SNVs in diseases is poorly understood, largely due to the fact that it is statistically very challenging to consider private SNVs in association testing. While it is generally impossible to make use of private SNVs in single-marker tests or in correlation-based tests like the popular SNP-set (Sequence) Kernel Association Test (SKAT), also burden tests are facing serious statistical issues.

We have proposed the *Position-Dependent Kernel Association Test,* which is designed for detecting associations of very rare and private SNVs with the trait under consideration even if the burden scores are not correlated with the trait. The test assumes that, the closer two SNVs are on the genome, the more likely they have similar effects on the trait under consideration. This assumption is fulfilled as long as deleterious, neutral, and protective variants are grouped sufficiently well along the genome.

This contribution highlights a recently released software package, PODKAT, that implements the position-dependent kernel association test along with the popular SKAT test and all necessary tools for defining regions of interest, multiple testing correction, filtering, and visualization of results. The package is based on the widely used scientific computing platform R. It is publicly and freely available via the Bioconductor project. It is able to read data directly from VCF files and facilitates easy parallelization on multi-processor systems or computing clusters. Due to the special memory management strategies, analyses of large studies can even be performed on regular desktop computers (at the cost of much longer computation times). We will present the PODKAT package in detail along with examples of analyses performed on the real-world data sets, such as, the UK10K whole-genome cohorts.

1289F

An extension of Conditional Inference Forest methodology for predictive biomarkers and personalized medicine applications. *B. Dizier[1,2], K. van Steen[1,2].* 1) Systems and Modeling Unit, Department of Electrical Engineering and Computer Science (Montefiore Institute), University of Liege, Quartier Polytech 1, Allee de la Decouverte 10, 4000 Liege, Belgium; 2) Systems Biology and Chemical Biology, GIGA-R, University of Liege, Quartier Hopital, Avenue de l'Hopital 11, 4000 Liege, Belgium.

Heterogeneity of treatment effect is a major hurdle for new drug development. When treatment effect is heterogeneous, average treatment effect estimation of new drug treatment effect is biased and requires very large trials to demonstrate a significant effect. This increases costs, time and risk of drug development. There is big hope that personalized medicine would bring solutions to this issue by enabling identification of patients more likely to benefit from the new drug to demonstrate efficacy. Generating actionable information from all the data generated by high-throughput technologies requires robust methodologies to avoid the pitfalls of exhaustive subgroup analysis. Recursive partitioning is a popular non-parametric modelling technique for regression and classification problems. Conditional Inference Forest is an appealing implementation of recursive partitioning that provides control of type I error during tree growth and unbiased variable selection. Here, we present a novel machine learning approach based on recursive partitioning for estimation of individual treatment benefit through counterfactual framework in time to event outcomes. The estimate of individual treatment benefit is used to assess heterogeneity of treatment effect and potential presence of subgroups that may derive benefit/harm from the new drug. Conditional on the presence of treatment effect heterogeneity, predictive biomarkers that could potentially become companion diagnostics for patient stratification or help understanding drug mode of action are identified through variable importance. The methodology allows for prognostic index or propensity score adjustments in case of strong prognostic effects (independent of treatment) and deviation from randomization between treatment arms. The performance of the novel methodology is further evaluated and validated via application to synthetic and real-life data.

1290W

Axiom™ HLA Analysis – a tool to impute HLA types. *Y. Lu, C. Patel, J. Danzer, M. Dong, M. McCrumb, R. Griffis, N. Hashmi, T. Hoang, J. Burrill, T. Webster, J. Schmidt.* Affymetrix Inc, Santa Clara, CA.

Thehuman leukocyte antigen (HLA) gene family encodes the human version of the major histocompatibility complex (MHC)- cell-surface proteins that regulate the humanimmune system. These genes play various roles in autoimmune/infectious diseases, severe drug responses, and transplant rejection, which are the focuses of many association studies. However, the highly polymorphic nature of this region and the prevalence of pseudogenes make it a very challenging task to accurately genotype or sequence the different HLA alleles with traditional approaches. Axiom™ HLA Analysis is an application designed for automated high-resolution HLA typing of Axiom® genotyping array data. Axiom HLA Analysis can be used for SNP genotypes from any platform, as long as the genotyping platform has sufficient density in the MHC region and provides high-accuracy genotypes. It uses a multi-population reference panel and an HLA type imputation model, HLA*IMP:02 (Alexander Dilthey, Stephen Leslie, *et al.*) to utilize the long-rang linkage disequilibrium between SNPs and HLA alleles to statistically infer HLA types. The direct SNP genotypes are used to impute HLA types for the three MHC class I genes (A, B, and C) and eight MHC class II genes (DPA1, DPB1, DQA1, DQB1, DRB1, DRB3, DRB4, DRB5). The integration of HLA typing with whole-genome data produced by high-density Axiom® genotyping arrays for GWAS enables a novel combination of direct and hypothesis-free discovery of immune system variation associations. We show that Axiom HLA Analysis produces accurate 2- and 4-digit resolution across multiple ethnic groups, when compared with HLA types produced by alternate technologies, and works robustly even when the number of directly genotyped SNPs in the MHC region is significantly lower than those present on Axiom arrays.

1291T

Tag SNP selection for low frequency variant imputation in populations of diverse ancestry. *G. L. Wojcik[1], C. R. Gignoux[1], C. Fuchsberger[2], D. Taliun[2], R. Welch[2], A. R. Martin[1], H. R. Johnston[3], S. Shringarpure[1], C. Pethiyagoda[7], J. O'Connell[7], L. McAuliffe[7], Z. S. Qin[3], K. C. Barnes[5], G. Abecasis[2], C. S. Carlson[4], H. M. Kang[2], M. Boehnke[2], C. D. Bustamante[1], E. E. Kenny[6].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor MI; 3) Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta GA; 4) Fred Hutchinson Cancer Research Center, Seattle WA; 5) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore MD; 6) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York NY; 7) Illumina, Inc, 5200 Illumina Way, San Diego CA.

The emergence of large sequenced reference panels in the past year has facilitated a new focus on accurate imputation of low frequency variants (LFV; 0. 005-0. 05 MAF). In tandem, a new generation of genotyping arrays are being developed targeting tag SNPs that will enrich for imputation at the rarer end of the frequency spectrum. Selection of these tag SNPs poses several challenges as LFVs tend to be continentally or even population specific, with fine-scale LD structure reflecting recent demographic events. To address these challenges, we have developed a novel algorithm to select tag SNPs considering both population-specific and trans-ethnic tagging performance, that maximizes imputation accuracy rather than pairwise coverage. This is achieved via a leave-one-out internal validation approach that allows direct comparison of tag SNP performance for each iteration of potential scaffold sites. This method was recently applied to boost diverse population coverage for the Multi-ethnic Genotyping Array (MEGA), a collaboration between Illumina and multiple consortia (PAGE, CAAPA, and T2D Genes), by leveraging the whole genome sequences available in the 1000 Genomes Project Phase 3 (TGP) release. We will present results that explore the effect of minimum r2 and minor allele frequency (MAF) threshold for tag SNP selection using various methods of prioritization across the 26 TGP populations. By prioritizing tags that contribute information across multiple populations, this method provides enhanced imputation accuracy compared to naive selection, for example, improving imputation accuracy by 9% for 0. 01-0. 02 MAF SNPs in East Asian populations. Consistent with demographic history, African ancestry populations contributed the largest proportion of tags informative to all populations, but required the densest number of tag SNPs to maximize LFV imputation. However, recent explosive growth in non-African populations means tag SNPs capture on average 30% less variation than in African populations [9. 8 SNPs/tag in East Asians and 13. 8 SNPs/tag in Africans]. When applied to the 1. 5M GWAS scaffold on MEGA, imputation accuracy was found to be ≥0. 89 for SNPs with MAF≥0. 005 and ≥0. 95 for MAF≥0. 05 in all 6 TGP continental populations. This unified framework for tag SNP selection and imputation evaluation will be useful for designing reference panels, large multi-ethnic epidemiological studies and biobanks, as well as future biological repositories.

1292F

A unifying method for multiple phenotype, multiple variance component mixed models. *A. Dahl, J. Marchini.* Department of Statistics, University of Oxford, Oxford, United Kingdom.

Mixed modelshaverecently re-emerged as a state-of-the-art approach to control forrelatedness, populationstructure andgenome-widepolygenicityinhumanGWAS. Moreover, mixed models are becoming even more prominent as the field moves beyond this standard analysis, as they can be broadly used to decompose phenotypic variation into biologically meaningful components. Recently, single phenotype, single random effectmixed modelshave been generalized in two distinct directions. First,single phenotypemixed models with multiple random effects have been applied to partition heritability amongst functionally meaningful categories, to test gene-based association and to improve phenotype and breeding value estimation. Second, mixed models with multiple phenotypesand one random effect have beenused to learn heritable phenotype networks, boost power in association studies and impute missing phenotype data.

We have generalized these approaches by developing the first efficient method to simultaneously model multiple phenotypes and multiple variance components. Furthermore, our method can utilize a wide range of likelihood penalty functions, providing additional statistical robustness, computational efficiency and biological interpretability. Even without penalization, our method can be run on tens of thousands of individuals, hundreds of variance components and dozens of phenotypes.

To demonstrate one possible application of our method, we performed a gene-based association test with multiple phenotypes whilecorrecting for confounding structure. Specifically, we found that the*FADS1*gene significantly associates with LDL and triglyceride traits in 5,163 samples from the NFBC19666 study (-log10(p)=8. 19). We willalsoillustrate the performance of this method on simulateddata.

1293W

MEGSA: A powerful and flexible framework for analyzing mutual exclusivity of tumor mutations. *X. Hua[1], P. L. Hyland[2], J. Huang[3], B. Zhu[1], N. E. Caporaso[2], M. T. Landi[2], N. Chatterjee[1], J. Shi[1].* 1) Biostatistics Branch, Division of Cancer Epidemiology & Genetics, National Cancer institute, National Institute of Health, Bethesda, MD; 2) Genetic Epidemiology Branch, Division of Cancer Epidemiology & Genetics, National Cancer institute, National Institute of Health, Bethesda, MD; 3) Laboratory of Cancer Biology and Genetics, Center for Cancer Research, National Cancer Institute, National Institute of Health, Bethesda, MD.

The central challenge in tumor sequencing studies is to identify driver genes and pathways, investigate their functional relationships and nominate drug targets. The efficiency of these analyses, particularly for infrequently mutated genes, is compromised when patients carry different combinations of driver mutations. Mutual exclusivity analysis helps address these challenges. To identify mutually exclusive gene sets (MEGS), we developed a powerful and flexible analytic framework based on a likelihood ratio test and a model selection procedure. Extensive simulations demonstrated that our method outperformed existing methods for both statistical power and the capability of identifying the exact MEGS, particularly for highly imbalanced MEGS. Our method can be used for *de novo* discovery, pathway-guided searches or for expanding established small MEGS. We applied our method to the whole exome sequencing data for fourteen cancer types from The Cancer Genome Atlas (TCGA). We identified multiple previously unreported non-pairwise MEGS in multiple cancer types. For acute myeloid leukemia, we identified a novel MEGS with five genes (*FLT3*, *IDH2*, *NRAS*, *KIT* and *TP53*) and a MEGS (*NPM1*, *TP53* and *RUX1*) whose mutation status was strongly associated with survival (*P*=6. 7×10-4). For breast cancer, we identified a significant MEGS consisting of *TP53* and four infrequently mutated genes (*ARID1A*, *AKT1*, *MED23* and *TBL1XR1*), providing support for their role as cancer drivers.

1294T

An improvement of the *funNorm* normalization method for methylation data from multiple cell or tissue types: *funtooNorm*. *C. M. T. Greenwood[1,2,13,17,18], K. Oros Klein[1,17], S. Grinek[1,17], S. Bernatsky[11], L. Bouchard[3,4], A. Ciampi[2], I. Colmegna[11], J. -P. Fortin[5], M. -F. Hivert[6,7], M. Hudson[1,10], M. S. Kobor[14,16], A. Labbe[1,8], J. L. MacIsaac[14], M. J. Meaney[8,9,14,15,17], A. M. Morin[14], K. J. O'Donnell[7], T. Pastinen[13], M. H. van Ijzendoorn[12], G. Voisin[1,17].* 1) Lady Davis Institute, Jewish General Hospital,, Montreal, QC, Canada; 2) Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, QC, Canada; 3) Centre de recherche du CHUS, Centre de santé et des services sociaux de Chicoutimi, Chicoutimi, QC, Canada; 4) Department of Biochemistry, Université de Sherbrooke, Sherbrooke, QC, Canada; 5) Department of Biostatistics, Johns Hopkins University, Baltimore, MD, USA; 6) Department of Medicine, Division of Endocrinology, Université de Sherbrooke, Sherbrooke, QC, Canada; 7) General Medicine Division, Massachusetts General Hospital, Boston, MA, USA; 8) Douglas Mental Health University Institute, McGill University, Montreal, QC, Canada; 9) Departments of Psychiatry and Neurology & Neurosurgey, McGill University, Montreal, QC, Canada; 10) Department of Medicine, McGill University, Montreal, QC, Canada; 11) McGill University Health Centre, McGill University, Montreal, QC, Canada; 12) Centre for Child and Family Studies, Leiden University, Leiden, The Netherlands; 13) Department of Human Genetics, McGill University, Montreal, QC, Canada; 14) Canadian Institute for Advanced Research, Child and Brain Development Program, Toronto, ON, Canada; 15) Singapore Institute for Clinical Sciences, Singapore; 16) Centre for Molecular Medicine and Therapeutics, Child and Family Research Institute, and Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 17) Ludmer Center for Neuroinformatics and Mental Health, McGill University, Montreal, QC, Canada; 18) Department of Oncology, McGill University, Montreal, QC, Canada.

Introduction: Since methylation patterns differ substantially across cell types or tissues, methods for normalization that take this into account may have improved performance. Here we extend the recently-published *funNorm* algorithm [Fortin et al. Genome Biology 2014], designed for normalization of methylation measurements from the Illumina Infinium HumanMethylation450 BeadChip, to allow normalization adjustments to vary by cell or tissue type. **Methods:**The relationships between quantiles of methylation signals and the Illumina control probes are allowed to vary by cell or tissue type, through the use of partial least squares regression models followed by smoothing across the quantiles. **Results:**We analyzed three data sets containing varying numbers of repeated samples and 2-3 different tissues or cell types each (Replication Data: 69 samples on 10 individuals, whole blood, blood spots, and buccal cells; Systemic Auto-Immune Disease Data: two samplings of T-cells (10 individuals) and Monocytes (8 individuals); Gestational Diabetes Data: Placenta and cord blood (one technical replicate each). For technical replicates, the variability between replicates was reduced by as much as 50% in some tissue types. Improvement between biological replicates was of the order of 10-20% in most cell types or tissues examined. Results were quite similar across probe annotations. Notably, agreement between replicates for the X chromosome was substantially improved. **Conclusions:**This normalization method, an extension of *funNorm*, has optimized performance for multiple tissues or cell types in a single data set. A Bioconductor package, *funtooNorm*, implements this approach, and also includes visualization of cross-validation results to facilitate selection of the best number of PLS components.

1295F

Leveraging the diploid genome to increase power in *QTL studies. *M. Subramaniam[1], C. J. Ye[1], N. Zaitlen[2].* 1) Department of Epidemiology and Biostatistics, University of California San Francisco, CA; 2) Department of Medicine, University of California San Francisco, CA.

Next generation sequencing coupled with molecular assays has enabled unprecedented opportunities to quantitatively measure genome function. When combined with dense genetic data, quantitative trait locus (QTL) mapping of functional genomic traits is a fundamental tool for understanding the genetic basis of processes such as transcription regulation. In standard *QTL analysis, genotypes are tested for association with estimated abundance of each genomic feature. While powerful, this approach ignores the diploid nature of our genomes, testing combined abundances across both alleles of every feature. In this work, we develop a new phase aware test for *QTL analysis (PhAT-QTL) leveraging allele specific estimates of genomic features. Briefly, we phase the genotypes of all individuals in the study and then test for association between the haploid count of each SNP and the allele specific expression level of the gene on the same haplotype. We use a linear mixed model to account for the correlated environments of haplotypes from the same individual. Through analytical derivations and simulations, we show that power increases relative to standard genotype based tests as a function of the number of heterozygotes and the noise correlation between haplotypes. Simulations show that phasing error and allele specific transcription quantification error results in a loss of power as opposed to bias. We apply PhAT-QTL to a new RNA-sequencing dataset in stimulated dendritic cells from 250 healthy human donors. In addition to having more power to detect eQTLs, and trQTLs, we scanned for specific cases where PhAT-QTL detects allele specific expression but no known eQTLs. In those cases, we consider a model of autoregulation in which the total expression level across both copies of a gene is constrained due to cellular feedback. In simulations and analysis of real data, PhAT-QTL obtained up to a 40% increase in effective sample size as compared to standard genotype based eQTL association test. With denser genetic maps and technological advances to obtain longer reads, we expect PhAT-QTL to be broadly applicable to a number of other *QTL analyses and will greatly impact the discovery of previously undetectable signals.

1296W

Firth Logistic Regression Reduces Estimation Bias for Rare Variant Analysis or Small-Sample Studies. *H. Zhou, V. Pradhan.* Pfizer, cambridge, MA.

Logistic regression has been widely used to test associations between genotypes and binary outcomes. The advantages of Logistic regression include ease of implementation and the ability to adjust for covariates. However, when rare variants are studied (even jointly) or when studies have limited sample size (such as pharmacogenetic studies of early stage clinical trials), separation occurs frequently. Separation is a phenomenon where one or a linear combination of the covariates completely separates the response and non-response hyperplanes; as a result, at least one of the parameter estimates diverges to infinity. In genetic studies, separation happens when only one type of events occurs in a certain genotype group. As a result, in fitting logistic regression model genetic effect size estimates and standard errors become infinite, and confidence interval becomes uninformative. Although Fisher's exact test has been used frequently to deal with the separation issue, in absence of using additional covariate information, it is less effective. In addition, as maximum likelihood estimation used by Logistic regression is only asymptotically unbiased, bias will be present when analyzing rare variants or small sample data. A simple approach called Firth Logistic regression provides an efficient solution and handles the issue of separation. By multiplying a penalty function to the likelihood known as Jeffreys invariant prior, Firth Logistic regression produces finite parameter estimates and substantially reduces bias due to small sample size (Heinze and Schemper (2002), Maiti and Pradhan (2007)). In the current study, we generate rare variant data with a wide range of disease prevalence, genetic effect sizes, and genetic structures, and show that Firth Logistic regression reduces estimate bias for both single marker analysis and joint analysis using the collapsing approach. In addition, we simulate a small-sample genome-wide association analysis utilizing the data from a real phase II pharmacogenetic study and show that compared to conventional Logistic regression, Firth Logistic regression preserves the type-I error rate and provides valid p-values. Because of the effectiveness of penalized maximum likelihood estimation, future work should focus on extend the approach to continuous outcome analysis.

1297T

Longitudinal Gaussian graphical model integrating gene expression and sequencing data for autism risk gene detection. *K. Lin[1], H. Liu[2], K. Roeder[1].* 1) Department of Statistics, Carnegie Mellon University, Pittsburgh, PA; 2) Department of Operations Research and Financial Engineering, Princeton University, Princeton, NJ.

Detection of genes influencing autism spectrum disorder (ASD) has been challenging since hundreds of risk genes affect neurodevelopment through complex interactions. Our approach to find ASD genes overcomes this challenge by integrating brain gene expression data with DNA sequence data, especially *de novo* variations. To do so we first estimate a gene interaction network in post-conception brains measured by gene expression. Under the assumption that genes interacting with bona fide ASD genes are themselves likely to affect ASD risk, we substantially augment the genetic signal by this integration. The brain microarray dataset comes from the BrainSpan data while the gene sequencing data are the marginal genetic signals measured by Transmission and De Novo Association (TADA) scores originating from data compiled by the Autism Sequencing Consortium. We build upon an existing technique, called DAWN, which was recently used to analyze network data from BrainSpan samples from the prefrontal cortex and primary motor-somatosensory cortex of 10-19 postconceptual weeks old. An advantage of this analysis is that it uses gene expression tissue known to be associated with a key developmental time and region of the brain associated with autism, but the number of samples available from this spatio-temporal period is extremely limited. We extend this method by 1) using a novel longitudinal transformation to incorporate more spatio-temporal regions of the BrainSpan data for our gene network estimated by neighborhood selection and 2) using modern high-dimensional hypothesis tests on every edge of the Gaussian graphical model to prune our gene network, making it more insensitive to tuning parameter mis-specifications. By improving the gene network estimation, we detect 200+ genes, roughly a fourth which are validated by new sequencing data -- an improvement over previous analyses. Our detected risk genes are significantly enriched for biological processes related to chromatin modification, neurogenesis, and others.

1298F

Allele Frequencies of Metabolic Syndrome-related SNPs in a Mexican Mestizo population. *R. Arguello[1,2], RI. De la Cruz-Granados[1], FF. Gonzalez-Galarza[1], JE. Gaytan-Arocha[1], RD. Arellano-PerezVertti[1], FO. Gonzalez-Galarza[2], AY. Garcia-Marin[1], EA. Ramirez-Ramirez[1,2], KL. Valdes-Morales[1], D. Delgadillo-Guzman[1].* 1) Faculty of Medicine, Autonomous University of Coahuila, Torreon, Coahuila, Mexico; 2) Instituto de Ciencia y Medicina Genómica. Torreón, Coahuila, México.

IntroductionMetabolic Syndrome is a combination of several diseases which include: central obesity, hypertriglyceridemia, low serum levels of high density lipoproteins, hyperglycemia, and hypertension; all of this is translated as an increased risk of cardiovascular disease and type 2 diabetes. In the past years, GWAS have contributed to reveal genetic variants with a probable association with this phenotype. Our goal was to mine the results of these GWAS to obtain a list of relevant disease-associated SNPs and assess the allele frequencies of risk variants in our sample and compare them to those reported in MEX Hap-Map population. **Methods**A sample of 100 Mexican volunteers of the geographic area known as Comarca Lagunera, which includes the territories of Coahuila and Durango states was analyzed with an Illumina HumanOmniExpress-24microarray. Patients were genotyped disregarding age, ethnicity, sex, or any known disease. Allele frequencies of 40 disease-related SNPs (rs10830962, rs174546, rs3099844, rs782590, rs10838681, rs1883025, rs3757840, rs7841189, rs1127065, rs2083637, rs3764261, rs7865146, rs12957347, rs2113334, rs4846922, rs7903146, rs13226650, rs2217332, rs496300, rs8060686, rs1532085, rs2266788, rs531676, rs886427, rs1535, rs2292354, rs560887, rs9315632, rs157582, rs247617, rs673548, rs964184, rs164898, rs268, rs6947830, rs9940128) were analyzed and compared with MEX Hap-Map population. **Summary of Results**38 out of 40 SNPs revealed similar allele frequencies to those of MEX Hap-Map population. The risk alleles had a low prevalence in both populations. When analyzing our patients individually, 5% patients had more risk alleles than non-risk alleles, 11% had the same number of risk and non-risk alleles, and 84% had more non-risk than risk alleles. This result is somewhat unexpected, since the prevalence of the disease is high amongst Mexican general population. **Conclusions**The allele frequencies of the 40 SNPs are similar amongst our study group and MEX Hap-Map population. The low number of risk alleles in the majority of our individuals does not explain the high prevalence of the disease. It has not scaped to our attention that unknown variants, non genetic factors, or both, may be contributing to the onset of disease. Nonetheless, this demonstrates the need for a wider study in Mexican Population.

1299W

Allele Frequencies of childhood-onset Acute Lymphoblastic Leukemia-associated SNPs in a Mexican Mestizo Population and their relationship with disease presentation. *HM. Arredondo[1,2], EA. Ramírez-Ramírez[1,2], KL. Valdés-Morales[2], AY. García-Marín[2], FF. González-Galarza[2], FO. González-Galarza[1], D. Delgadillo-Guzmán[2], RD. Arrellano-Pérez Vertti[2], JR. Arguello-Astorga[1,2].* 1) Instituto de Ciencia y Medicina Genómica, Torreón, Coahuila, Torreón, Mexico; 2) Facultad de Medicina, Universidad Autónoma de Coahuila, Torreón, México.

Objective: To asses the frequency of the Acute Lymphoblastic Leukemia-associated SNPs in a well characterized Mexican population, in comparison with the MEX Hap-map population and correlate it with the frequency of the disease in México. Methodology: In this study, we extracted DNA from saliva samples of 50 healthy volunteers and analyzed them in a microarray plataform with a customized Illumina Omin-expreess chip. 32 SNPs related with Childhood onset Acute Lymphoblastic Leukaemia were selected from NCBIs GWAS catalog (rs3824662 rs10828317, rs10821936, rs6964969, rs4982731, rs7142143, rs17079534, rs17837497, rs10170236, rs6683977, rs1496766, rs9958208, rs7578361, rs41322152, rs546784, rs4132601, rs7089424, rs2239633, rs10821936, rs11978267, rs2089222, rs11155133, rs2191566, rs7554607, rs12621643, rs10873876, rs9290663, rs6428370, rs1881797, rs563507, rs10849033, rs1879352), and the Allele frequencies of them compared with the MEX Hap-map population. The SNPs in which we didn't have the allele frequencies of both datasets were excluded, as well as those in which the Risk allele had a lower frequency than the MAF. A t-student test was performed to compare the allele differences in both populations. Also, Frequency analysis was performed to count the SNPs in Hardy-Weinberg disequilibrium and denote the ones in which the risk allele had a higher frequency than the wild type. Results: The allele frequencies in our study group were consistent with the MEX Hap-map population; without a significant variation in 9 out of 10 SNPs. From the studied SNPs, 5/10 were not in Hardy-Weinberg equilibrium and of those 5, only 1 had higher frequencies on the risk allele in comparison with the wild type allele. Conclusions: Our small cohort closely resembles the MEX Hap-Map population. Mexicans have been reported to have a higher incidence of Childhood-onset Acute Lymphoblastic Leukemia in comparison with Worldwide incidence. The fact that half of the SNPs are not in Hardy-Weinberg equilibrium and only one of those has a higher frequency in the risk allele than in the wild-type one does not explain the high incidence of the disease in Mexican population. The low coverage and the small number of subjects in both cohorts may not help in making a statistically significant result, but denote the importance of making a larger case-control study in a Mexican population in order to demonstrate the existence (or non existence) of a derivative genetic cause of this disease.

1300T

Deep learning and the prediction of human disease risk. *N. Furlotte, D. Hinds.* 23andMe, Mountain View, CA.

The accurate prediction of disease risk using genetic data remains one of the key challenges in human genetics research. Regression and whole-genome based methods that model the phenotype of interest as a linear function of a set of genetic variants have shown much promise, but generally their predictive power remains limited. One of the frequent criticisms of these methods is that they do not account for higher order interaction effects or more generally that they assume a restrictive genetic architecture when constructing the predictive model.

Artificial neural networks (ANNs) have experienced a resurgence in interest due to the success of deep learning in image and speech processing. These machines do not enforce a rigid relationship between the predictors and the target and can be thought of as general function approximators. Motivated by this feature and the general success of ANNs, recent work has suggested that ANNs might have an advantage over traditional methods of phenotype prediction and promising results have been demonstrated for genome-enabled trait prediction in cattle. However, there has been little recent work in developing and applying neural network-based approaches to the problem of predicting phenotypes and disease risk in human populations.

The development of deep learning based methods for human phenotype prediction has a number of challenges. First, deep learning methods require a large amount of training data. Second, there is not a standard methodology for applying deep learning techniques to prediction tasks or more practically there is not a well-defined network structure given the input data. Finally, large multi-layered networks can be computationally cumbersome to train and it can be difficult to control for overfitting.

In this project, we investigate these challenges and assess the ability of deep learning and ANN based methods to predict phenotypes with genetics by utilizing the large-scale 23andMe database of more than one million customers, 80 percent of whom consent to research. We compare the performance of standard methods like linear, logistic and ordinal regression with whole-genome-based and ANN-based predictive methods and evaluate performance across a spectrum of morphological and disease-related traits. In addition, we compare performance across different network architectures. Our results suggest a potential for applying deep learning methods to improve disease risk prediction.

1301F

Summary Statistics Imputation utilizing Fine-mapping Approach. *Y. Wu[1], E. Eskin[1,2].* 1) Computer Science, UCLA, Los Angeles, CA; 2) Human Genetics, UCLA, Los Angeles, CA.

Genotype imputation has been widely utilized for two reasons in the analysis of Genome Wide Association Studies (GWAS). One reasonis to increase the power for association studies when the causalSNPs are not collected in the GWAS study. The second reason is to aid in the interpretation of a GWAS result by predicting the association statistics at untyped variants. In this project, we show that the prediction of association statistic at untyped variants which do in fact have an influence on the trait is overly conservative. This is because current imputation methods assume that none of the variants in a region affect the trait which is often inconsistent with the observed data. In this paper, we propose a new method, Summary-Imp, which can impute the associationstatistics at untyped variants taking into account that some of the variants in the region may affect the trait. Our method builds on recent methods that impute the marginal statistics for GWAS utilizingthe fact that the marginal statistics follows a multivariate normal distribution. We utilize both simulated and real data sets toassess the performance of our method. We show that traditional imputation approaches underestimate the association statistics for variants involved in the trait while our approach provides less biased estimates of these association statistics.

1302W

The causal effect of adiposity on vascular dysfunction in healthy adolescents. *K. H. Wade[1,2], T. Khan[3], J. E. Deanfield[3], A. D. Hughes[4], N. Chaturvedi[3], A. Fraser[1,2], D. A. Lawlor[1,2], G. Davey Smith[1,2], N. J. Timpson[1,2].* 1) Integrative Epidemiology Unit, University of Bristol, Bristol, United Kingdom; 2) School of Social and Community Medicine, Faculty of Medicine and Dentistry, University of Bristol, Bristol, BS8 2BN, UK; 3) Vascular Physiology Unit, Institute of Cardiovascular Science, University College London, London W1T 7HA; 4) Cardiometabolic Phenotyping Group, Institute of Cardiovascular Science, University College London, London, W1T 7HA.

Adiposity is a known risk factor for vascular dysfunction in adults, with some evidence that this effect may be apparent in early life. However, conventional cross-sectional and case-control studies may suffer from reverse causation, confounding and other sources of bias. In the Avon Longitudinal Study of Parents and Children (ALSPAC), a large UK birth cohort, we used an allelic score comprising 97 genetic variants as an instrument to test the causal effect of body mass index (BMI) on a range of detailed vascular phenotypes measured at 11 and 18 years. Alongside confirmatory evidence for a positive causal role of adiposity on systolic and diastolic blood pressure at 11 and 18 years of age, we found positive effects of BMI on pulse rate (0. 70bpm; 95% CI: 0. 25, 1. 16; *P*=2. 51x10-03), brachial artery diameter (0. 04cm; 95% CI: 0. 02, 0. 05; *P*=5. 49x10-09) and brachial artery compliance (0. 02mm2/mmHgx10-3: 95% CI: -0. 003, 0. 04; *P*=0. 09) at age 11, and on left atrial size (0. 08cm; 95% CI: 0. 05, 0. 11; *P*=1. 03x10-08), left ventricular diameter during diastole (0. 04cm; 95% CI: 0. 01, 0. 07; *P*=0. 004), left ventricular mass indexed to height in meters2. 7 (1. 00g/m2. 7; 95% CI: 0. 64, 1. 36; *P*=5. 10x10-08) and e', the peak velocity of the lateral mitral annulus in early diastole, (0. 01cm/s; 95% CI: 0. 0004, 0. 03; *P*=0. 04) at age 18. We found inverse associations (not in the hypothesised causal direction) of BMI on pulse wave velocity (PWV) at 11 and 18 (-0. 01m/s; 95% CI: -0. 01, -0. 002; *P*=0. 01 at age 11 and -0. 01m/s; 95% CI: -0. 02, -0. 002; *P*=0. 01 at age 18). Although our results suggest that increased BMI is causally related to most measures of adverse cardiovascular health in childhood and adolescence, we also found a paradoxical benefit of greater BMI on brachial artery compliance and PWV, suggesting that the aetiology of cardiovascular health is complex. Therefore, caution may be needed when assessing likely implications of policy approaches targeting adiposity on the cardiovascular risk profile throughout the lifecourse. Further replication of these findings using Mendelian randomization studies across all ages, together with two-step mediation analyses of the association between adiposity and cardiovascular intermediates, might help to better understand how adiposity influences cardiovascular health across the lifecourse.

1303T

Estimating aggregate penetrance of actionable genomic findings in European American and African American exomes. *R. C. Green[1,2,3,4], P. Natarajan[2,3,5], N. B. Gold[2,6], A. G. Bick[2,3,7], H. McLaughlin[2,4,8], P. Kraft[2,9], H. L. Rehm[2,4,8], G. M. Peloso[2,3], J. G. Wilson[10], A. Correa[11], J. G. Seidman[2,7], C. E Seidman[1,2,7,12], S. Kathiresan[2,3,5].* 1) Department of Medicine, Brigham and Women's Hospital, Boston MA; 2) Harvard Medical School, Boston, MA; 3) Broad Institute of MIT and Harvard, Cambridge, MA; 4) Partners HealthCare Personalized Medicine, Boston, MA; 5) Department of Medicine, Massachusetts General Hospital, Boston, MA; 6) Boston Children's Hospital, Boston, MA; 7) Department of Genetics, Harvard Medical School, Boston, MA; 8) Department of Pathology, Brigham and Women's Hospital, Boston, MA; 9) Departments of Statistics and Epidemiology, Harvard School of Public Health, Boston, MA; 10) Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, MI; 11) Departments of Pediatrics and Medicine, University of Mississippi Medical Center, Jackson, MI; 12) Howard Hughes Medical Institute, Harvard Medical School, Boston, MA.

Background: The American College of Medical Genetics and Genomics has recommended that pathogenic (PVs) and likely pathogenic variants (LPVs) in 56 genes, corresponding to 24 actionable diseases, should be returned as secondary findings with clinical sequencing. However, the penetrance of these variants in populations unselected for family history is largely unknown. We tested the hypothesis that persons carrying PVs and LPVs in these genes are more likely to manifest suggestive clinical features (SCFs) of corresponding actionable diseases. Methods: Among 462 European American Framingham Heart Study (FHS) participants and 3,223 African American Jackson Heart Study (JHS) participants, 642 and 4,429 unique variants, respectively, were classified across the 56 genes, without knowledge of clinical information, as either PVs, LPVs, of uncertain significance, or benign. Clinical history of cancer, and data from the most recent echocardiograms, electrocardiograms, and lipid measurements on all participants were ascertained without knowledge of genetic variants. Within each cohort, we compared the observed number of SCFs among those with corresponding PVs/LPVs to the number expected, assuming rates of SCFs were identical among individuals with and without PVs/LPVs. Results: PVs/LPVs in the 56 genes were found in 8 FHS (1. 7%) and 67 JHS participants (2. 1%). Persons with PVs/LPVs were more likely than expected to have SCFs in both the FHS (50. 0% vs 10. 8%) and the JHS (19. 6% vs 7. 3%). Standardized incidence ratios of SCFs among FHS and JHS participants with PVs/LPVs were 4. 63 (95% CI 1. 25-11. 9;*P*= 0. 006) and 2. 69 (95% CI 1. 34-4. 82; *P*= 0. 001), respectively. Conclusion: Unselected participants in population-based studies carrying PV/LPVs in genes for actionable conditions have an increased risk of displaying features associated with the corresponding clinical diseases.

1304F

Improved Association Testing with Linear Mixed Models by Modeling Non-Genetic Phenotypic Covariance Structures. *M. Conomos[1], S. M. Gogarten[1], T. Sofer[1], A. M. Stilp[1], A. Szpiro[1], T. Lumley[2], K. M Rice[1], C. C. Laurie[1], T. A. Thornton[1].* 1) Biostatistics, University of Washington, Seattle, WA; 2) Statistics, The University of Auckland, Auckland, New Zealand.

Linear mixed models (LMMs) have emerged as a prevalent analysis approach for genome-wide association studies (GWAS) of complex traits. LMMs account for phenotypic correlation of sample individuals due to genetic similarity by including an empirical genetic relationship matrix (GRM) constructed from observed genotype data in the model covariance structure. However, it is plausible that shared environments, such as household or geographic location, may also result in correlated phenotypes among sample individuals, and these structures are often overlooked in the GWAS setting. We developed a computationally efficient LMM for GWAS that allows for the modeling of additional phenotypic covariance structure, beyond that due to shared polygenic effects. We applied our LMM to quantitative traits measured in the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), a cohort study of approximately 13,000 self-identified Hispanics/Latinos from the U. S. collected via a community-based household sampling design. Through variance component estimation, we found that shared household often explains a significant proportion of phenotypic variability, even after accounting for that attributed to shared genetic effects. Furthermore, standard implementations of LMMs for GWAS make the implicit assumption of residual homoskedasticity, which may be violated in samples with complex population structure. Individuals in HCHS/SOL self-reported membership to six ethnic background groups, and we found strong evidence of non-constant variance across these groups for nearly all phenotypes. Traditionally, the use of robust standard errors and GEE can be used to alleviate issues of non-constant variance; however, in the context of GWAS, these approaches result in a loss of efficiency to detect associations as well as severe inflation of test statistics at markers with low minor allele frequencies. To address this issue, a further adaptation to our LMM allows for group-specific residual variances. Through the comparison of genetic association testing results from LMMs with different covariance structure models, we demonstrate that accounting for phenotypic similarity due to shared environment and allowing for heteroskedasticity by background group can lead to both better control of type-I error and improved signal at previously identified associations. These results illustrate the potential value of collecting environmental and demographic information in future genetic studies.

1305W

Using multi-way admixture mapping to elucidate TB susceptibility in the South African Coloured population. *M. Daya[1], L. van der Merwe[1], C. R. Gignoux[2], P. D. van Helden[1], M. Moller[1], E. G. Hoal[1].* 1) Human Gen & Molecular Biol, Stellenbosch Univ, Cape Town, Tygerberg, South Africa; 2) Department of Genetics, Stanford University, Stanford, California, USA.

The admixed South African Coloured population is ideally suited to the discovery of tuberculosis susceptibility genetic variants and their probable ethnic origins, but previous attempts at finding such variants using genome-wide admixture mapping were hampered by the inaccuracy of local ancestry inference. In this study, we infer local ancestry using the novel algorithm implemented in RFMix, with the emphasis on identifying regions of excess San or Bantu ancestry, which we hypothesize may harbour TB susceptibility genes.

Using simulated data, we demonstrate reasonable accuracy of local ancestry inference by RFMix, with a tendency towards miss-calling San ancestry as Bantu. Regions with either excess San ancestry or excess African (San or Bantu) ancestry are less likely to be affected by this bias, and we therefore proceeded to identify such regions, found in cases but not in controls (642 cases and 91 controls). A number of promising regions were found (overall p-values of 7.19×10^{-5} for San ancestry and less than 2.00×10^{-16} for African ancestry), including chromosomes 15q15 and 17q22, which are close to genomic regions previously implicated in TB. Promising immune-related susceptibility genes such as the *GADD45A*, *OSM* and *B7-H5* genes are also harboured in the identified regions.

In conclusion, admixture mapping is feasible in the South African Coloured population and a number of novel TB susceptibility genomic regions were uncovered.

1306T

Multivariate genome-wide association study for metabolic syndrome. *Y. Lee[1,2], S. Park[1,3], BJ. Kim[1], J. Lee[1], S. Won[1,4].* 1) Korea Center for Disease Control and Prevention, Osong, South Korea; 2) Healthcare System Gangnam Center, Seoul, South Korea; 3) Soonchunhyang University, Seoul, South Korea; 4) Seoul National University, Seoul, South Korea.

Metabolic syndrome (MetS) is a kind of disease which is determined co-occurrence of three of five following medical conditions: abdominal obesity, elevated blood pressure, elevated fasting plasma glucose, high serum triglycerides, and low high-density cholesterol levels. It is well known that people with MetS become more dangerous to cardiovascular disease (CAD) or type 2 diabets mellitus (T2D). For several years, many genome-wide association studies (GWASs) have been conducted, but the genetic determinants of the MetS were still limited and most of variants were related with each metabolic trait not MetS. Recently, various programs which can conduct multivariate GWAS were developed and we considered Genome-wide Efficient Mixed Model Association (GEMMA) and More powerful Family based Quasi Likelihood Score test (MFQLS) programs for detecting new genetic variants. We identified rs4742875 on 9q31. 1 (P = 1. 23 x 10–7) with some previously reported variants related with MetS or metabolic traits by using Korean population-based data. These findings may provide additional insights into understanding disease pathology.

1307F

Meta-analysis of genome-wide association studies with correlated individuals: application to the Hispanic Community Health Study / Study of Latinos (HCHS/SOL). *T. Sofer[1], J. Shaffer[2], M. Graff[3], Q. Qi[4], A. M. Stilp[1], S. M. Gogarten[1], K. E. North[3], C. R. Isasi[4], B. S. Weir[1], C. C. Laurie[1], A. A. Szpiro[1].* 1) Department of Biostatistics, University of Washington, Seattle, WA; 2) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; 3) Department of Epidemiology, University of North Carolina, Chapel Hill, NC; 4) Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY.

Motivation and application: Multiple GWAS are often meta-analyzed to increase the power to detect associations of traits with genetic variants. Meta-analysis also may be performed within a single study that is stratified by, e. g. , genetic groups of the study participants. Having correlated individuals between the studies or strata may complicate meta-analyses of GWAS results, potentially limiting power and inflating Type 1 error. In the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), there are 12,803 genotyped individuals, each belonging to one of 6 genetic strata defined by both self-reported Hispanic/Latino subgroup and genetic similarity. Sources of correlations between study individuals include genetic relatedness, shared household, and shared community. **Methods:** We propose a linear mixed effect model that yields estimates of the covariance between the effect estimates across strata and implement a computationally efficient method for meta-analysis ("MetaCor") that uses these covariance estimators. **Simulations:** We generated datasets with individuals from 3-member clusters, stratified by sex. Sources of correlations were genetic similarity and correlated residuals. We compared MetaCor to (1) no stratification, (2) not accounting for correlations, and (3) removing people to generate independent strata ("StratInd"). **Results:** In simulations, MetaCor and StratInd always protected type 1 error; however, MetaCor was more powerful than the alternatives. In the data analysis we compared MetaCor to StratInd. To generate 6 genetic strata with low correlations, we created a data set of 11,788 individuals such that any genetic group did not have a person living in the same household with someone, or a relative of up to 3rd degree, from another genetic group. In BMI analysis, StratInd had high inflation (χ=1. 065) compared to MetaCor (χ=1. 034 on the entire data set). Given the surprisingly high inflation of StratInd, we applied MetaCor to the same set of 11,788 individuals, yielding χ=1. 032. For power, we studied MetaCor in GWAS of dental caries. The top hit, rs7791001, had p-value=4. 65x10-8 (StratInd), and 3. 66x10-8 (MetaCor). **Conclusion:** Simulations and data analyses show that our test is powerful, and effective in control of inflation. An important consequence of our investigation is that even having distant relatives (4th degree) between GWAS studies can inflate meta-analysis results, and that our proposed method results in notable improvement.

1308W

GWAS-based Mendelian-Randomization and Path Analyses of Causal Effects of Lipid Risk Factors on Coronary Artery Disease. *Y. Tan[1,2], D. Liu[1,2,3].* 1) Biochemistry and Molecular Biology , Pennsylvania State University College of Medicine, Hershey, PA, 17033; 2) Institute of Personalized Medicine, Pennsylvania State University College of Medicine, Hershey, PA,17033; 3) Deptment of Public Health Sciences, Pennsylvania State University College of Medicine, Hershey, PA, 17033.

Plasma lipid levels haven been identified as a inheritable and heritable risk factors for coronary artery disease (CAD), one of the leading causes of death. Due to some confounders, it is difficult to infer if the high risk is attributed to causal effects of these risk factors on CAD. For this reason, genetic variants associated with lipid levels may qualify as instrument variables and be used to carry out Mendelian randomization (MR) analysis. To define causal effects of risk factors on a disease, we developed a method called GWAS-based MR path analysis (GMRP). GWAS or GWAS meta-analysis provides genome-wide SNPs associated with risk factors for choice of instrumental variables and MR analyses exclude confounders and quantize bivariate outcome variable, which provide a way to path analysis of diseases. To choose valid SNPs for MR analyses, we formulized Do et al's ad hoc criteria. We obtained 368-768 SNPs from Willer et al's GWAS meta-analyzed data using our method and performed GMRP analyses of beta values of these SNPs on LDL, HDL, triglycerides (TG) and CAD. The results show that LDL and TG had significant direct causal risks for CAD while HDL had significant direct protection effect against CAD, which is opposed to Do et al's result of MR analysis of 185 SNPs. Adding total cholesterols (TC) into the multivariate regression models changed HDL significant protection effect against CAD to weak risk, but significantly strengthened direct causal effects of LDL and TG on CAD. However, indirect causal effects of HDL from TG to CAD, from TC to CAD and from TG to TC and to CAD made HDL be still significantly negatively correlated to risk for CAD. We used direct path coefficients of SNPs onto these risk factors and CAD to execute GMRP analyses and found that the beta values of these lipid variables became very small but their direct and indirect path coefficients onto CAD were not significantly altered, meaning that path coefficients can be defined as causal effects on CAD. Linkage disequilibrium had obvious impacts on analysis of causal effects of CAD.

1309T

A Mendelian Randomization Study of the Effect of Type-2 Diabetes on Bone Mineral Density. *O. S. Ahmad[1,2], J. A. Miller[3], M. Mujammami[1,2], J. B. Richards[1,2,3,4].* 1) Department of Medicine, McGill University, Montreal, Quebec, Canada; 2) Centre for Clinical Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal, Quebec, Canada; 3) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 4) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK.

Although patients with type-2 diabetes (T2D) have on average a higher bone mineral density (BMD) than patients without the disease,observational studies provide evidence for both poorer bone qualityand higher fracture risk for a given BMDin patients with T2D compared to those without this condition. Because of confounding in observational studies between T2D and other metabolic factors known to influence bone homeostasis, the precise effect of T2D on BMD is unclear. Here, we analyze genome-wide association study (GWAS) data to obtain quantitative estimates of the causal effect of T2D (with data from 34,840 T2D cases, 114,981 controls)and fasting glucose (data from 133,010 non-diabetic individuals)on BMD (data from 50,933 individuals)using *Mendelian randomization* (MR), an application of the method of instrumental variables to the analysis of genetic data. Using multiple genetic variants associated with T2D and related metabolic traits, we find that a genetically increased risk of T2D increases femoral-neck BMD (+0. 03 standard deviation (SD) increase in BMD [95%CI: 0. 01 - 0. 06, P = 0. 017] per unit increase in log-odds of T2D). We find also that an increase in fasting glucose increases femoral-neck BMD in a pooled sample of non-diabetic individuals (+0. 13 SD increase in BMD [95%CI: 0. 02 - 0. 25, P = 0. 027] per 1 mmol/Lincrease in fasting glucose). These findings provide evidence for a causal relationship between genetically increased risk of T2D, fasting glucose, and BMD.

1310F

Contribution of variants at genes expressed in fetal and adult pancreatic islets to type 2 diabetes risk. *K. Gaulton[1], D. Calderon[2], D. Golan[3,4], M. Cruz[5], A. Valladares-Salgado[5], E. Parra[6], T. Raj[7,8], J. Below[9], C. Hanis[9], A. Morris[10], M. McCarthy[1,11,12], J. Pritchard[3,13,14], GoT2D consortium, AGEN-T2D consortium, MAT2D consortium, SAT2D consortium.* 1) Wellcome Trust Centre for Human Genetics, Oxford, UK; 2) Biomedical Informatics, Stanford University, Stanford, CA; 3) Department of Genetics, Stanford University, Stanford, CA; 4) Department of Statistics, Stanford University, Stanford, CA; 5) Unidad de Investigación Médica en Bioquímica, Hospital de Especialidades, Centro Médico Nacional Siglo XXI, IMSS, Mexico City, Mexico; 6) Department of Anthropology, University of Toronto at Mississauga, Mississauga, Ontario, Canada; 7) Departments of Neurology and Medicine, Brigham and Women's Hospital, Boston, MA; 8) The Broad Institute of Harvard and MIT, Cambridge, MA; 9) School of Public Health, The University of Texas Health Science Center at Houston; 10) Department of Biostatistics, University of Liverpool, Liverpool UK; 11) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK; 12) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK; 13) Department of Biology, Stanford University, Stanford, CA; 14) Howard Hughes Medical Institute, Stanford University, Stanford, CA.

Type 2 diabetes (T2D) is a complex disorder that affects over 250 million individuals worldwide. Recent studies have suggested that population risk of T2D is mediated through a substantial number of variants of modest effect, many of which alter pancreatic islet function. We sought to more precisely characterize the contribution of altered islet function to T2D by combining genome-wide T2D association data in Europeans, South and East Asians, and Mexican-Americans and glycemic association data in Europeans with gene expression from FACS-sorted alpha and beta cells from fetal and adult pancreas. We first weighted variants in each expression dataset as a function of the inverse variant distance to the closest gene and the relative cell-type expression of that gene, and then applied a polygenic regression model (polyTest) to variants genome-wide to estimate the contribution of each cell-type to the observed phenotypic variance. We identified a significant effect of European T2D in genes expressed in beta cells from both adult and fetal pancreas (b=6. 2, *P*=. 0017; b=3. 7, *P*=. 009), which remained even when removing variants with strong T2D association (*P*<. 0001). We further found significant effect of fasting glucose in beta cells from adult pancreas only (b=22. 0, *P*<1x10-10), and no effect of fasting insulin for all cell types. Finally, South Asian T2D samples showed stronger effect in beta cells from adult pancreas (b=17. 7, *P*=. 001), whereas Mexican-American T2D samples showed stronger effect in alpha and beta cells from fetal pancreas (b=20. 1, *P*=. 0002; b=16. 1, *P*=. 0003). Our preliminary results thus suggest that genes expressed in both fetal and adult islets contribute to T2D risk, and further that exploiting the highly polygenic inheritance of T2D using detailed genomic annotation can uncover disease mechanisms.

1311W

Single-variant and Gene-based Tests to Detect Genetic Loci Associated with Diabetic Retinopathy: 1000G Imputation and Whole Exome Sequencing Data. *J. Kim[1], J. E. Below[1], R. E. Klein[2], B. E. K. Klein[2], C. L. Hanis[1].* 1) Epidemiology, University of Texas School of Public Health, Houston, TX; 2) Ophthalmology and Visual Sciences, The University of Wisconsin School of Medicine and Public Health, Madison, WI.

Diabetic retinopathy (DR) is the most common microvascular complication of diabetes and can lead to blindness. Familial aggregation of severe DR in family studies estimate heritability at 24%~52%. Linkage studies and genome-wide association studies (GWAS) have found genetic loci predictive of severe DR risk, but findings are largely inconsistent across studies. To identify common and rare genetic loci contributing to severe DR, we used three approaches; 1) GWAS of common SNPs using Affymetrix Genome-wide Human SNP Array 6. 0 data imputed to the 1000 Genome Phase 1 reference (7. 9 million SNPs passing quality control), 2) single variant tests of common genetic variation from whole exome sequencing (WES) data, and 3) gene-based tests of rare variation from the WES data. A total of 747 Mexican Americans with type 2 diabetes from Starr County, Texas, were included in the GWAS analyses. WES data were available on a subset of 675. Cases included subjects who had moderate or severe non-proliferative DR or proliferative DR. Tests were implemented with three comparisons; 1) comparing cases to controls with no DR (total n=462, logistic model), 2) changing the control definition to include those with mild retinopathy (total n=747, logistic model), and 3) using an ordinal classification of none, mild, and severe DR (total n=747, proportional odds model). WES data were analyzed with comparisons 1 and 2. We evaluated score statistics and optional sequence kernel association tests (SKAT-O) for single variant and gene-based testing, respectively. All analyses were adjusted for age, duration of diabetes, body mass index, systolic blood pressure, fasting glucose, and cholesterol. No loci showed a significant association after Bonferroni correction; however, we found suggestive signals in nine loci on chromosome 3, 4, 6, 11, and 21 in all the GWASs (p=5. 42E-8~8. 27E-6). In singe variant tests using WES data, *IRAK2* and *DAZL* on chromosome 3 and *NOX3* on chromosome 6 were associated with severe DR (P=. 02~. 03). *DAZL* had a DR-related variant in the 5' UTR in the single variant tests and also showed associations with severe DR in gene-based test (P=. 03~. 05). *IRAK2* and *NOX3* have biological relevance to diabetic retinopathy through the interleukin-1 receptor (*IL1R*) and inflammatory cytokines. *DAZL* is a novel gene not previously reported. These preliminary findings provide additional evidence implicating genes in the pathology of DR.

1312T

Identification of Causal Mechanisms in Diabetes Pathophysiology in the Risk of Primary Open-Angle Glaucoma. *L. Shen[1], S. Walter[2], R. Melles[3], M. Glymore[2], E. Jorgenson[1].* 1) Division of Research, Kaiser Permanente, Oakland, CA; 2) Department of Epidemiology and Biostatistics, University of California, San Francisco,CA; 3) Department of Ophthalmology, Kaiser Permanente Northern California Redwood City Medical Center, Redwood City, CA.

Although type 2 diabetes (T2D) has been implicated as a risk factor for glaucoma, the potential for unmeasured confounding, unambiguous temporal order and possibility of shared causes have hampered causal conclusions. We performed separate sample Mendelian Randomization using the Genetic Epidemiology Research Study on Adult Health and Aging (GERA) cohort (n=69,685) to estimate effects of T2D on primary open angle glaucoma (POAG; 3,554 cases). Genetic instrumental variables (IVs) for overall and mechanism-specific (i. e. , linked to T2D via influences on adiposity, β-cell function, insulin regulation, or other metabolic processes) T2D risk were constructed using 39 genetic polymorphisms established to predict T2D in large-scale genome-wide association study of T2D. IV estimates indicated T2D increased POAG risk (OR=2. 53, 95% CI: 1. 04, 6. 11). The IV for β-cell dysregulation also significantly predicted POAG (ORβ-cell=5. 26, 95% CI: 1. 75, 15. 85), even among individuals without diagnosed T2D, suggesting that metabolic dysregulation may increase POAG risk prior to T2D diagnosis. T2D risk variant in the *MTNR1B* gene predicted risk of POAG independently of T2D status, indicating possible pleiotropic physiological functions of melatonin, but IV effect estimates were significant even excluding *MTNR1B* variants. This is the first genetic IV study of T2D and glaucoma, providing a powerful approach to evaluating this hypothesized effect. Our findings substantially bolster observational evidence that T2D increases POAG risk.

1313F

Playing musical chairs in multi-phenotype studies improves power and identifies novel associations. *H. Aschard[1], P. Kraft[1], N. Zaitlen[2].* 1) Department of Epidemiology, Harvard School of Public Health, Boston, MA; 2) Department of Medicine Lung Biology Center, University of California San Francisco, San Francisco, CA.

Variability in complex human traits is associated with many factors, including exposures, biomarkers and genetic variants, all of which are increasingly collected in large-scale cohorts. Identifying the genetic variants that are causally associated with human phenotypes among the tens of millions of variants that are typically tested remains a challenge. Current strategies to improve power to identify modest genetic associations mostly consist of applying univariate statistical approaches such as linear or logistic regression (LR) and increasing study sample sizes. While successful, these approaches do not leverage the environmental and genetic factors shared between the multiple phenotypes typically collected in contemporary cohorts. Here we develop a method called Musical Chairs (MC) that improves identification of small effects in studies where a large number of correlated variables have been measured on the same samples. When testing a specific variant for association with a phenotype of interest, including additional correlated phenotypes as covariates can increase or decrease power depending on the underlying causal relationship between the genetic variant, the phenotype, and the covariates. MC is a data-driven approach that leverages our previous work (Zaitlen et al PG 2012, Aschard et al AJHG 2015) to select covariates that will increase power for each SNP-phenotype pair considered. MC is computationally efficient and maintains a controlled false positive rate even in the presence of thousands of phenotypes. Simulations based on phenotypic correlation structures from real cohorts provide direct support that large sets of correlated variables can be leveraged to achieve dramatic increases in statistical power equivalent to a two or even three or four fold increase in sample size. To demonstrate the power of our approach in real data with thousands of phenotypes, we performed a genome-wide screen for cis expression QTLs in the GEUVADIS cohort. We examined the expression of 12,167 genes in 375 individuals of European descent. At a stringent FDR of 0. 1% standard LR identified 1,660 genes with at least one cis-eQTL within 50kb of the gene while our MC approach identified 2,154 genes, and increase of 30%. As cohorts move toward large-scale phenotypes collections as in the electronic health records of the UK Biobank, Million Veterans, Kaiser, and Mt. Sinai cohorts, MC will improve the ability of the community to identify associated genetic variants.

1314W

Reducing false-negative exclusion of GWAS by using redescription and computational topological analysis to identify complex phenotypes describing pathways. *D. E. Platt[1], P. Zalloua[2, 3], S. Basu[4], L. Parida[1].* 1) IBM T. J. Watson Research Center, Yorktown Hgts, NY, USA; 2) Lebanese-American University, Byblos, Lebanon; 3) Harvard School of Public Health, Boston, MA; 4) Dept. of Mathematics, Purdue University, West Laffayette, IN.

Coronary artery disease (CAD) progression starts damage to arterial epithelial layers, which may take years to develop. CAD risk factors form a highly correlated cluster of conditions called metabolic syndrome with diverse pathways. Odds ratio associated with some pathway producing CAD is identical to the odds ratio that CAD was caused by that pathway, leading to dilution. GWAS promised to reveal clear genetic causes of disease. Yet, identified SNPs only account for≈10% of CAD, leaving the remaining heritability unexplained. One gap is the large probability of false negatives given the genome-wide false-positive threshold. Also, genome-wide logistic regressions subtract other risk factors by including "adjustments," though, isolating genetic impact specific to CAD excludes the genetics of pathway-induced pathogenic etiology. In this study, we applied pattern discovery seeking significant phenotypic associations to select statistically significant combinations of factors that would yield significant logistic regressions. Pathway mechanisms link putative phenotypes: condition *A* implies the presence of condition *B*, or *A⇒B*, and therefore also that *A=A⇒B*. Therefore, the list of subjects *S(A)* associated with condition *A* will be the same as the list *S(A⇒B)*. Therefore, we sought clusters of patterns sharing similar sets of subjects measured by Jaccard distances commensurate with epidemiological odds ratios. Such clusters represent the same groups of subjects identified by multiple different patterns, or "redescriptions," reflecting underlying pathway-specific etiology. As such, the patterns become multicomponent phenotypes suitable for GWAS. The structure of these clusters was further explored using computational topological analysis, seeking to identify Vietoris-Rips complexes, where the patterns are vertices, the Jaccard distances provide the filtration, and the lifetimes of the generators of interest are within the Jaccard threshold. The generator complexes are also candidate phenotypes. We applied logistic regression to SNPs drawn from the RAAS complex, as well as GWAS predicting these complex phenotypes. We also identified SNPs individually predicting all conditions comprising complex phenotypes, requiring the joint significance to be genome wide. This greatly increases sensitivity, reducing the threshold excluding false negatives, but at the cost of requiring significance of for multiple factors.

1315T

Using imputed genotype data in joint score tests for genetic association and gene-environment interactions in case-control studies. *M. Song, N. Chatterjee.* Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville , MD.

BackgroundGenome-wide association studies (GWAS) are now routinely imputed for untyped SNPs with powerful tools such as IMPUTE2 and various reference panels such as HapMap2 or 1000Genome. The use of predicted allele count for imputed SNPs as the dosage variable is known to produce validscore-test for genetic association. **Methods**In this report, we investigate how to best handle imputed SNPs in various modern complex tests for genetic association incorporating gene-environment interactions in case-control studies. We focus testing under a logistic regression framework using various alternative methods that rely on varying degree on the assumption of gene-environment independence in the underlying population. As increasingly large scale GWAS are being performed through consortia effort, where it is preferable to share only summary-level information across studies, we also explore how these methods could be implemented in the context of meta-analysis. The methods are evaluated using simulation studies and through analysis of a GWAS of lung cancer accounting for interaction of genotypes with smoking status. **Results**Both simulation studies and the real data application illustrate that the proposed methods can maintain type-I errors for joint tests for genetic association and interactions for common SNPs. For less frequent variants, however, application of some of these methods, whether the genotype data is imputed or not, requires caution because of inaccuracy of the asymptotic standard error calculations in the presence of sparse genotype-exposure combinations. Methods exploiting gene-environment independence showed stronger enrichment for signal for joint test of genetic association and interaction for the analysis of lung cancer GWAS with the pattern being consistent with the simulation studies of power.

1316F

Statistical properties of spatial regression based longitudinal analysis. *T. Schwantes-An, H. Sung, A. F. Wilson.* Genometrics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, MD.

Large-scale longitudinal genetic studies of complex diseases with multiple measures of the same phenotype(s), coupled with low-cost genotyping and next generation sequencing technologies, are now available for genetic analysis. At least three different approaches can be taken in the analysis of these data. In the first, repeated phenotypic measures can be summarized into a single measure in order to detect associations between causal variants and phenotypes. In the second, the repeated measures can be used to reduce the variation of the phenotype, in effect, increasing the relative heritability of the trait. In the third, these measures can be used to measure changing effects of causal variants on the same phenotype over time. In this study we investigate the statistical properties of a method based on spatial regression to order to identify causal variants that have changing effects on phenotypes over time. Ordinarily, spatial regression is used to adjust for the fact that points that are closer together in a three dimensional space are more correlated than points that are far apart. Therefore, instead of summarizing repeated measures into a single value to identify a static association, a "map" of association results across the genome can be defined with the map distance of single nucleotide polymorphisms (SNPs) as one dimension and time as the other dimension. Moran's I statistic is used to test autocorrelation to determine if the observed patterns of associations are significant over time. Two thousand replications of simulated quantitative phenotypes for 2000 individuals were used to compare the power and type I error of this approach with two commonly used longitudinal analysis methods -- slope (the rate of change between first and last exams for a given trait) and Max-Min (the absolute difference between maximum and minimum trait values). At a critical value of 0. 05, the special regression method has an inflated type I error (18. 2% at a p-value threshold of 0. 05) while the type I error for slope and Max-Min are 5. 3% and 5. 5%, respectively. However, after adjusting the critical level (0. 004) to reflect the nominal type I error rate the power to identify a locus with locus specific heritability of 0. 01 was 87. 5% compared to methods using the slope (72. 1%) and Max-Min (8. 3%).

1317W

Exploring genetic variants for anthropometric traits in Hispanic/Latino populations. *M. Graff[1], L. Fernandez-Rhodes[1], H. M. Highland[1,16], C. Schurmann[2], Y. Feng[3], M. Cruz[4], A. Valladares-Salgado[4], J. Rangarajan[3], C. Rojo[5], X. Guo[6], J. Rotter[7], M. Allison[8], R. P. Igo[9], A. Pereira[10], E. John[11], T. Tusie-Luna[12], C. A. Aguilar Salinas[13], C. Gonzalez[14], C. Gignoux[15], S. Iyengar[10], C. Hanis[16], J. E. Below[16], F. Hartwig[17], A. E. Justice[1], K. L. Young[1], C. R. Isasi[18], Q. Qi[19], C. C. Laurie[20], A. M. Stilp[20], S. M. Gogarten[20], V. S. Voruganti[21], G. Chittoor[21], R. Duggirala[22], S. Puppala[22], T. Dyer[22], S. S. Albrecht[23], L. Raffel[24], D. Conti[3], E. Ziv[5], E. Parra[25],E. Kenny[26], R. J. F. Loos[2], C. Haiman[3], K. E. North[1].* 1) Epidemiology, University of North Carolina, Chapel Hill, NC; 2) The Genetics of Obesity and Related Metabolic Traits Program, The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA; 3) Department of Preventive Medicine, Keck School of Medicine, University of Southern California, San Diego, CA, USA; 4) Unidad de Investigacion Medica en Bioquimica, Hospital de Especialidades, CMNSXX1-IMSS, Mexico City, Mexico; 5) Department of Medicine, University of California San Francisco, San Francisco, CA, USA; 6) Director, Laboratory of Statistical and Mathematical Genetics, University of California Los Angeles, Los Angeles, CA, USA; 7) Director, Institute for Translational Genomics and Population Sciences, University of California Los Angeles, Los Angeles, CA, USA; 8) Department of Family Medicine and Public Health, University of California San Diego, CA, USA; 9) Department of Epidemiology & Biostatistics, Case Western Reserve University, Cleveland, OH, USA; 10) Laboratory of Genetics and Molecular Cardiology, Heart Institute, Medical School of University of Sao Paulo, Sao Paulo, Brazil; 11) Cancer Prevention Institute of California in Fremont CA, Fremont, CA, USA; 12) Instituto de Investigaciones Biomedicas, Universidad Nacional Autonoma de Mexico, Mexico City, Mexico; 13) Instituto Nacional de Ciencias Medicas y Nutricion, Mexico City, Mexico; 14) Instituto Nacional de Salud Publica, Mexico City, Mexico; 15) Department of Genetics, Stanford University School of Medicine, Stanford CA, USA; 16) Department of Epidemiology, Human Genetics and Environmental Sciences,The University of Texas Health Science Center at Houston, Houston, TX, USA; 17) Postgraduate Program in Epidemiology, Federal University of Pelotas, Brazil; 18) Department of Epidemiology & Population Health, Department of Pediatrics, Albert Einstein College of Medicine, Bronx, NY, USA; 19) Department of Epidemiology & Population Health, Albert Einstein College of Medicine, Bronx, NY, USA; 20) Department of Biostatistics, University of Washington, Seattle, WA, USA; 21) Department of Nutrition and UNC Nutrition Research Institute, University of North Carolina at Chapel Hill, Kannapolis, NC, USA; 22) South Texas Diabetes and Obesity Institute, Brownsville, TX, USA; 23) Department of Nutrition, University of North Carolina, Chapel Hill, NC, USA; 24) Medical Genetics Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA; 25) Department of Anthropology, University of Toronto at Mississauga, Missisauga, ON, Canada; 26) Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

Despite the increased prevalence of obesity in Hispanic/Latino populations, no large-scale GWAS for any obesity-related traits have been performed. Hispanic/Latino populations are diverse and genetically admixed with recent origins from Europe, Africa and the Americas. Using genome-wide imputation to multiethnic 1000 Genomes Phase 1 reference panels in up to 44,405 individuals of Hispanic/Latino descent across 11 studies, we performed a genome-wide search for loci associated with anthropometric traits, including body mass index (BMI), waist-to-hip ratio adjusted for BMI (WHRadjBMI) and height. Each study adjusted traits for age, sex, study specific covariates, and principal components to correct for population stratification. Assuming an additive genetic model, we tested the association of over 20 million variants using fixed-effects inverse variance weighted meta-analyses in men and women combined. For BMI, we observed genome-wide significant (P<5E-8) associations at 11 established BMI loci (near *SEC16B, TCF7L2, LMX1B, OLFM4, TMEM18, MTCH2, C6orf106, ELAVL4, MC4R, ETV5,* and *FTO*) and two novel loci (in/near *PPP1R3B* and *PAX3*) both driven by common variants (minor allele frequencies>5%), that showed the largest minor allele frequencies among Native American reference samples. *PPP1R3B* (a serine/threonine-protein phosphatase) is expressed in liver and skeletal muscle and regulates glycogen synthesis and has previously been associated with lipid traits in European ancestry populations. *PAX3* encodes a transcription factor that aids in converting white to brown-like adipo-

cyte progenitors. For WHRadjBMI, we did not confirm any known loci at genome-wide significance but we identified a novel locus near *TAOK3* (a serine/threonine-protein kinase) driven by a rare variant (minor allele frequency=0. 8%, polymorphic in African ancestry only) and beta=0. 02 WHR unit change, p=4. 8x10-08. For height, we observed 37 known loci and identified two novel loci near *BRK1* and *RABGAP1,* which were also driven by variants with minor allele frequencies that were largest among Native American reference samples. Follow-up in replication cohorts is underway and any replicated loci will be further interrogated in analyses to fine-map loci and identify causal variants. These results show strong preliminary evidence that future discovery efforts in Hispanic/Latino populations can further inform the genetic architecture of anthropometric traits.

1318T

A recall-by-genotype study linking *ZNF804A* variants to sleep neurophysiology phenotypes. *L. J. Corbin[1], U. Bartsch[2], C. Hellmich[2], C. Durant[3], M. W. Jones[2], N. J. Timpson[1].* 1) MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK; 2) School of Physiology and Pharmacology, University of Bristol, Bristol, UK; 3) Clinical Research and Imaging Centre (CRICBristol), University of Bristol, Bristol, UK.

Some patients suffering from schizophrenia have been shown to have changes in sleep patterns compared with healthy subjects. Deficits in memory processing have also been noted. Since current schizophrenia treatments largely fail to treat cognitive symptoms, these sleep abnormalities may constitute important targets for novel therapeutic intervention. However, the underlying biological processes are not well understood. Family studies have shown that more than 40% of the variation in incidence of schizophrenia may be attributable to differences in genes. Initial genome-wide studies have shown evidence for association of common variants at several loci, of which rs1344706 (in *ZNF804A*) was amongst the strongest. Brain imaging studies suggest rs1344706 may be associated with differences in neuronal connectivity during wakefulness in individuals unaffected by psychiatric illness. To further explore the role of *ZNF804A* in healthy people, a recall-by-genotype (RBG) study was designed to investigate neuronal activity during sleep. 200 male participants from the Avon Longitudinal Study of Parents and Children were invited to take part on the basis of genotype at rs1344706. An equal number of major and minor allele homozygotes were invited with frequencies of 54% and 46%, respectively, achieved in the final study sample (n=26), comparing to 36% and 17% in the overall cohort. Participants completed 2 overnight stays, approximately 2 weeks apart, during which sleep was monitored using polysomnography (PSG). Participants also performed a motor sequence task. Information about habitual sleep behavior was collected using sleep questionnaires, sleep diaries and actigraphy. This study has generated ~400 hours of PSG data and ~8700 hours of actigraphy data, with a very low rate of missing data. In addition to the standard sleep scoring of data, analyses will be performed to assess the individual characteristics of slow-waves and spindles, as well as their coordination, and how these differ across the genetic stratum. The relationship between these derived sleep phenotypes and task performance will be investigated. This RBG experiment has generated dense, high quality phenotypic measures related to sleep neurophysiology and behavior in a subset of individuals selected according to genotype. Using an RBG approach, it is possible to explore genes relevant to disease in healthy individuals, in a financially and pragmatically feasible sample size.

1319F

A practical guide for GWAS based Mendelian randomization analysis. *Y. Hu, D. Hinds.* 23andMe Inc, Mountain View, CA.

GWAS identify an ever-increasing number of trait-associated genotypes. These genotypes are randomly assigned at meiosis and independent of non-genetic confounding, and so can be used as instrument variables in Mendelian randomization (MR) to infer causal roles of the GWAS studied traits. The underlying MR model has been built on continuous instruments and many currently published MR studies are on continuous traits such as lipid levels and body mass index (BMI).

Although the existing method and experience cannot be directly applied to binary GWAS traits, a majority of the currently published GWAS studies are on binary traits. Understanding how to perform a MR analysis on binary intermediate traits would greatly extend the application of MR in genetic studies. Here, we evaluate ways to test and estimate causal effect. We demonstrate with simulations that the causal effect, sample size and strength of the instrument together influence the power of MR analysis. For example, with a normally distributed genetic risk, when the odds of the binary intermediate trait increase by 14%, a cohort of sample size 60,000 has a power of 76% to detect a causal effect of odds ratio (OR) 2. 5 between the binary intermediate trait and the outcome. The power decreases to 27% when the causal effect of OR reduces to 1. 5, illustrating that a large sample size is needed for typical GWAS identified SNPs with small genetic effect size and when the causal effect is small.

Finally, we apply our strategies on 23andMe's large cohort of research participants to evaluate the causal role of morning person to BMI and depression, and high cholesterol to heart disease. We found no evidence for the protective role of morning person for low BMI and less depression, but we replicated previous findings that high cholesterol causes heart disease. Taken together, these findings suggest we have extended the MR methodology to determine causal effect for binary traits, greatly enhancing the utility of the approach.

1320W

Genetically elevated fasting glucose levels and risk of hypertension in African Americans: a Mendelian randomization analysis. *S. Tajuddin[1], M. Nalls[2], A. Zonderman[1], M. Evans[1].* 1) Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD; 2) Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD.

Hypertension (HTN) causes significant morbidity and mortality, and disproportionately affects African Americans (AAs) and low socioeconomic status populations. Elevated fasting glucose (FG) levels are associated with increased risk of HTN in observational analyses while individuals with HTN are at higher risk of developing diabetes mellitus. Whether the association between elevated FG and HTN is truly causal is unknown and could be the result of confounding or reverse causality. The aim of the study was to test the hypothesis that elevated FG is causally associated with HTN risk. We performed a Mendelian randomization (MR) analysis in 942 AA participants from the Healthy Aging in Neighborhoods of Diversity across the Life Span study (HANDLS) by using FG associated allele score as an instrumental variable (IV). We constructed the allele score using five significantly associated SNPs located at *PTPRD, TMEM163-MGAT5*, and *LOC285692 l*oci selected from a genome-wide association analysis of FG. For observational analysis, we fitted a multivariable adjusted logistic regression model to assess the association between FG and HTN risk. Causal estimate of fasting glucose on HTN risk was derived by logistic control IV function estimator – a two-stage regression analysis comprising (1) linear regression of FG on allele score adjusted for potential confounders, and (2) logistic regression to assess the association between genetically predicted FG values and risk of HTN adjusting for residuals from the first-stage regression. The allele score was positively associated with FG levels (beta=0. 143, standard error=0. 023, *P*=1. 45E-9, F-statistic=37. 3) but not with HTN risk (odds ratio (OR)=0. 90, 95% confidence interval (CI) 0. 56-1. 46), which together suggest the allele score is a strong instrumental variable. In observational analysis, increased FG levels were significantly associated with increased risk of HTN [OR=2. 03, 95%CI 1. 13-3. 65, *P*=0. 02]. In MR analysis, genetically elevated FG levels were significantly associated with increased risk of HTN (causal OR=15. 08, 95%CI 12. 98-17. 19, *P*=0. 01). In summary, our findings provide the first evidence that the genetic loci that elevate FG are causally associated with increased risk of HTN among community dwelling AA adults, suggesting that improved glycemic control may reduce risk of HTN. Replication study in independent AA cohorts is warranted.

1321T

Test of Genotypic Association Allowing for Sequencing Misclassification. *L. Zhou, W. Kim, D. Londono, A. Musolf, T. Matise, D. Gordon.* Genetics Dept, Rutgers, the State University of NJ, Piscataway, NJ.

Genome-wide association studies (GWAS) have led to identification of an ever-increasing number of single nucleotide polymorphisms (SNPs) for further studies. However, the presence of genotype sequencing misclassification, differential or non-differential, among cases and controls may cause either an increase in the false positive rate or a decrease in the power of statistical tests. A commonly used statistic to test for association between SNPs in cases and controls is the chi-square test of independence. This statistic tests whether genotype frequencies (single or multi-locus) differ between case and control groups. This is commonly referred to as a test of association. Regions of the genome where frequencies significantly differ are regions that may harbor a disease locus or loci. Several researchers have documented that genotype misclassifications may alter either the null or alternative distribution of the chi-square test. Our overall objective is the development of a statistical test of association that uses NGS data and is robust to random sequencing misclassifications, both non-differential and differential. More concretely, we develop a chi-square test of independence that uses parameters such as the observed alternative variant reads at a given polymorphic site, the coverage per individual, the individual's phenotype, and error model parameters. This statistic is developed in a log-likelihood framework and the Bootstrapping approach is applied to approximate the type I error. We also apply the factorial design to determine the performance of our statistic under specific situations.

1322F

Novel Association Testing Based on Genetic Heterogeneity in GWAS. *W. Pan, Z. Xu.* Division of Biostatistics, University of minnesota, Minneapolis, MN.

The commonly used association tests in GWAS all aim to detect single SNPs with allele frequency (AF) differences between cases and controls. Although many disease-associated SNPs have been identified, they can only account for a small proportion of the genetic variance for a common disease. Two of several possible reasons are small effect sizes and genetic (or locus) heterogeneity, which are related to each other and have been confirmed by completed GWAS. Under genetic heterogeneity it is assumed that a complex disease is not caused by a single SNP or single gene, but by multiple variants in multiple functionally related genes forming a pathway, motivating the application of pathway analyses. However, as for single SNP tests for AF differences, in addition to incomplete annotations of biological pathways, existing pathway methods do not directly take advantage of existing genetic heterogeneity to boost statistical power. Here we propose and apply a single SNP-based test to explicitly account for genetic heterogeneity: it is assumed that a patient population of a complex disease can be decomposed into multiple subpopulations, each with a possibly different set of causal SNPs; under this model, even if the AF of a causal SNP for the whole patient sample is the same as that of the controls, in which the conventional AF tests have no power, our proposed test retains power to detect the causal SNP. We apply the proposed test to the WTCCC data, demonstrating that it can detect novel SNPs and loci missed by the conventional AF tests. For example, when applied to the WTCCC Crohn's disease data, the proposed test detected 128 significant SNPs, while the conventional trend test identified only 60 significant SNPs, among which 46 were common; compared to 32 new loci detected by the meta-analysis of Franke et al (2010) that included the WTCCC data as a subsample and contained a much larger sample size, the new test and the conventional test (based on the WTCCC sample) identified 21 and 8 loci respectively, which overlapped with those of the meta-analysis at 7 and 5 loci respectively. In addition to annotated pathways, the new test can be generalized to discover de novo (i. e. un-annotated) disease-associated pathways. Its relationships with (and advantages over) some existing tests will be discussed.

1323W

Admixture Mapping with Mixed Models. *L. Brown, T. Thornton.* Biostatistics, University of Washington, Seattle, WA.

Genetic studies in recently admixed populations can provide insight into risk factors contributing to disease. Population admixture results in combined genomes from previously isolated ancestral populations that may have discernible allele frequency differences due to natural selection and genetic drift. Genetic variants that show differential risk by ancestry can be identified using admixture mapping. Due to long range correlation in local ancestry values, admixture mapping is not limited to identifying causal variants tagged by single nucleotide polymorphisms alone because local ancestry captures both common and rare variation. When applied to genome-wide association data, mixed model methods have been shown to protect against spurious associations in structured samples by directly accounting for sources of dependence including cryptic relatedness and population stratification. We present a mixed model approach for admixture mapping in the presence of population structure and known or cryptic relatedness. We demonstrate that our proposed method provides a substantial improvement over widely used admixture mapping approaches and identifies variants not found by association studies applied to the same data.

1324T

Novel Genetic Risk Prediction Model using Multiple Secondary Traits. *W. Chung[1], J. Chen[2], C. Chen[1], P. Kraft[1,3], L. Liang[1,3].* 1) Program in Genetic Epidemiology and Statistical Genetics, Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA; 2) Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN; 3) Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, MA.

Despite the success of recent genome-wide association studies (GWAS), for most human complex diseases and traits, a large fraction of overall heritability remain unexplained and the accuracy of genetic risk prediction is generally low because most causal variants have small effect size. Recent studies have shown that human complex traits may share similar genetic architecture with other clinical or molecular phenotypes, such as disease-related traits or gene expression levels. Combining information across such secondary traits might increase the effective sample size for model training, and therefore improve prediction accuracy of the primary trait of interest. In this paper, we develop a new statistical framework for genotype-based risk prediction of complex traits using multiple secondary phenotypes and whole-genome single nucleotide polymorphism (SNP) data. Based on penalized least squares methods, we developed a novel cross trait penalty function to incorporate the shared genetic effects across traits. Our approach has several advantages (1) extracting information from the secondary traits that is useful for primary trait but tuning down information that is not useful; (2) the primary trait and secondary traits can come from same set of subjects or different sets of subjects; (3) our novel implementation of a parallel computing algorithm utilizing Message Passing Interface (MPI) makes it feasible to apply penalized least squares methods on millions of SNPs from thousands of samples. We evaluated our method on GWAS datasets using the Lasso and the minimax concave penalty (MCP) for inducing a sparse solution. We performed large-scale simulation studies to illustrate the excellent performance of our multi-trait approach. For example, when the cross-trait heritability is 50%, the prediction R2 improved from 16% (single trait model) to 21% (two-trait model) when using 1 million SNPs on 7400 training samples and 5900 testing samples. We further applied our method to real GWAS data using body mass index (BMI) as the primary trait and age at menarche as the secondary trait. With 12,000 individuals and 2,000,000 SNPs, the gain in prediction accuracy of our multi-trait approach is 15. 7% using Lasso penalty and 16. 0% using MCP penalty. Finally, comparing to prediction using SNPs within 1kb of 96 BMI loci curated at the NHGRI GWAS category (access 2015. 6. 10), our method has more than 10 times increment in prediction R2 using the same training and testing dataset.

1325F

Functional-Trait GWAS—Going Beyond Single-Value Traits. *N. Fusi, J. Listgarten.* Microsoft Research, CAMBRIDGE, MA.

Practically all genome-wide association studies examine either one trait at a time, or several independent/related traits where the relationship among traits does not contain particularly rich structure (i. e. , none beyond simple pairwise correlations). However, with the advent of wearables for health, or “quantified health”, and the approaching ubiquity of electronic health records, traits collected over time are becoming widely available. Longitudinal traits are one example of functional traits—traits which can be viewed as a function of some variable, such as time (e. g. , clinical history), or space (e. g. , position along the genome). Such structured, functional traits offer new opportunities to dissect the genetics of disease by offering richer traits than were previously available. However, truly benefiting from such opportunities requires that the rich structure within these traits can be leveraged by the statistical model of choice. Rich trait structure arises from constraints in the physical world, such as that time moves forward and is smoothly varying, or that the correlation between positions on the genome is slowly decreasing according to linkage disequilibrium. Appropriate modelling approaches in such a setting not only assume that such constraints exist, but, also, critically, that genetic effects can dramatically alter this structure, even while the mean structure for one genotype alone is, for example, smoothly varying. Although several approaches for analyzing longitudinal traits have been presented in recent years, these are computationally slow, and yet, do not generally make effective use of the rich trait structure to increase power. In our work, we show that (i) with a clever insight, a fairly simple model can outperform significantly more complex and computationally expensive models, and (ii) if one is willing to incur the cost of expensive models then one can do much better than existing approaches; in particular, our more expensive models can handle missing function values (e. g. , missing time points in the trait), unevenly sampled functional values (e. g. , not the same time points for all individuals), and can take in to account function domain distance between the observed points (e. g. , points closer in time should appear more similar than points further away in time).

1326W

A general and flexible framework for meta-analyzing dependent studies with overlapping subjects in association mapping. *B. Han[1], D. Duong[2], J. H. Sul[3,4], P. I. W. de Bakker[5,6], E. Eskin[2,7], S. Raychaudhuri[3,4,8,9,10].* 1) Asan Institute for Life Sciences, University of Ulsan College of Medicine, Asan Medical Center, Seoul, 138-736, Republic of Korea; 2) Computer Science Department, University of California, Los Angeles, California, 90095, USA; 3) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, 02115, USA; 4) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, 02142, USA; 5) Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands; 6) Department of Medical Genetics, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands; 7) Department of Human Genetics, University of California, Los Angeles, California, 90095, USA; 8) Division of Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, 02115, USA; 9) Partners Center for Personalized Genetic Medicine, Boston, Massachusetts, 02115, USA; 10) Faculty of Medical and Human Sciences, University of Manchester, Manchester, M13 9PL, UK.

Meta-analysis strategies have become critical to augment power of genome-wide association studies (GWAS). To reduce genotyping or sequencing cost, many studies today utilize shared controls, and these individuals can inadvertently overlap among multiple studies. If these overlapping individuals are not removed prior to meta-analysis or taken into account during meta-analysis, they can induce dependence between studies, resulting in spurious associations. A simple solution would be to manually split the overlapping subjects into separate studies, but this can be sub-optimal in terms of statistical power. Moreover, splitting genotype data is often not possible in practice if the primary genotype data are not shared. Previous studies addressed this challenge, but only focusing on specific meta-analysis models such as the fixed effects model (FE), but not on the random effects model (RE). However, application of the RE models to meta-analyses recently became crucial because of the expected heterogeneity in cross-disease analysis, cross-population analysis, and cross-tissue eQTL analysis. In this paper, we propose a flexible and general framework for meta-analyzing dependent studies. The philosophy of our framework is unique, in that our framework does not compete with existing methods; it enables them to account for overlapping subjects. Given dependent studies, our framework transforms the data so it can be used in methods that strictly assume independence between studies. Specifically, our method *decouples* dependent studies into independent studies, adjusting association statistics to account for uncertainties in dependent studies. As a result, our approach enables general meta-analysis methods, including the FE and RE models, to account for overlapping subjects. Moreover, to apply our approach, existing software packages and pipelines for meta-analysis can be re-used without changes, because our method only involves a transformation of the data. We performed a meta-analysis of three autoimmune diseases with shared controls using the Wellcome Trust Case Control Consortium (WTCCC) data, to demonstrate that our framework improved statistical significance of known candidate loci associated to multiple autoimmune diseases. We also show that in a multi-tissue mouse eQTL study, cross-tissue analysis using our method increases the number of discovered eQTLs by up to 19% compared to existing methods.

1327T

Semi-parametric Allelic Tests For Mapping Multiple Phenotypes: Binomial Regression And Mahalanobis Distance. *A. Majumdar[1,2], J. Witte[1], S. Ghosh[2].* 1) Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, CA; 2) Human Genetics Unit, Indian Statistical Institute, Kolkata.

Binary phenotypes commonly arise due to multiple underlying quantitative precursors and genetic variants may impact multiple traits in a pleiotropic manner. Hence, simultaneous analysis of such correlatedtraits may be more powerful compared to analyzing individual traits. Various genotype-level methods oftesting for association, such as MultiPhen [O'Reilly et al. , 2012] have been developed in order to identifygenetic factors underlying a multivariate phenotype. For univariate phenotypes, the usefulness and applicability of allele-level tests have been investigated [Guedj et al. , 2006, 2008;Zheng, 2008], and for a case-control phenotype, the chi-square test of allele frequencydifference among cases and controls is a commonly used method for identifying genetic associations. However, allelic methods for mapping association with multivariate traits have not been studiedmuch in the literature. In this article, we explore two allele-level tests of association for analyzing multivariate phenotypes: one based on a Binomial regression model in the framework of inverted regressionof genotype on phenotype(termed BAMP, Binomial regression-based Association of Multivariate Phenotypes) and the other based on the Mahalanobis distance between the two sample means of vectors ofthe multivariate phenotype corresponding to the two alleles at a SNP adopting the set-upof regressing phenotype on genotype (termed DAMP, Distance-based Association of Multivariate Phenotypes). These methods inherit the flexibility of incorporating both discrete as well as continuous traits inthe multivariate phenotype vector. Some desirable theoretical properties for BAMP are studied. Usingsimulations, the potential of the methods in enhancing the power of detecting genetic association withmultiple traits is evaluated in comparison with MultiPhen, which is based on a genotype-level test. Wefind that the allelic tests yield marginally higher power compared to MultiPhen for multivariate phenotypes. We also find that, for one or two binary traits under a recessive mode of inheritance, allelic testsare substantially more powerful than MultiPhen. The allelic tests and MultiPhen are applied to two realdata and the results offer some support for the simulation study. Since the allelic tests are applicableunder Hardy-Weinberg Equilibrium (HWE), we propose a hybrid approach for association mapping ofmultivariate phenotypes that implements MultiPhen when HWE is violated and BAMP otherwise.

1328F

Evaluation of GWAS-identified 55 SNPs for association with body mass index (BMI) in north Indian population. *B. Mittal[1], A. Srivastava[1], J. Prakash[2], N. Srivastava[2].* 1) Dept Med Gen, Sanjay Gandhi Med Inst, Lucknow, UP, Uttar Pradesh, India; 2) Dept of Physiology, King George Medical University, Lucknow, UP, India.

Background: Several GWAS have identified large number of risk SNP with obesity but require replication in different populations. **Aims and objectives:** We evaluated association of 55 recently identified SNPs across 28 genes in GWAS and additional candidate genes with BMI in general population of northern India. **Materials and methods:** Overall 480 individuals out of 809 were selected befitting strict inclusion/exclusion criteria. Genotyping were performed by Sequenom Mass ARRAY platform or Taqman probes. Statistical analysis was conducted using SPSS ver. 19, SNPStats and GMDR softwares. **Results:** Single locus analysis revealed significant associations of 50 SNPs in FTO, IRX3, BDNF, TMEM18, GHSR, GHRL, KCNJ11, CDKAL1, UCP3, TCF7L2, ADRB1, ADRB2, FAIM2, SH2B1, NPC1, MC4R, POMC, LEPR, GNPDA2, SEC16B, LYPLA1, NEGR1 and MAF with BMI score in Indian population. A strong linkage disequilibrium was observed between FTO and IRX3 SNPs specified as *FTOrs17817449-IRX3rs3751723* (D'=0. 84), *FTOrs8050136-FTOrs1421085* (D'=0. 95); *FTOrs8050136-FTOrs9939609* (D'=0. 94) and *FTOrs1421085-FTOrs9939609* (D'=0. 90). In GMDR analysis of FTO and IRX3 genetic variants revealed the best candidate interaction model- *FTOrs9939609; FTOrs17817449;* IRX3rs3751723 (P< 0. 0001; OR (95%CI) 3. 5(2. 1-5. 9). CART analysis counting all investigated genetic variants of the study revealed the risk associated with obesity using the terminal nodes, individuals carrying the combination of genotypes exhibited a significantly higher risk for obesity such as *FTO rs9939609 (TT) + TMEM18 rs6548238 (CC) + POMC rs1042571 (CT+TT)* [Node 2] (OR=4. 0(1. 8-8. 9); p=0. 0003); *FTO rs9939609 (CC+TC) + TCF7L2 rs7903146(TT)* [Node 3] (OR=6. 4(3. 2-13. 39); p=0. 0000001) and *FTOrs9939609 (TT) + TMEM18rs6548238 (TT+CT)* [Node 4] (OR=6. 6(3. 3-13. 85); p=0. 0000001). The combined genotypes *FTO rs9939609 (CC+TC) and TCF7L2 rs7903146 (TT+CT)* [Node 5] revealed the highest risk (OR=15. 92(7. 7-34. 25); p=0. 0000001). Pathway wise estimated G-Scores for risk alleles was also calculated for obesity (P=0. 001). In silico analysis of interaction network generated by GENEMANIA revealed direct and indirect connections among them via co-expression, gene interactions and physical interactions. **Conclusions:** FTOrs9939609; FTOrs17817449; IRX3rs3751723, TCF7L2 rs7903146 and TMEM18rs6548238 emerged as major SNPs contributing to higher BMI risk in North Indian population. **Financial support from DBT and ICMR Govt. of India.**

1329W

The use of Complimentary Pairs Stability Selection as an approximation to analysis with replication data. *J. A. Sabourin[1], C. D. Cropp[2,3], Y. Kim[1,4], L. C. Brody[5], J. E. Bailey-Wilson[2], A. F. Wilson[1].* 1) Genometrics Section, CSGB / NHGRI / NIH, Baltimore, MD; 2) Statistical Genetics Section, CSGB / NHGRI / NIH, Baltimore, MD; 3) Translational Genomics Research Institute (TGen)/Integrated Cancer Genomics Division/Phoenix, AZ; 4) Macrogen Clinical Laboratory/Bioinformatics/Rockville, MD; 5) MGMGB NHGRI / NIH, Bethesda, MD.

One of the standard ways to validate an association study for biologically meaningful findings is to replicate the findings in an independent dataset. Although replication datasets may be available for the primary phenotypes in a given study, it is likely that many of the findings for phenotypes of secondary interest will not be reported due to the lack of available resources for replication data and/or functional validation. In this study, we investigate an alternative analysis approach using complimentary pairs stability selection that may be viewed as an approximation to replication. Specifically, the sample is randomly split into two halves, which are conditionally independent, and a simple regression-based GWAS is performed on each half. A SNP is defined as corroborated if its p-value is found to be significant in both halves of the data, similar to seeking significance in discovery and replication data. This procedure is performed for many random splits of the data, and SNPs that are corroborated in a high proportion of the random splits are considered as candidates for association with the phenotype. Since only one half of the samples are analyzed at a given time, a loss in power is expected with this type of approach. Due to the conditional independence of the two halves of the data, the square of the significance threshold characterizes the nominal type I error rate (rather than the significance threshold in a standard GWAS); this allows for the use of a less stringent significance threshold that counters the loss in power due to the smaller sample size. The Trinity data is comprised of 2,232 healthy young students at Trinity College Dublin in Ireland with 3 primary and 35 secondary interest quantitative phenotypes. Genotyping with an Illumina 1M HumanOmni1-Quad chip was performed at the Center for Inherited Disease Research (CIDR). A simulation study was performed to evaluate type I error and power of this approach using the Trinity genotypes and simulated phenotypes that followed either genetically null or simple single SNP models. Simulation results demonstrated that this approach maintains the same power as the standard GWAS while reducing the type I error. We analyzed all phenotypes from the Trinity data with this method. In order to validate this method, the primary phenotype results were compared and found consistent with those obtained using independent replication data and/or functional studies.

1330T

Using the Coriell Personalized Medicine Collaborative to illustrate the challenges in translating GWAS results to clinical care. *L. B. Scheinfeldt, N. P. Gerry, M. F. Christman.* Coriell Institute for Medical Research, Camden, NJ.

Several factors contribute to health-related quality of life, including: healthcare quality and access, behavior, environment, and genetics. Genetic information contributes through research that focuses on the underlying biology and inherited component of disease and implementation that applies research knowledge in a clinical setting to improve health outcomes. Clinical genetic testing for Mendelian disorders is standard of care in many cases; however, it is less clear to what extent and in which situations clinical genetic testing will improve prevention, diagnosis and/or prognosis of complex disease. One of the challenges in implementing clinical genetic testing in the management of complex disease is that much of the research reported to date relies on tag SNPs to act as proxies for assumed underlying functional variants that are not yet known. This assumption is especially problematic when reported studies do not include representative population samples, and the results are generalized. Here we use the Coriell Personalize Medicine Collaborative to evaluate how well the reported genetic risk variant for coronary artery disease (rs1333049) tags surrounding variants in the 1000 genomes data across 26 continental population samples. We evaluated variants within 4 distance windows, 5kb, 10kb, 50kb, 100kb, and performed one million simulations for each distance window where we randomly assigned a 'functional' variant within each window that was in high linkage disequilibrium (R2≥0. 8) with rs1333049 in the CEU. We evaluated how often this simulated functional variant was in high LD (R2≥0. 8) with rs1333049 in all of the other population samples. Asexpected, we found good replication in the FIN and GBR; however, we only correctly tagged the simulated functional variant 56% of the time in the IBS and 19% of the time in the TSI. Results were similarly low or lower in other continental groups (2-53%). Our results indicate a relatively large error rate when generalizing increased genetic risk (using rs1333049) of coronary artery disease across diverse population samples, even when generalizing within continents. Taken together, our results highlight the importance of including diverse populations in GWAS. Future work focused on identifying functional variants will eliminate the need for tag SNPs; however, until functional variants are known, caution should be used in the interpretation of genetic risk for complex disease using tag SNPs.

.

1331F

Identification of a possible susceptibility locus for UVB-induced skin tanning phenotype in Korean females using genome-wide association study. *YA. Shin[2], TJ. Kwak[1], YH. Chang[1], JM. Shin[3], JH. Kim[2], SK. Lim[3], SH. Lee[1], MG. Lee[4], TJ. Yoon[5], CD. Kim[3], JH. Lee[3], JS. Koh[6], YK. Seo[6], MY. Chang[1], Y. Lee[3].* 1) LG Household and Healthcare, Daejeon, Korea; 2) Theragen-Etex Bio Institute, Advanced Institute of Convergence Technology, Kwanggyo Technovalley, Korea; 3) Departmentof Dermatology, School of Medicine, Chungnam National University, Daejeon, Korea; 4) Departmentof Dermatology, School of Medicine, Yonsei University, Seoul, Korea; 5) Departmentof Dermatology, School of Medicine, Gyeongsang National University, Jinju, Korea; 6) Dermapro skin research center, Seoul. Korea.

Human skin pigmentation is regulated by various factors including genetic, environmental and endocrine components that modulate the amount, type and distribution of melanin. Among these factors, genetic contribution is one of the most potent determinants of the skin pigmentation phenotype, which displays polygenic quantitative traits. To investigate the genetic factors influencing ultraviolet (UV)-induced skin pigmentation, A multistage genome-wide association (GWA) analysis was conducted in Korean females after UV exposure. Previously, GWA studies evaluating ~500,000 single nucleotide polymorphisms (SNPs) in 99 Korean females identified eight SNPs that were highly associated with tanning ability. To confirm these associations, we genotyped the SNPs in an independent replication study (112 Korean females). We found that a novel intronic SNP yielded significant replicated associations with skin tanning ability. To understand the functional consequences of this locus located in the non-coding region, we investigated the role of this gene in human melanocytes using a recombinant adenovirus expressing a microRNA specific. Inhibition of gene expression significantly increased the expression and activity of tyrosinase in human melanocytes. Taken together, our results suggest that the genetic variants could be determinants in the UV-induced tanning ability of Korean females. The gene represents a new candidate gene to evaluate the molecular basis of the UV-induced tanning ability in individuals.

1332W

Joint analysis of multiple traits using optimal principal components of heritability. *Z. Wang, Q. Sha, S. Zhang.* Mathematical Sciences, Michigan Technological University, Houghton, MI.

The joint analysis of multiple traits has recently become popular since it can increase statistical power to detect genetic variants, and there is increasing evidence showing that pleiotropy is a widespread phenomenon in complex diseases. Currently, several statistical methods have been proposed to jointly analyze multiple traits in genome-wide association studies (GWAS). However, each of these existing methods has good performance only on some specific scenarios, and no method has acceptable performance in all scenarios. In this paper, we proposed an Optimal Principal Components of Heritability (PCH-O) method to jointly analyze multiple traits in GWAS. We used extensive simulation studies to evaluate the performance of PCH-O and compared the power of PCH-O with some existing methods. Our simulations showed that PCH-O has correct type I error rates. Power comparisons showed that PCH-O either is the most powerful one or has comparable power with the most powerful one. Thus, PCH-O has good performance in a wide range of scenarios.

1333T

Privacy Leaks in Quality Control for Meta-analysis and Effective Countermeasures. *W. Xie[1], M. Kantarcioglu[2], J. C. Denny[3,4], T. L. Edwards[5], N. J. Cox[6], B. A. Malin[1,3].* 1) Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN; 2) Department of Computer Science, University of Texas at Dallas, Richardson, TX; 3) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 4) Department of Medicine, Vanderbilt University, Nashville, TN; 5) Division of Epidemiology, Department of Medicine, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 6) Division of Genetic Medicine, Vanderbilt University, Nashville, TN.

Quality control (QC) is fundamental to genome research. This is especially the case for genome-wide association studies (GWAS) conducted via meta-analysis in consortia, where data are contributed by disparate and heterogeneous cohorts. Often, QC on GWA meta-analysis proceeds by sharing and contrasting summary statistics beyond their originating cohorts, under the belief that such summaries are privacy-preserving and ethically safer to disclose than individual-level genome. This work challenges such a belief by pinpointing privacy vulnerabilities of standard QC practices that could leak sensitive information of individual participants. In particular, by systematically assessing privacy implications of various QC procedures, we empirically demonstrate that private traits (e. g. , disease status or study participation) can be ascertained with high accuracy leveraging QC summary statistics such allele frequency, effect size, effect direction, etc. We later demonstrate how to mitigate the privacy risks through novel technological methods based on distributed computing and Cryptography, and present an end-to-end QC pipeline that guarantees genome privacy. Empirical evaluations on several consortia studies show that our secure pipeline enhances privacy effectively, and incurs negligible overhead in terms of runtime and monetary cost. We believe that our privacy assessment and proposed protections are highly relevant to future genome research, in which genome privacy is increasingly a concern due to the formation of even larger collaborative studies/consortia and inclusion of even more cohorts on a national or global scale.

1334F

Bayesian large-scale regression with GWAS summary statistics. *X. Zhu[1], M. Stephens[1,2].* 1) Department of Statistics, The University of Chicago, Chicago, IL; 2) Department of Human Genetics, The University of Chicago, Chicago, IL.

Recent work has revealed potential merits of performing multiple-SNP analysis. Existing multiple-SNP methods require individual phenotypes and genotypes, which are not available for large GWAS. In contrast, single-SNP association summary statistics are often released in public domain. Here we present a new Bayesian regression method for large-scale multiple-SNP analysis, using only GWAS summary statistics and estimated population LD from a public reference panel. Our method takes advantage of the fact that, under standard genetic association model, the posterior of multiple-SNP effect sizes depends on the individual-level GWAS data only through the single-SNP summary statistics and the cohort sample LD. We estimate the unknown cohort LD by its posterior mean given the public panel. Using (singular) Wishart distribution theory, we prove that the estimated cohort LD is actually an estimated population LD, and we use a shrinkage estimator based on genetic maps (Wen and Stephens, 2010) to facilitate fast computation. We approximate the likelihood by a multivariate normal distribution, and derive its connection with previous work, including BIMBAM (Servin and Stephens, 2007), ABF (Wakefield, 2009), CAVIAR (Hormozdiari et al, 2014) and LDSC (Bulik-Sullivan et al, 2015). We specify two types of flexible prior on the effect sizes, and develop MCMC algorithms for posterior inference. We consider three applications of our method: fine mapping, detecting genome-wide associations, and estimating the proportion of phenotypic variation explained (PVE) by available genotypes. Simulations show that our method is robust to a wide range of genetic architectures with varying proportion of causal variants and amount of true PVE, and comparable to existing tools that require individual-level data. We illustrate the method on the summary statistics of 1,064,575 SNPs from 253,288 individuals in a GWAS of human height (Wood et al, 2014). Our estimated total PVE is 60. 4% (95% C. I. [54. 9%, 65. 5%]), and we observe a significant correlation between the estimated PVE of each chromosome and chromosome length (R2=0. 779, p=5. 5e-8). We include 384 previous GWAS significant SNPs identified by GCTA-COJO (Yang et al, 2012) in our analysis, and detect 125 of them with strong evidence (posterior inclusion probability > 0. 9). We also identify that 314/360/371 of these 384 previous hits are located in 10/20/40-Kb regions that show strong signal (posterior expected number of included SNPs >= 1).

1335W

Genetic determined lipoprotein cholesterol levels and age-related macular degeneration. *C. Y. Cheng[1,2,3], Q. Fan[1], E. S. Tai[4,5], G. Cheung[1,2], X. Sim[5], C. C. Khor[6], Y. Y. Teo[4], K. Park[7], C. Pang[8], N. Yoshimura[9], T. Y. Wong[1,2,3].* 1) Singapore Eye Research Institute, Singapore National Eye Centre, Singapore; 2) Duke-NUS Graduate School of Medicine, Singapore; 3) Department of Ophthalmology, National University of Singapore, Singapore; 4) Department of Medicine, National University of Singapore, Singapore; 5) Saw Swee Hock School of Public Health, National University of Singapore, Singapore; 6) Division of Human Genetics, Genome Institute of Singapore, Singapore; 7) Department of Ophthalmology, Seoul National University Bundang Hospital, Korea; 8) Department of Ophthalmology and Visual Sciences, The Chinese University of Hong Kong, Hong Kong; 9) Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, Japan.

The nature and underlying mechanisms of associations between plasma lipoprotein cholesterol levels and the risk of age-related macular degeneration (AMD) are unclear. Using a genetic approach, we aimed to investigate whether genetic determined plasma lipoprotein cholesterol levels were associated with an increased risk of neovascular AMD in East Asians. We assessed genetic variants influencing plasma lipoprotein cholesterol levels in 23,520 East Asians from the Asian Genetic Epidemiology Network (AGEN). The diagnosis of neovascular was made based on clinical examinations and retinal photographs. We identified 62 cholesterol-associated common SNPs, including 31 SNPs for high-density lipoprotein (HDL) cholesterol, 16 for low-density lipoprotein (LDL) cholesterol, and 15 for triglycerides. Using these cholesterol-associated as instrumental variables, we tested the association between a change in genetically determined lipoprotein cholesterol levels of 1 standard deviation (SD) with the risk of AMD in 2,219 AMD cases and 5,275 controls from the Genetics of AMD in Asians (GAMA) Consortium. The association was quantified by combining the instrument variable estimators of each SNP, using fixed-effects models with inverse-variance weighting. We found that higher plasma HDL cholesterol was associated with an increased risk of neovascular AMD (odds ratio [OR] = 1. 48 per SD increase in plasma HDL levels; P = 0. 002) using fixed-effects models. After removing *CETP* (MIM 118470) rs2303790 (which was previously identified as AMD risk locus in East Asians) from the analysis, there was no more significant association between HDL levels and the risk of neovascular AMD (D442G; OR = 1. 31, P = 0. 375). LDL cholesterol and triglycerides levels did not increase the susceptibility to neovascular AMD (OR = 1. 11, P = 0. 724; and OR =1. 04, P = 0. 760, respectively). The moderate association between genetically determined HDL levels and an increased risk of neovascular AMD in East Asians was mainly driven by the *CEPT* D442G variant. Genetically determined LDL cholesterol and triglycerides do not affect the risk of neovascular AMD.

1336T

Detecting allele specific expression in large scale genetic expression studies using overdispersed Poisson linear models. *G. Gliner[1], Y. Park[2], C. Brown[2], B. E. Engelhardt[3].* 1) Operations Research and Financial Engineering, Princeton University, Princeton, NJ, USA; 2) Department of Genetics, Perelman School of Medicine University of Pennsylvania, Philadelphia, PA 19104, USA; 3) Computer Science Department and Center for Statistics and Machine Learning, Princeton University, Princeton, NJ 08540, USA.

Non-coding cis-regulatory mechanisms underlying variation in gene expression levels often show signatures of differential expression of the alleles at coding polymorphisms within the gene, a phenomenon known as allele specific expression (ASE). Detection of significant association between non-coding regulatory variants and gene expression levels (eQTLs) has proven to be a powerful method to study the genetic origins of phenotypic variation. Furthermore, high-resolution genotype and RNA-seq study data motivates the development of systematic statistical models that identify and quantify ASE on a genome-wide scale. To search for significant differential expression levels at bi-allelic single nucleotide polymorphisms (SNPs), current methods test the relative abundance of expression levels of two coding alleles at a coding SNP in each heterozygous sample, where the null model is that the two alleles have equal levels of expression. We present a Bayesian ASE detection method that applies an overdispersed Poisson linear model (OPLM) to recover coding SNPs that show evidence of ASE using phased genotype and mapped RNA-seq data. The OPLM identifies ASE by quantifying the imbalance in gene expression levels between reference and alternative alleles at a given coding locus. This statistical model quantifies the statistical significance of deviations from the null hypothesis of no ASE. While most state-of-the-art methods control for the effects of RNA-seq mapping bias, they only search for ASE locally. We extend the OPLM framework to include all non-coding genetic variants in a gene as candidates for ASE in a single model, combining evidence for ASE across loci and homozygous samples to better control for mapping bias and share statistical support. To empirically validate our method, OPLMs to extract ASE, or OPLEASE, we apply it to publicly available data from the GEUVADIS consortium and compare results to leading ASE detection methods RASQUAL [Kumasaka et al. 2015] and Battle's ASE method [Battle et al. 2014]. We apply it to the Genotype Tissue Expression (GTEx) consortium v6 data to identify ASE loci across tissues. We find that our approach is highly complementary to related methods by improving ASE detection when multiple heterozygous loci are available in a gene and by providing a model to interpret the allele-specific effects of the collective cis-eQTLs on gene transcription by sharing effects across genetic loci, genes, and individuals.

1337F

A unified framework of association mapping for admixed samples. *Y. Guan.* Pediatrics and Molecular & Human Genetics, Baylor College of Medicine, Houston, TX.

Admixed samples provide great opportunities in disease association mapping. First, the variation of local ancestry across individuals provide "free" genetic variates to be associatd with disease phenotypes. Second, the local ancestry of each marker serves as a background forgenotypes to have differential effects conditioning on a specific ancestry. Our working hypothesis is that an allele can have positve, zero, and negative effect sizes toward a phenotype depending on its ancestral background. Thus incorportating the local ancestry to test for genetic association evidently increase power when the allele does have differential effect sizes. Current literature separates the mapping by admixture and mapping by marker, and ignores the possibility of differential effects. We developed a unified framework for association mapping for admixed samples. Our approach based on a highly efficient, model-based, local ancestry inference method ELAI. The statistical model underlying ELAI is a two-layer cluster model where each layer represents ancestral hapltoypes of different age. Our statistical model links the phenotype with genetic markers and all cluster loadings (both old and young haplotypes). If jointly significant, we further disect association signals into genetic markes and old and young cluster loadings. If genetic marker is significant, it is significant after controlling for local ancestry. If old cluster loadings are significant, it is signal for mapping by admixture. If young cluster loadings are significant, it is signal for haplotype associations after controlling for local ancestry. BecauseELAI automatically incorporate training samples, the imputation based association mapping can be performed simultaneously -- both single-marker test and haplotype test. We illustrate our methodusing two Mexican GWAS datasets.

1338W

Impact of gene-environment correlation on genome-wide analysis. *C. Chen[1,2], J. Smoller[1,2], B. Neale[1,2].* 1) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA.

Gene-environment correlation (rGE) refers to the phenomenon of genetic influences on environmental exposure. For diseases with strong environmental risk factors, rGE may play an important role in discover genetic effects, direct or mediated through environmental factors. We investigate the impact of rGE on genome-wide analyses including association tests, SNP-based heritability (h2g) estimation, and stress-diathesis model (SDM) analysis. SDM analysis refers to detecting the induced inverse relationship between genetic risk loading and environmental risk loading among disease cases under the liability-threshold model. To examine the impact of rGE, we use simulated and real genome-wide genotype data, and simulated phenotypes and environmental exposures under the polygenic model with given h2g and prevalence (for binary variables). The simulations were done in 3 settings: 1) causal genetic variants and environmental exposures affect disease risk independently (no rGE); 2) causal genetic variants affect disease risk only through their influences on environmental risk exposures (complete rGE); 3) genetic variants affect disease risk both directly and through effects on environmental risk exposures (partial rGE)We applied genome-wide analyses on simulated datasets with different modeling strategies. We showed that conditioning on environmental factors will remove the genetic effects through rGE and reduce power to detect genetic risk variants but increase power when these is no rGE. We also showed that conditioning on environmental factors will lower the h2g estimates when rGE is present, but increase h2g estimates when there is no rGE. Finally, we examined the SDM analysis under different simulation settings. We used a polygenic risk score as the measure of genetic loading for disease risk. We demonstrated that the expected inverse relationship between disease genetic loading and environmental loading will not always hold under different combinations of rGE and direct genetic effects. In conclusion, potential rGE should be considered when conditioning on environmental risk factors in genome-wide association tests and h2g estimation. The SNP-based heritability of the environmental factors can be a measure of genome-wide rGE and potentially inform on the impact of genome-wide analysis.

1339T

Assortative mating causes biases in SNP-heritability estimates. *R. Tahmasbi[1], T. de Candia[1,2], M. Jones[1,2], D. Evans[4], G. Carey[1,2], L. Eaves[3], M. Keller[1,2].* 1) Institute for Behavioral Genetics (IBG), University of Colorado, Boulder, CO; 2) Department of Psychology and Neuroscience, University of Colorado, Boulder, CO; 3) Departments of Human Genetics and Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Richmond VA; 4) Diamanti Institute, University of Queensland, Australia.

Assortative mating occurs when individuals mate with people similar to themselves on a given trait. Assortative mating is known to lead to downward biases in estimated heritability in twin studies, but its effects on heritability estimated from genome-wide similarity at measured SNPs (SNP-heritability) in unrelated individuals have not yet been quantified. Here, we derive the expected biases in SNP-heritability under assortative mating, and show that these expectations match those observed in simulations. In general, assortative mating causes long-range statistical associations between causal variants (gametic phase disequilibrium), which alters the phenotypic (co)variance matrix in a way that is unaccounted for in the genomic relationship matrix. This causes SNP-heritability to be biased upwards as a function of the degree of mate similarity and the equilibrium heritability of the trait. For highly heritable traits (h2~. 60-. 80) with modest spousal similarity (r~. 20-. 30), such as height or IQ, we find that the estimated SNP-heritability is expected to be ~20% to 30% higher than the true equilibrium SNP-heritability. We discuss several unsuccessful attempts to mitigate this bias. Finally, we describe several signatures of estimation bias due to assortative mating and demonstrate evidence of these using height as a test phenotype. We conclude that, to the degree that assortative mating has occurred, SNP-heritability estimates are expected to be biased upwards.

1340F

Multi-Trait Analysis of Genomic heritability and genomic correlations in the Training Interventions and Genetics of Exercise Response (TIGER) study. *A. I. Vazquez[1], H. Kim[2], J. R. Fernandez[3], M. S. Bray[4].* 1) Epidemiology and Biostatistics, Michigan State University, Lansing, MI; 2) Biostatistics, University of Alabama at Birmingham, Birmingham, AL; 3) Nutrition Sciences, University of Alabama at Birmingham, Birmingham, AL; 4) Nutritional Sciences, University of Texas at Austin, Austin, TX.

While the incidence of obesity continues to rise, fat accumulation and sedentary life are established risk factors for many chronic diseases. Though the genetic basis of obesity and body composition has been well-established, current approaches to gene discovery have explained a very small fraction of the variability in these traits. Alternatively, by simultaneously regressing all available molecular markers, we can explain a sizeable proportion of the inter-individual differences observed in these complex traits because small-effect markers not captured by GWAS can be taken into account. We estimated SNP-derived genomic heritabilities (GH) of body composition traits. Additionally, since these traits are highly correlated, we further differentiated phenotypic correlations from genomic and environmental correlations (GC, EC) observed between these traits. We used 991 non-Hispanic white (W) and 701 African American (AA) young adults genotyped using the MetaboChip (Illumina, Inc.). Measures of BMI, lean mass (Lean), fat mass (Fat) and percent fat (%Fat) were obtained. GH, GCand ECamong variables were estimated though bivariate analysis in a Bayesian context. GHestimates and standard deviations (in parentheses) for each trait were: BMI=0. 23(0. 09); Lean=0. 31(0. 09); Fat=0. 18(0. 07); %Fat=0. 32(0. 07) in Whites, and BMI=0. 18(0. 07); Lean=0. 27(0. 08); Fat=0. 21(0. 08); %Fat=0. 35(0. 09) in Blacks. These estimates represent the amount of inter-individual variation attributable to markers on the MetaboChip. The GCand EC, respectively, between BMI and other traits were as follows: Lean=0. 19 (0. 08) and 0. 32 (0. 08), Fat=0. 15 (0. 06) and 0. 69 (0. 06), and %Fat=0. 03 (0. 06) and 0. 17 (0. 06) in W. In AA, the estimated correlations with BMI and related traits were Lean=0. 11 (0. 07) and 0. 38 (0. 07), Fat=0. 20 (0. 08) and 0. 73 (0. 08), and %Fat=0. 05 (0. 06) and 0. 09 (0. 06). Genomic correlations result from pleiotropic effects of SNPs related to more than one trait. Our estimates of GHin BMI explain more than seven times the variability explained by all markers identified through GWAS. 11-20% of co-variability in BMI and fat and lean mass was accounted for by shared genetic effects, while shared genes jointly accounted for only 3-5% of co-variability in BMI and %fat, suggesting non-pleiotropic effects between genes influencing these two traits. Our results support the hypothesis that genes related to body mass are distinct from those encoding adiposity.

1341W

Bivariate Analysis and Prediction of AMD progression Using Genetic Scores. *Y. Ding[1], Q. Yan[2], Y. Liu[1,2], L. G. Fritsche[3], G. R. Abecasis[3], A. Swaroop[4], E. Y. Chew[4], D. E. Weeks[5,1], W. Chen[2,1,5].* 1) Biostatistics, University of Pittsburgh, Pittsburgh, PA; 2) Pediatrics, University of Pittsburgh, Pittsburgh, PA; 3) Biostatistics, University of Michigan, Ann Arbor, MI; 4) National Eye Institute (NIH/NEI), DC; 5) Human Genetics, University of Pittsburgh, Pittsburgh, PA.

Age-related Macular Degeneration (AMD), the leading cause of blindness in the developed world, accounts for more than 50% of all blindness in the United States. Multiple large-scale genetic studies have had remarkable successes in identifying disease-susceptibility genes for AMD. However, the genetic causes for disease progression have not been well studied yet. In this National Eye Institute (NEI) funded research project, using the data from two large multi-center randomized clinical trials -- AREDS (Age-Related Eye Disease Study) and AREDS2, we perform a novel analysis to evaluate the effect of the known AMD risk SNPs on disease progression and establish prediction models for AMD progression. Specifically, we calculate the time-to-advanced AMD (either choroidal neovascularization or geographic atrophy) from the baseline visit time for each eye of each patient, and model this eye-level progression time using a bivariate approach, which appropriately accounts for between-eye correlations. For the top SNPs that are significantly associated with AMD progression, we additionally model their effects on transitions between multiple disease progression states through a multistate Markov model, where the transitions among four AMD states are examined. Finally, we develop prediction models on AMD progression based on the genetic score (i. e. , the weighted effects of the known AMD risk variants), demographic factors, and eye-level macular information. We find that the model with genetic score and demographic factors (but without macular information) predicts the progression well, suggesting an effective and efficient screening process which uses only genetic and demographic factors to identify patients at high risk of progression for early interventions.

1342T

A novel gene-based analysis method based on MB-MDR. *R. Fouladi[1, 2], C. Schurmann[3], K. Bessonov[1, 2], JP. Vert[4,5,6], R. J. F Loos[3], K. Van Steen[1, 2].* 1) Systems and Modeling Unit, Department of Electrical Engineering and Computer Science (Montefiore Institute), University of Liège, Quartier Polytech 1, Allée de la Découverte 10, 4000 Liège, Belgium; 2) Systems Biology and Chemical Biology, GIGA-R, University of Liège, Quartier Hôpital, Avenue de l'Hôpital 11, 4000 Liège, Belgium; 3) Icahn School of Medicine at Mount Sinai, New York, NY; 4) MINES ParisTech, PSL-Research University, CBIO-Centre for Computational Biology, 35 rue St Honoré, 77300 Fontainebleau, France; 5) Institut Curie, 75248 Paris Cedex ,France; 6) INSERM U900, 75248 Paris Cedex, France.

Here we present a novel gene-based test that builds upon Model-Based Multifactor dimensionality reduction (MB-MDR). This approach was initially developed to investigate SNP-based interactions from GWAS data. It relies on a natural organization of individuals in multi-locus genotype categories, followed by a labelling of these categories using trait information and appropriate association tests. In a gene-based setting, any set of features (e. g. , SNPs, structural variants, epigenetic markers) that can be mapped to a gene is submitted to a clustering algorithm to find groups of individuals (clusters) with similar "gene-based" profiles. The resulting profile types for a gene can be considered to be category levels of a gene-based summary variable. These categories can be labeled by MB-MDR, as before. Earlier, we obtained promising results when taking rare and common variants as features and genes as regions of interest derived from exome sequencing data from GAW17. Here we show the utility of the approach on common variants derived from GWAS data, provided optional parameters are optimized to the new analysis context, including those related to kernel selection prior to clustering. In particular, we explore the pros and cons of using diffusion kernels as a special case of exponential kernels based on the heat equation. This allows the incorporation of biological information via defining a meaningful graph structure on the input features to each gene. In addition, special attention is given to the automatic detection of the optimal number of aforementioned clusters and the optimal minimal cluster size. The newly proposed gene-based tool is evaluated on synthetic data via extensive simulations. We furthermore show its practical use on real-life data from the BioMe Biobank (Mount Sinai school of medicine, USA) for Type II diabetes.

1343F

Identification of significant genetic variants via SLOPE and Group SLOPE. *A. Gossmann[1,3], S. Cao[2,3], Y. -P. Wang[2,3].* 1) Department of Mathematics, Tulane University, New Orleans, LA; 2) Department of Biomedical Engineering, Tulane University, New Orleans, LA; 3) Center for Bioinformatics and Genomics, Tulane University, New Orleans, LA.

The method of *Sorted L-One Penalized Estimation*, or SLOPE, is a novel sparse regression method for model selection introduced in a sequence of recent papers by Bogdan, van den Berg, Sabatti, Su and Candes (Bogdan *et al* 2013 arXiv, Bogdan *et al* 2014 arXiv, Candes and Su 2015 arXiv). It estimates the coefficients of a linear model that possibly has more unknown parameters than observations. In a sense, the SLOPE method combines the ideas of the LASSO method (Tibshirani 1994 Journal of the Royal Statistical Society) and the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995 Journal of the Royal Statistical Society). In many settings SLOPE is shown to successfully control the false discovery rate (the proportion of the irrelevant among all selected predictors) at a user specified level. We evaluate its performance on realistic simulated genetic data, and show its superior feature selection capabilities over LASSO and similar penalized regression methods. Often in genetic data sets, group structures among the predictor variables are given as prior knowledge, such as SNPs in a gene or genes in a pathway. Following this motivation we extend SLOPE in the spirit of Group LASSO (Yuan and Lin 2006Journal of the Royal Statistical Society) to *Group SLOPE*, a method that can handle group structures between the predictor variables, which are ubiquitous in real genetic data. Our simulation results show that the proposed Group SLOPE method is capable of controlling the false discovery rate at a specified level, and its superior variable detection capabilities over Group LASSO.

1344W

Allele Specific Expression Can Substantially Limit Genotype/Phenotype Associations. *J. Dannemiller.* Psychology, Rice University, Houston, TX.

Allele Specific Expression (ASE) can significantly alter quantitative traits. *Purpose:* To derive expressions for and simulate the effects of ASE on the percentage of phenotypic variance explained (PVE) by a single, hypothetical protein-coding SNP. *Methods:* Statistical expressions were derived to estimate the impact of ASE on PVE. Gillespie's (1976) *Monte Carlo* method of simulating transcription and translation was also used as a check on the derived expressions with biologically realistic rates for mRNA transcription and degradation and for protein translation and degradation. ASE was assumed to be produced by a single bi-allelic, regulatory SNP (T/C for convenience) in *linkage equilibrium* with the coding SNP (A/G for convenience) that produced phenotypic variation. All [C/T,A/G] haplotypes were assumed present in the population. It was assumed that the T allele at the regulatory locus produced ASE, and that the G allele at the coding locus produced a protein that was 25% more biologically active than the protein produced by the A allele. Both loci were assumed to be independently in H-W equilibrium with a minor allele frequency (MAF) of 0. 5 for the regulatory SNP and 0. 2 and 0. 5 for the coding SNP (in separate experiments). The expression rate for any coding allele in *cis* with a T allele was randomly drawn either from a lognormal ASE density distribution or from an empirical distribution of ASE ratios modeled on the data in Valle et al. (2008) for the gene *TGFBR1*. The phenotype was the product of the final number of protein molecules times the activity level of each of the protein molecules (1. 00 or 1. 25) summed across the products of the two alleles and averaged across 200 cells for each of 600 subjects per experiment. With no ASE, PVE was varied using phenotype-perturbing, additive Gaussian noise from near 0% (polygenic trait) to 100% (monogenic, Mendelian trait). The impact of ASE was defined as the difference in the PVEs with and without ASE for a fixed level of noise. *Results:* As the PVE *without* ASE increased from near 0% to 100%, the PVE *with* ASE increased but quickly reached asymptote near 12%. When the PVE by the coding SNP was set at approximately 5%, adding the genotype of the regulatory SNP as a predictor still only doubled the PVE to near 10%. *Conclusion:* ASE and potentially other regulatory polymorphisms may limit severely genotype/phenotype associations because of the apparently random nature of the allelic expression ratios across subjects.

1345T

Improved fine mapping using generalized likelihood ratios. *W. Li[1,2,3], L. Strug[1,3].* 1) Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada; 2) The Center for Applied Genomics, Hospital for Sick Children, Toronto, Canada; 3) Program in Genetics and Genome Biology, Research Institute, Hospital for Sick Children, Toronto, Canada.

As genomic dataset continues to grow, so do the number of statistical tests required to fully interrogate them. Statistical methodologies with more sensible solutions to the multiple hypothesis testing problem are required. The Evidential Paradigm (EP) measures statistical evidence using the likelihood ratio at two *simple* hypotheses. EP is attractive for use in 'big data' applications because it decouples statistical evidence from error probabilities, and as a consequence does not require conventional multiple test adjustments (Strug & Hodge, 2006). Although the simple vs. simple hypothesis set-up is applicable to many study designs, some research questions such as genetic association or bioequivalence trials are naturally formulated as composite hypotheses. To extend EP to accommodate composite hypotheses (EPC), Bickel (2012) and Zhang & Zhang (2013) proposed the Generalized Likelihood Ratio (GLR). We analytically derive the operational characteristics for the GLR and show that it has small and bounded error probabilities for normal data and fixed-dimensional parametric models asymptotically. Unlike conventional approaches, GLR can glean additional information from the data such as whether there is strong evidence supporting the null hypothesis. We provide guidelines for study planning under EPC, and compare sample size calculations to the Neyman-Pearson approach. The sample size required for a high chance of observing strong evidence in support of the correct hypothesis depends on the size of the null composite hypothesis, and one generally needs larger sample sizes in EPC. We propose the 'London' plot for easy visualization and separation of association evidence across the genome. Using simulation, we show how the data can support either composite hypothesis with increasing sample size, allowing regions of interest to be more clearly demarcated than in Manhattan plots. Furthermore, specifying an interval/composite null hypothesis, instead of a point null, can guard against statistically significant but practically insignificant findings, a problem often seen in p-value based methods. We applied the EPC to a genetic modifier study of Cystic Fibrosis lung disease within a previously reported associated region. The GLR provided confirmatory results at the associated markers, and markers previously showing weak/inconclusive association evidence began to show support for the hypothesis of no association, allowing for improved fine mapping of the locus.

1346F

A new and scaleable Bayesian framework for joint re-analysis of marginal SNP effects. *P. J. Newcombe[1], D. V. Conti[2], S. Richardson[1].* 1) MRC Biostatistics Unit, Cambridge, United Kingdom; 2) Division of Biostatistics, Department of Preventive Medicine, Zilkha Neurogenetic Institute, University of Southern California, US.

Recently, large scale GWAS meta-analyses - accumulating information over tens of thousands of people - have boosted the number of known signals for some traits into the tens and hundreds. However, the availability of many correlated single nucleotide polymorphisms (SNPs) presents analytical challenges and typically variants are only analysed one-at-a-time. This complicates the ability of fine-mapping to identify a small set of SNPs for further functional follow up. We describe a new and scaleable algorithm for the re-analysis of published marginal associations under joint multi-SNP models, in which correlation is accounted for according to estimates from a reference dataset. SNPs which best explain the joint pattern of effects are highlighted via an integrated Bayesian penalized regression framework. Through a realistic simulation study, including an application to 10,000 SNPs, we demonstrate substantial gains in the proportion of true signals among top ranked SNPs (positive predictive value) using our multivariate framework. We also present a real data application to published results from MAGIC (Meta-Analysis of Glucose and Insulin Related Traits Consortium) - a GWAS meta-analysis of more than 15,000 people, in which we re-analyse several top loci associated with glucose levels two hours after oral stimulation. Our algorithm was able to rule out many SNPs as false positives and for one gene, *ADCY5*, joint modeling of the pattern of effects across the locus highlighted an alternative, and more plausible, SNP to the reported index.

1347W

Test rare variants for multiple traits in admixed populations. *X. Wang, X. Zhao, C. Huang.* Joseph J. Zilber School of Public Health, P. O. Box 413, University of Wisconsin-Milwaukee, Milwaukee, WI.

Pleiotropy, effect of one variant on multiple traits, is a widespread phenomenon in complex diseases. Joint analysis of multiple traits such as systolic and diastolic blood pressures evaluated in hypertension can increase statistical power to detect disease susceptible genetic variants. However, testing rare variants for multiple traits in admixed populations such as African Americans and Hispanic Americans is challenging due to the extreme rarity of individual variants, allelic heterogeneity, and the confounding issue of population stratification. Population stratification has long been recognized as an issue in genetic association studies because unrecognized population stratification can lead to both false-positive and false-negative findings and can obscure true association signals if not appropriately corrected. This issue can be even worse in rare variant association analyses because rare variants often demonstrate stronger and potentially different patterns of stratification than common variants. To correct for population stratification for multiple traits in genetic association studies, we proposed two novel methods to test the gene- or pathway-based effect of genetic variants for admixed populations. For the first method, using generalized linear model, we treat the combination of local ancestry weighted variants as response variable and the combination of traits as predictors. For the second method, using generalized linear model, we treat the ancestry-based weighed dosage score as response variable and the combination of traits as predictors. To test association between the multiple traits and the test genomic region, we have developed a score test in the first method and a bootstrap confidence interval test in the second method. We used simulated sequence data to show that our proposed tests have controlled type I error rates, whereas naïve application of existing rare variants tests for multiple traits leads to inflated type I error rates. We will evaluate power of the novel methods in simulation studies and real data analysis.

1348T

Investigation of heterogeneity of genetic effects in severe malaria sub-phenotypes across sub-Saharan Africa, Asia and Oceania. *G. M. Clarke[1], G. Band[1], Q. S. Le[1], C. C. A. Spencer[1], K. A. Rockett[1,2], D. P. Kwiatkowski[1,2], MalariaGEN Consortium.* 1) Kwiatkowski Group, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Warwickshire, UK, United Kingdom; 2) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SA.

Recent studies of genetic association with severe malaria sub-phenotypes have found evidence of heterogeneity of genetic effect across phenotypes suggesting that many studies for the more general severe malaria outcome could fail to identify relevant genetic variants. We use genome-wide data in 11,625 cases of severe malaria and 11,435 population controls in three African populations to examine association between severe malaria sub-phenotypes severe malarial anaemia and cerebral malaria allowing for heterogeneity of genetic effects. We highlight a complex pattern of effects underlying severe malarial anaemia and cerebral malaria at the *G6PD* locus and show that allowing for heterogeneity of effects increases the ability to detect signals of association that would otherwise go unnoticed. Our analyses also allow dissection of signals of association at known severe malaria loci.

1349F

A High-Throughput Platform for Functional Validation of CFTR Variants. *A. A. Sockell[1], A. Rychkova[1], H. A. Costa[1], A. S. Verkman[2], M. C. Bassik[1], C. D. Bustamante[1].* 1) Department of Genetics, Stanford University, School of Medicine, Stanford, CA; 2) Department of Medicine, University of California San Francisco, San Francisco, CA.

Rapid, accurate, and inexpensive genome sequencing promises to transform medical care, but a critical hurdle to enabling personalized genomic medicine is predicting the functional impact of novel genomic variation. Several computational tools currently exist for differentiating “benign” from “pathogenic” variants based on genomic context and evolutionary conservation, but all existing approaches suffer from low overall accuracy rates. As a solution to this problem, we are creating a high-throughput validation platform for directly testing the effect of novel variation in disease genes. We are focusing initially on CFTR, a chloride ion channel commonly mutated in cystic fibrosis. Using both public and internal CFTR variant databases, we have developed a tree-based machine learning algorithm with increased variant classification accuracy compared to current computational tools. With these variant classifications in hand, we are now developing an experimental system for functional validation. First, we are synthesizing a library of CFTR variants, which will be stably integrated into Fischer Rat Thyroid (FRT) cells expressing a halide-sensitive YFP reporter gene. FRT cells resemble the epithelial cells that are primarily affected by cystic fibrosis, but do not express any endogenous cAMP-activated transmembrane channels. To verify the validity of variant function in human tissues, we will also use the CRISPR/Cas9 system to install individual variants into the genomes of human lung epithelial cells. With mutant copies of CFTR installed into these cells, we will determine ion conductance properties for each variant by measuring fluorescence of the halide-sensitive YFP reporter. We will also measure transcript levels, and will take advantage of a surface-expressed epitope included in our CFTR constructs to determine the cell-surface expression for each CFTR variant. Together, we expect that these data will allow for robust functional annotation of a broad panel of CFTR variants, which we will use to validate and refine our computational predictions of variant pathogenicity.

1350W

The Prevalence of Nonclassic Congenital Adrenal Hyperplasia Mutations in the Ashkenazi Jewish Population. *M. S. Elman[1], R. Morissette[1], T. R. Prezant[2], A. Pulver[3], D. P. Merke[1,4].* 1) NIH Clinical Center, Bethesda, MD; 2) Endocrine Sciences, Laboratory Corporation of America® Holdings, Calabasas, CA; 3) Department of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD; 4) The Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD.

Background: Congenital adrenal hyperplasia (CAH) is an autosomal recessive disorder of cortisol biosynthesis most commonly caused by *CYP21A2* mutations (chromosome 6p, HLA-linked) resulting in 21-hydroxylase deficiency. Two major phenotypes exist: classic (severe) and nonclassic (mild, NC-CAH). Classic CAH is part of neonatal screening in the U. S. , occurs in 1 in 15,000 live births, and is life-threatening if not diagnosed and treated in a timely manner. NC-CAH may be asymptomatic but can result in female infertility and pregnancy loss. Neonatal screening does not accurately detect NC-CAH. A study conducted in 1985 using HLA linkage found the prevalence of NC-CAH to be 1 in 1,000 in the general population and more common in certain ethnic groups, such as Ashkenazi Jews (3. 7%), Hispanics (1. 9%), and Yugoslavs (1. 6%) compared to general US Caucasians (0. 1%). We sought to reevaluate the prevalence and carrier frequency of NC-CAH in a representative cohort of Ashkenazi Jews and healthy Caucasians. **Methods:** DNA from healthy Ashkenazi subjects was obtained from The Foundation for Jewish Genetic Medicine, Inc. Johns Hopkins University School of Medicine. Healthy Caucasian DNA was acquired from the Coriell Biorepository. Ethnicity is defined as having four Ashkenazi Jewish grandparents. *CYP21A2* genes were amplified using a gene-specific PCR reaction and analyzed by multiplex mini-sequencing assays for the 12 most common CAH mutations, large gene conversions, deletions, and multiple single nucleotide polymorphisms. Statistics were calculated using the Hardy-Weinberg Law for a population at equilibrium. **Results:** Our results suggest that the 1985 study over-estimated the prevalence of NC-CAH mutations and disease in Ashkenazi Jews. Calculations based on our cohort indicate that 14% of Ashkenazi Jews carry a NC-CAH mutation with 0. 8% affected compared to previous estimates that 31% of Ashkenazi Jews carry a NC-CAH mutation with ~4% affected. Our results also suggest that the life-threatening classic form of CAH is more common in Ashkenazi Jews than in the general population, where the carrier rate is estimated to be 1:60. **Conclusion:** *CYP21A2* genotyping should be considered for prenatal/preconception screening in Ashkenazi Jews, especially females experiencing infertility or miscarriage. Further analysis to clarify potential deletion and duplication genotypes as well as the carrier frequency and prevalence of classic CAH in the Ashkenazi population is underway.

1351T

Risk Prediction Modeling of Sequencing Data Using Random Field. *Y. Wen[1], Q. Lu[2].* 1) Statistics, The University of Auckland, Auckland, New Zealand; 2) Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, USA.

The translation of human genome discoveries into individualized prediction and prevention represents one of the major challenges in the coming decades. With the advance in high-throughput sequencing technology, it is feasible to investigate the role of both common and rare variants in disease risk prediction. While the new technology holds great promise to improve individualized disease prediction, the massive amount of data and low frequency of rare variants pose great analytical challenges on risk prediction. In this paper, we develop a forward random field method (FRF) for risk prediction using sequencing data. Subjects' phenotypes are treated as stochastic realizations of a random field on a genetic space, and an individual's phenotype is predicted by adjacent subjects with similar genetic and environment predictors. The FRF method allows for the specification of various similarity measures and the selection algorithm adopted can determine the best similarity measure and reduce the effect of noise, both of which can improve the robustness and prediction accuracy of the method. Like other random filed based methods, the FRF method avoids the specification of the threshold of rare variants and it allows for multiple variants acting with different magnitudes. Through simulations, we demonstrate the proposed FRF method attains a higher or comparable accuracy over commonly used support vector machine based methods under various disease models. We further illustrate the FRF method with an application to the sequencing data obtained from the Dallas Heart Study.

1352F

Evaluation of stepwise and penalized regression methods in selecting a set of genetic predictors explaining quantitative trait variation. *A. J. M. Sorant, A. F. Wilson.* Genometrics Section, Computational and Statistical Genomics Branch, NHGRI, NIH, Baltimore, MD.

In seeking to understand genetic causes of variation in a quantitative trait, one may attempt to fit a regressive model describing the additive effects of a set of genetic variants acting together on the trait as a dependent variable. There are a number of different methods for selecting the proper set of variants to include in the model, as well as a number of different criteria for defining the best result. In a simulation study we considered various methods of choosing causal variants and judged them according to different criteria. Each simulated sample included 500 single nucleotide variants (SNVs) in linkage equilibrium, each having a minor allele frequency (MAF) at least 0. 01. A quantitative trait was determined by five SNVs with MAF 0. 01 with locus-specific heritabilities of 0. 01, 0. 02, 0. 03, 0. 04 and 0. 05, producing an overall trait heritability of 0. 15 in the theoretical population sampled. A total of 750 unrelated observations with complete data were generated for each of 2000 samples. Each sample was analyzed with traditional stepwise regression using p-value selection criteria of 0. 0001 and 0. 05, stepwise regression with the Bayes Information Criterion (BIC), LASSO regression with cross-validation and three different penalty choices, LASSO regression minimizing BIC and LASSO regression with permutations. Because LASSO methods with cross-validation or permutations involve the use of a random number sequence, each of these methods was performed with three different starting seeds, and results were very similar across seeds, especially for the permutation method. Each method was evaluated in terms of power to detect causal SNVs and type I error rate, as well as model p-value, proportion of trait variance explained, BIC and Akaike Information Criterion (AIC) for the selected model. The "best" method, of course, depended on the evaluation criterion used. Stepwise regression with 0. 05 p-value criterion achieved both the best (highest) power over all causal variants (0. 95) and worst (highest) type I error rate (0. 06). The same method using 0. 0001 criterion had the worst (lowest) power (0. 69) and best (lowest) type I error rate (0. 0001). The other methods had power ranging from 0. 70 to 0. 94 and type I error rate ranging from 0. 0008 to 0. 04. The sets selected by each method were also compared for similarity and overlap. The main difference appeared to be the number of SNVs included in the model. Overlap between sets was otherwise quite good.

1353W

Characterizing power of Adjusted-trait Regression in GWAS with quantitative traits. *P. Yajnik, M. Boehnke.* Biostatistics - School of Public Health, University of Michigan, Ann Arbor, MI.

Multiple linear regression (MLR) is commonly used to test for association between genetic variants and quantitative traits and to estimate genetic effect sizes while controlling for the effects of other variables (covariates). Covariates may be included to account for confounding (due to population structure, batch effects or other variants in linkage disequilibrium), to reduce trait variability (when relevant explanatory variables have been measured), or to negate associations that are driven through the action of the variant on an intermediate phenotype. Modern genetic association studies assay a large number of variants and typically fit separate MLR models for each variant. As the dependent variable and covariates are the same across all models, many analysts use a two-stage approach. In the first stage, an 'adjusted' trait is obtained as the residuals from the regression of the trait on the covariates. In the second stage, simple linear regressions are performed to test for association between the adjusted trait and each genetic variant. Analysts may prefer this two-stage adjusted-trait regression (ATR) over MLR since it is faster and makes data-management easier. However, ATR and MLR results are not equivalent. Previous work has shown that ATR produces null-biased estimates and is less powerful than MLR when the genetic variant is correlated with the covariates by deriving approximate relationships between the ATR and MLR test-statistics. These studies used simulations to assess power of ATR when the acceptable type 1 error rate (α) is 0. 05. We extend these results by deriving the exact relationship between the ATR and MLR test-statistics and study the ATR hypothesis test when $\alpha$is at the stringent levels typically used in GWAS and other large-scale genetic association studies. We show that the discrepancy of power between ATR and MLR increases as $\alpha$is made more stringent. We confirm that ATR is conservative when the correlation between genetic variant and covariates is high but is anti-conservative when the correlation is low and the ratio of covariates to sample size is large. We note that the covariates typically used in GWAS rarely exhibit correlations with variants that are large enough to cause substantial differences between ATR and MLR. However, ATR may exhibit substantial loss of power for variants that are in high linkage-disequilibrium with variants included as covariates during regional conditional analysis.

1354T

Estimates of genetically regulated gene expression and associated genetic predictors explain variance of type 2 diabetes. *J. Torres[1], H. Wheeler[1], E. Gamzon[2], K. Shah[1], N. Knoblauch[3], J. Below[4], C. Hanis[4], D. Nicolae[1], N. Cox[2], H. Im[1], The GTEx Project.* 1) Medicine/Genetic Medicine, The University of Chicago, Chicago, IL; 2) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 3) Committee of Genetics, Genomics and Systems Biology, The University of Chicago, Chicago, IL; 4) Department of Epidemiology, Human Genetics and Environmental Sciences, The University of Texas Health Sciences Center at Houston, Houston, TX.

Expression quantitative trait loci (eQTLs) mapped in insulin-responsive peripheral tissues are not only enriched among the set of variants that associate most strongly with type 2 diabetes (T2D) but they also account for a disproportionate share of disease heritability. This highlights an important role for variants regulating gene expression in the genetic architecture of T2D risk and provides an avenue for improved disease gene mapping. One approach that leverages the plethora of functional genetic data from large-scale collaborative efforts such as the Genotype-Tissue Expression Project (GTEx) is Predicted Expression Scan (PrediXcan), whereby machine learning approaches are used to train models that estimate the genetic component of gene expression for each gene expressed in a tissue. These predicted expression traits are then used as the basis for a gene-based association test that provides tissue-level and directional information. Here, we apply penalized regression to genotype and RNA-seq data for 14 T2D-relevant tissues (including subcutaneous adipose, liver, pancreas, and skeletal muscle) from the GTEx project to identify a set of SNP predictors of tissue-dependent gene expression. We investigated both the T2D heritability accounted for by these predictors in the WTCCC and Starr County, TX, T2D datasets as well as the phenotypic variance explained collectively by the predicted transcriptome – a measure we term "regulability" (in contrast to heritability). We show that SNP sets of predictors of gene expression explain disproportionately high (relative to the number of markers in each SNP set) shares of disease heritability (7-fold enrichment over SNP set proportion for Whole Blood eQTLs) and that regulability estimates account for significant portions of the phenotypic variance of T2D for each of the evaluated tissues (across tissues, the mean proportion of explained variance = 0. 15 ±0. 03 and 0. 38 ±0. 09 for WTCCC and Starr County, respectively). These results support the utility of applying predicted expression to gene mapping and implicate an important role for eQTLs that elicit effects across tissues.

1355F

Genetic Risk Score for Essential Hypertension and Risk of Preeclampsia. *C. Smith[1], A. Saftlas[1], C. Spracklen[1,2], E. Triche[3], A. Bjonnes[4,5], B. Keating[6,7], R. Saxena[4,5], P. Breheny[8], A. Dewan[9], J. Robinson[1], J. Hoh[10], K. Ryckman[1].* 1) Epidemiology, College of Public Health, University of Iowa, Iowa City, IA; 2) Department of Genetics, University of North Carolina-Chapel Hill, 5100 Genetic Medicine Building, CB #7264, 120 Mason Farm Road, Chapel Hill, NC 27599; 3) Department of Epidemiology, School of Public Health, Brown University, 121 S. Main St. , 2nd floor, Box G-S121-2, Providence, Rhode Island; 4) Center for Human Genetic Research and Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, 185 Cambridge Street, CPZN 5, Boston, MA 02114; 5) Program in Medical and Population Genetics, Broad Institute, 415 Main Street, Cambridge, MA 02142; 6) Department of Surgery, Penn Transplant Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; 7) Center for Applied Genomics, Abramson Research Center, The Children's Hospital of Philadelphia, Philadelphia, PA 19104; 8) Department of Biostatistics, College of Public Health, University of Iowa, Iowa City, IA; 9) Department of Chronic Disease Epidemiology, Yale School of Public Health, 60 College Street, Room 403, New Haven, CT, 06520; 10) Department of Environmental Health Sciences, Yale School of Public Health, 60 College Street, New Haven, CT, 06520.

Background: Preeclampsia is a hypertensive complication of pregnancy characterized by novel onset of hypertension after 20 weeks gestation, accompanied by proteinuria. Affecting over 5% of pregnancies, preeclampsia represents a substantial burden to maternal and neonatal health. Epidemiological evidence suggests that genetic susceptibility exists for preeclampsia; however, whether preeclampsia is the result of underlying genetic risk for essential hypertension has yet to be investigated. Based on the hypertensive state that is characteristic of preeclampsia, we aimed to determine if established genetic risk scores (GRS) for hypertension and blood pressure are associated with preeclampsia. **Methods:** Subjects consisted of 162 preeclamptic cases and 108 normotensive pregnant controls, all of Iowa residence. Subjects' DNA was extracted from buccal swab samples and genotyped on the Affymetrix Genome-wide Human SNP Array 6. 0 (Affymetrix, Santa Clara, CA). Missing genotypes were imputed using MaCH and Minimac software. GRS were calculated for hypertension, systolic blood pressure (SBP), diastolic blood pressure (DBP) and mean arterial pressure (MAP) using established genetic risk loci for each outcome. Regression analyses were performed to determine the association between GRS and risk of preeclampsia. These analyses were replicated in an independent US population of 516 cases and 1,097 controls of European ancestry. **Results:** GRS for SBP and DBP were associated with their respective blood pressure measurements, demonstrating that these GRS are valid indicators of genetic contribution to blood pressure. GRS for hypertension, SBP, DBP and MAP were not significantly associated with risk for preeclampsia ($p>0.189$). The results of the replication analysis also yielded non-significant associations between GRS and risk for preeclampsia. **Conclusions:** Genetic risk scores for hypertension and blood pressure are not associated with preeclampsia, suggesting that an underlying predisposition to essential hypertension is not on the causal pathway of preeclampsia. These results suggest that preeclampsia is not an early manifestation of increased genetic risk for chronic hypertension, but rather that it represents a unique hypertensive state. These results also suggest that the increased risk for chronic hypertension following a preeclamptic pregnancy is due to the vascular damage accrued during preeclampsia.

1356W

Isovaleryl-CoA dehydrogenase as a drug target for fibrotic idiopathic interstitial pneumonias. *S. Ross[1], M. Narahara[2,3], N. Soranzo[4,5], J. B. Richards[1,2,6,7].* 1) Centre for Clinical Epidemiology, Department of Epidemiology, Lady Davis Institute for Medical Research, Jewish General Hospital, McGill University, Montreal, Quebec, Canada; 2) Department of Human Genetics, McGill University, Montréal, Québec, Canada; 3) Genome Quebec Innovation, Centre Montreal, Quebec, Canada; 4) Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton, UK; 5) Department of Hematology, University of Cambridge, Cambridge, UK; 6) Department of Medicine, McGill University Montreal, Quebec, Canada; 7) Department of Twin Research and Genetic Epidemiology, King's College London, UK.

Background: Drug development for complex disease has a high failure rate, due in part to a lack of validated drug targets. Human genetics provides the opportunity to identify causal drug targets in humans, and this paradigm is simplified in the case of metabolomics, where the genetic variants influencing metabolite levels are controlled by few, or even a single enzyme. Using fibrotic idiopathic interstitial pneumonias (IIPs) as an exemplar complex disease, we sought to determine whether the genetic variants underpinning IIPs also directly influence metabolite levels, using the largest GWAS datasets available to date. **Methods and Results**: Effect size estimates of common genetic variants associated with serum metabolite levels were obtained from our previously published GWAS of 7,824 Europeans participants. The corresponding effect estimates of these SNPs on the risk of IIPs were obtained from a GWAS of 9,065 participants of European descent. The rs2034650 [A] polymorphism (near *IVD*) was associated with reduced levels of isovalerylcarnitine (b= -0. 0184, 95% confidence interval (CI): -0. 024, -0. 013; P= 7. 5x10-12) and an increased risk of IIP (odds ratio (OR): 1. 30, 95% CI: 1. 19-1. 41; P= 1. 0x10-11). *IVD* is a mitochondrial matrix enzyme that encodes for isovaleryl-CoA dehydrogenase, which is involved in the production of isovalerylcarnitine. Furthermore, the IVD eQTL (rs2034650; P= 1. 0x10-6, n=142) was significantly associated with skeletal muscle, a major source of mitochondrial metabolites. **Conclusion**: Our results demonstrate the potential role of IVD in the etiology of IIP and provide evidence that IVD is a drug target for the treatment of IIP. These results also provide evidence that metabolomics are a rich source of information for validating drug targets in humans.

1357T

Polygenic risk prediction model based on winner's curse correction and multidimensional thresholding. *N. Chatterjee, J. Shi.* National Cancer Institute, Rockville, MD.

BackgroundEstimates of heritability suggest that genome-wide association studies (GWAS) have the potential to improve genetic risk prediction models, but SNPs discovered at the genome-wide significance level typically have low discriminatory power. **Methods**We explore methods for optimal thresholding and weighting SNPs to improve performance of models exploiting heritability in SNPs that do not achieve genomewide significance. We propose simple modifications to the widely used polygenic risk-score (PRS) modeling techniques that required availability of only GWAS summary-level data. First, motivated from the widely used algorithm lasso, we propose a threshold dependent winner's curse adjustment for marginal association coefficients that are used to weight the SNPs in PRS. Second, to exploit various external functional/annotation knowledge that might identify subset of SNPs highly enriched for association signals, we consider using multiple thresholds for SNPs selection based on group membership and identify optimal set of thresholds based on independent validation dataset. We use summary-level results from available from large GWAS of a wide variety of traits and independent validation datasets to assess the performance of these methods. We use a variety of recently available annotation database, expression and methylation trait analysis to identify groups of SNPs that are likely to be enriched with association and evaluate the utility of this information for risk-prediction for respective outcomes based on the proposed methods. We also study performance of the algorithm using simulation studies that incorporate realistic genetic architecture, linkage disequilibrium pattern and enrichment factor for underlying functional SNPs. **Results**Our analysis show that while a simple winner's curse correction uniformly lead to enhancement of performance of the models across traits, incorporation of functional SNPs was beneficial for only selected traits. Compared to standard PRS algorithm, the proposed methods in combination led to substantial efficiency gain (25-50% increase in the value Nagelkerk R2 statistics) for five out of fifteen different diseases we studied. Our simulation studies provided further clarification why differential treatment of certain category of functional SNPs, even when shown to be highly enriched for GWAS-heritability, does not lead to proportionate improvement in genetic risk-prediction due to non-uniform linkage disequilibrium structure.

1358F

Impact of Surrogate Variable Analysis on eQTL Discovery and Application to RNA-Seq Data. *C. Lu[1], L. Klei[2], K. Roeder[1], B. Devlin[2], The CommonMind Consortium.* 1) Department of Statistics, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA; 2) Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA.

Many things affect gene expression, not all of them will have been measured in a genetic study of such expression. For example hidden variables could influence detection of quantitative trait loci (eQTLs) affecting gene expression. Leek and Storey (2007) proposed a statistical model, Surrogate Variable Analysis (SVA), to estimate the variability of gene expression that is accounted for by these hidden or surrogate variables, and in theory this approach should increase the power of identifying eQTLs. Surprisingly we are unaware of a comprehensive study of the performance of SVA. Hence we studied the performance of SVA for eQTL discovery and other features of expression analysis, evaluating its ability to detect and accurately account for such variables. To do so, we conducted simulation studies and applied SVA to the RNA-Seq data of gene expression from the CommonMind Consortium (CMC), which measured expression from >600 human brains. Simulations generated data involving both confounders and biological gene co-expression, all linear effects, and included cis-eQTLs, trans-eQTLs and differential expression according to disease status. We explored a variety of scenarios for hidden variables and gene co-expression, evaluating how the results of SVA compared to the true generating model (e. g. , coefficient estimates, standard error of coefficient estimates, p-values of estimates, gene clustering structure, False Discovery Rate (FDR), power). These simulations help interpret results from CMC RNA-Seq data, especially those from models with and without SVA. We find that SVA models gene co-expression as well as other hidden factors and thus its results can be challenging to interpret. On the other hand in most situations SVA performed well at estimating surrogate variables and did increase the power of eQTL discovery.

1359W

Estimating clinical outcomes and classifying *CFTR* variants of unknown significance in children with a positive newborn screening for Cystic Fibrosis. *D. Conti[1], C. Azen[2], D. Thomas[1], D. Salinas[2].* 1) Preventive Medicine, Univ Southern California, Los Angeles, CA; 2) Pediatrics, Univ Southern California, Los Angeles, CA.

Cystic Fibrosis (CF) is an autosomal recessive disorder caused by a defective trans-epithelial channel called Cystic Fibrosis Transmembrane Conductance Regulator (CFTR). The defective CFTR causes mucus secretions to become thick and viscous, obstructing the lungs and pancreas and consequently causing life-threatening lung infections and malnutrition. Each year approximately 3,700 infants in the US are identified through newborn screening (NBS) to be at risk for developing CF. While approximately 70% of CF cases of European ancestry are due to carrying two copies of the most common variant, F508del, other combinations of any two CF-causing variants can cause disease. With information on over 39,000 patients with CF, the Clinical and Functional TRanslation of CFTR (CFTR2) mutation database has assigned disease liability to 206 of the near 2,000 *CFTR* variants identified. Only 174 have been classified as CF-causing. Thus, there is a great need to predict the clinical outcomes in children with positive newborn screening and to predict the penetrance of variants of unknown significance (VUS). We present a Bayesian hierarchical model, with one level for the phenotypes (e. g. immunoreactive trypsinogen, sweat chloride concentration, pancreatic sufficiency status, *Pseudomonas aeruginosa* colonization), a level for the risk of CF given the disease-causing status of the joint genotype, and a level for the potential for disease-causing of each variant given external information (e. g. genomic annotation, population allele frequency, and CTFR2 database characteristics). We apply our model to data from 848 children with a positive screening test and 2 to 6 year of clinical follow up from over 2 million births from the California Newborn Screening Program. Of the 269 joint genotypes with a VUS, ~7-10% are estimated to have an elevated posterior probability of being classified a CF-causing variant. This translates to ~5% of the 593 individuals carrying a VUS having an elevated posterior probability of having CF. We demonstrate the sensitivity of estimation to various models forms (e. g. with and without phenotype information) and prior specifications (e. g. with and without genomic annotation). We discuss the use of this model to classify previously unobserved VUS, predict clinical status for newborns with a positive screening test, and to impact clinical care protocols.

1360T

Associations of high-risk genetic factors with dementia in the Korean elderly population. *J. J. Lee[1,6], K. Y. Choi[1,2], M. J. Kim[1], N. H. Kim[1,3], J. S. Lee[1,3], S. R. Kang[1,3], Y. Y. Choi[1,2], K. H. Lee[1,3,4,5].* 1) National Research Center for Dementia, Chosun University, Gwangju, South Korea; 2) Department of Premedics, College of Medicine, Chosun University; 3) Department of Life Science, College of Natural Sciences, Chosun University; 4) Department of Biomedical Science, College of Natural Sciences, Chosun University; 5) Department of Marine Life Science, College of Natural Sciences, Chosun University; 6) Department of Computer Science and Statistics, College of Natural Sciences, Chosun University, Gwangju, South Korea.

Dementia is a neurodegenerative disorder that causes cognitive dysfunction and behavioral symptoms. As the aging population in the world increases, the interest in early diagnosis of dementia is becoming a necessity. Genes associated with dementia can be found through Genome Wide Association Study(GWAS) with a large cohort. Although there are many GWAS studies with the Caucasian population, the studies for Asians, in particular Koreans are not sufficient. We have associated the previously reported high-risk single nucleotide polymorphisms (SNPs) with dementia in Korean population. This cohort study was designed with residents aged over 65 years and under 85 years in Gwangju city, Korea. 434 subjects were divided into the normal control group and the dementia group. With highly AD-associated 29 SNPs as reported previously, we have examined SNP genotyping for these subjects by Taqman Assay and Fluidigm using DNA extracted from the blood. We have analyzed the association of each SNP for dementia by logistic regression models using additive model after adjustment of the age, sex, APOE-ϵ4. The associated SNPs were rs11767557 (EPHA1). Odds ratio of rs11767557 is 0. 583 (95% CI:0. 353-0. 961). These results indicate the potential use of some SNPs as genetic markers for determining susceptibility to dementia in Korean population. The further study is necessary to validate the results of this study by using a different Korean elderly population. In addition, a molecular functional study will define the role of genetic polymorphisms in an animal model or *in vitro*.

1361F

A generalized joint location-scale association test for uncertain genotypes and related individuals. *D. Soave[1,2], L. Sun[1].* 1) University of Toronto, Toronto, Ontario, Canada; 2) The Hospital for Sick Children, Toronto, Ontario, Canada.

In genetic association studies, it has been pointed out that a number of biologically meaningful scenarios, including GxG and GxE interactions, can lead to variance (scale) heterogeneity of a quantitative trait across genotype groups, and corresponding scale tests have been developed. Recently, a joint location-scale testing framework was proposed for association analysis of genotyped SNPs in a population sample, combining evidence from both the traditional mean (location) test and the newer scale test to achieve better power [Soave, et al. 2015 AJHG in press]. Genotype uncertainty, however, is inherent in both imputed and sequenced data, and these DNA data are also commonly obtained on samples that include related individuals from families or cryptic relationships. A number of modified location tests have been proposed for these more complex data, but the lack of a generalized scale test remains the bottleneck in broader application of the joint location-scale association test. Among many testing procedures for variance heterogeneity, Levene's test is noted for its simplicity and robustness to departures from normality. Extending the work of Glejser [1969] and Iachine et al. [2010], we propose an easy-to-implement extension to Levene's method that formulates the variance test as a generalized least squares regression problem. Unlike Levene's original test, the regression framework is flexible to directly incorporate the probabilistic data associated with genotype uncertainty and the correlation structure associated with related individuals. Performance of the proposed method, including its validity and enhanced power, is demonstrated through simulations and applications.

1362W

Meta-Analysis for Discovering Rare-Variant Associations:Statistical Methods and Software Programs. *Z. Tang[1], D. Lin[2].* 1) Biostatistics, Vanderbilt University, Nashville, TN; 2) Biostatistics, University of North Carolina, Chapel Hill, NC.

There is heightened interest in using next-generation sequencing technologies to identify rare variants that influence complex human diseases and traits. Meta-analysis is essential to this endeavor because large sample sizes are required to detect associations with rare variants. In this work, we provide a comprehensive overview of statistical methods for meta-analysis of sequencing studies to discover rare-variant associations. Specifically, we discuss the calculation of relevant summary statistics from participating studies, the construction of gene-level association tests, the choice of transformation for quantitative traits, the use of fixed-effects versus random-effects models, and the removal of shadow association signals through conditional analysis. We also show that meta-analysis based on properly calculated summary statistics is as powerful as joint analysis of individual-participant data. In addition, we demonstrate the performance of different meta-analysis methods using both simulated and empirical data. We then compare four major software packages for meta-analysis of rare-variant associations -- MASS, RAREMETAL, MetaSKAT, and seqMeta -- in terms of the underlying statistical methodology, analysis pipeline, and software interface. Finally, we present a software interface, PreMeta, that integrates the four meta-analysis packages and allows a consortium to combine otherwise incompatible summary statistics.

1363T

Non-Parametric Analysis reveals Novel Genetic Associations with variation in Plasminogen Activator Inhibitor-1 levels. *M. J. White[1,2,3], N. M. Kodaman[2,3], R. H. Harder[3], F. W. Asselbergs[4,5,6], D. E. Vaughan[7], N. J. Brown[8], J. H. Moore[3], S. M. Williams[3].* 1) Department of Medicine, University of California, San Francisco, San Francisco, CA; 2) Center for Human Genetics Research, Vanderbilt University, Nashville, TN; 3) Department of Genetics and Institute of Quantitative Biomedical Sciences, Dartmouth College, Hanover, NH; 4) Department of Heart & Lungs, University Medical Center Utrecht, Utrecht, The Netherlands; 5) Institute of Cardiovascular Science, University College London, London, UK; 6) Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, The Netherlands; 7) Department of Medicine, Northwestern University, Feinberg School of Medicine, Chicago, IL; 8) Department of Medicine Vanderbilt University, Nashville, TN.

Background: Cardiovascular disease (CVD) consists of multiple conditions with overlapping environmental and genetic risk factors, symptoms, and etiologies. It is a leading cause of death both in the United States and globally. Plasminogen activator inhibitor 1 (PAI-1), a major modulator of the fibrinolytic system, is an important factor in cardiovascular disease CVD susceptibility and severity. PAI-1 is highly heritable, but the few genes associated with it explain only a small portion of its variation. Studies of PAI-1 typically employ linear regression to estimate effects of genetic variants on levels, but PAI-1 is not normally distributed, even after transformation. Therefore, alternative statistical methods may provide greater power to identify important genetic variants. Also, the majority of studies aimed at evaluating the role of genetic variation in PAI-1 have been conducted in Asian and Caucasian populations and in these studies results indicate that genetic impact on PAI-1 levels may be population-specific. To our knowledge, our study represents one of the only large scale evaluation of the impact of genetic variants on PAI-1 in an African population. **Methods and Results:** We analyzed >30,000 variants for association with PAI-1 in a Ghanaian population (n=1105), using median regression, a non-parametric alternative to linear regression. Three variants associated with median PAI-1, the most significant of which was in the gene *arylsulfatase B* (*ARSB*) (p=1. 09 x 10-7). We also analyzed the upper quartile of PAI-1, the most clinically relevant part of the distribution, and found 19 additional SNPs significantly associated in this quartile. Of note, a significant association was discovered between elevated PAI-1 levels and a variant in *period circadian clock 3* (*PER3*). **Conclusions:** Our results reveal novel associations with median and elevated PAI-1 in an understudied population. The lack of overlap between the two analyses indicates that the genetic effects on PAI-1 are not uniform across its distribution; this may impact the development of future therapies and clinical interventions. We also provide evidence of a possibly generalizable circadian clock pathway effect on PAI-1, as a recent meta-analysis performed in Caucasian populations identified another circadian clock gene (*ARNTL*).

1364F

Evaluation of alternative imputation strategies for mega-analysis of multiple genotyping cohorts: maximizing imputation yield while minimizing artifacts. *S. C. Nelson[1], D. M. Levine[1], B. Storer[2], L. P. Zhao[2], E. H. Warren[2], P. J. Martin[2], J. A. Hansen[2].* 1) Biostatistics, University of Washington, Seattle, WA; 2) Fred Hutchinson Cancer Research Center, Seattle, WA.

A key consideration in modern GWAS is how to merge datasets genotyped on different arrays. Genome-wide imputation is an attractive solution that provides a set of variants common to each dataset. However, artifacts leading to false-positive associations are a significant concern and no single multi-platform imputation strategy has emerged as a standard. We have extensively assessed three imputation strategies in a hematopoietic stem cell transplantation study comprising three genotyping arrays: Affymetrix 5. 0 (n=2,999), Illumina Omni1M (n=4,114) and Illumina Omni2. 5 (n=2,219). Using the SHAPEIT2/IMPUTE2 software suite and a 1000 Genomes (1000G) Project Phase 3 reference panel, we assessed the following three imputation strategies: (1) imputing the cohorts together to 1000G using only the three-way intersection of array variants; (2) a two-stage approach in which the Omni2. 5 cohort was first used as an internal reference panel to impute the Affymetrix and Omni1M cohorts, followed by imputation to 1000G; and (3) imputing each cohort separately to 1000G. We sought to identify the strategy that maximized imputation yield (i. e. , the number of imputed variants passing a minimum quality threshold) while minimizing imputation artifacts between batches. We evaluated imputation yield by determining the number of imputed variants passing increasingly stringent quality metrics thresholds (IMPUTE2 “info” score cutoffs). We evaluated imputation artifacts between batches by both (1) calculating the correlation in imputed dosages between cross cohort duplicate sample pairs and (2) examining false positive rates from pseudo-association tests where case/control status was defined by genotyping cohort. While strategy 1 yielded the least imputation artifacts, the relatively sparse imputation basis resulted in the lowest imputation yield. Strategy 2 with a stringent filter on internally-imputed variants yielded well-imputed variants while minimizing batch effects; however, the multi-stage strategy introduced some bias from the Affymetrix cohort having a lower stage 1 call rate compared to Omni1M. Notably, strategy 3 produced relatively few artifacts and reasonable imputation yield. We show that while genotyping cohorts can successfully be merged, considerable attention should be paid to selecting imputation basis variants and that the most straightforward method of imputing cohorts separately and combining for mega analysis (strategy 3) may indeed be the most viable option.

1365W

Finding and characterizing latent genetic sub-structure for imprecisely defined phenotypes. *H. Finucane[1,2], N. Fusi[1], A. Price[2], J. Listgarten[1].* 1) Microsoft Research New England; 2) Harvard School of Public Health.

Some case-control phenotypes, labelled only with 0/1, may actually contain distinct underlying sub-phenotypes; for example, depression is thought to be such a phenotype. Analysis of such imprecisely defined phenotypes has negative consequences. First, it obscures genetic signal, decreasing apparent heritability as well as power for association analysis. Second, it obscures understanding of disease, because only a union of causal genes and pathways can be uncovered. In this work, we apply latent variable models to extract sub-phenotypes jointly from both genotypic and phenotypic data in order overcome these problems. In particular, we seek to identify sub-phenotypes that have higher heritability individually than when grouped together as a single phenotype. We compare several newly-developed methods on simulated data, assessing the relative power and computational complexity of the methods as a function of sample size, level of shared genetic architecture among the sub-phenotypes, polygenicity of the trait, and existence of available auxiliary phenotype information.

1366T

Implications of the Co-Dominance Model for Hardy-Weinberg Testingin Genetic Association Studies. *S. Wellek.* Central Institute of Mental Health, University of Heidelberg, Heidelberg, Germany.

The standard way of using tests for compatibility of genetic markers with HWE as a means of quality control in GWAS is to carry out this step of preliminary data analysis with the sample of non-diseased individuals only. We show that this strategy has no rational basis whenever the marker under consideration satisfies the assumption of co-dominance. The mathematical argument behind this statement is made precise and proven as a lemma. The major practical consequence of this theoretical result is that under the co-dominance model, testing for HWE should be done both for cases and controls aiming to establish the combined (intersection) hypothesis of compatibility of both underlying genotype distributions with the HWE assumption. A particularly useful procedure serving this purpose is obtained through applying the confidence-interval inclusion rule derived by Wellek, Goddard & Ziegler (Biom J. 2010; 52:253-270) to both samples separately and combining these two tests by means of the intersection-union principle.

1367F

An approach combining the construction of haplotype collection and the design of population-optimized SNP array on the large-scale whole genome-sequencing project in Japan. *Y. Kawai[1,2], T. Mimori[1], K. Kojima[1,2,4], N. Nariai[1,2], I. Danjoh[1,2], R. Saito[1], M. Yamamoto[1. 2], M. Nagasaki[1,2,3,4].* 1) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 2) Graduate School of Medicine, Tohoku University, Sendai, Miyagi, Japan; 3) Graduate School of Information Sciences, Tohoku University, Sendai, Miyagi, Japan; 4) Department of Cohort Genome Information Analysis, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan.

Genotype imputation is now prerequisite for genome-wide association study (GWAS). Accurate genotype imputation enables to test millions of variants that have been discovered in large-scale whole-genome sequencing project in array-based GWAS. We conducted the whole-genome sequencing of healthy individual in Tohoku region in Japan in the part of prospective genome cohort study and constructed haplotype collection of 1,070 people of cohort participants for the reference panel of whole-genome imputation (referred to as the 1KJPN panel). Furthermore, we developed the customized SNP array "Japonica array" that is optimized for the genotype imputation using the 1KJPN panel. Japonica array contains 659,253 SNPs, including tag SNPs for imputation, SNPs of Y chromosome and mitochondria, and SNPs related to previously reported genome-wide association studies and pharmacogenomics. The imputation performance of Japonica array was compared to the existing GWAS arrays using the 1KJPN panel and the international 1000 genomes project panel. The Japonica array with the 1KJPN panel exhibited better imputation performance over other combinations of SNP array and reference panel. The genomic coverage of imputed genotypes ($r2$ >0. 8) with the Japonica array were 96. 9% and 67. 2% for common (MAF > 5%) and low-frequency SNPs (0. 5% < MAF ≤ 5%), respectively, when the 1KJPN panel is used as reference panel. We demonstrated that the creation of custom-made SNP arrays based upon a population-specific reference panel is a practical way to facilitate further association studies through genome-wide genotype imputations.

1368W

Detecting the source of DNA contamination in genotyping arrays. *G. J. M. Zajac[1,2], L. G. Fritsche[1,2], S. L. Dagenais[3], R. H. Lyons[3], C. M. Brummett[4], G. Abecasis[1,2].* 1) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 3) Department of Biological Chemistry and DNA Sequencing Core, University of Michigan, Ann Arbor, MI; 4) Department of Anesthesiology, Division of Pain Medicine, University of Michigan Medical School, Ann Arbor, MI.

Genotyping arrays measure the relative intensities of allele specific probes, which are then used to call the genotype at each locus. The mixture of DNA samples prior to genotyping, known as contamination, increases the probability that the calling software fails to call a genotype or makes an incorrect call, ultimately reducing the power and accuracy of follow-up genetic analyses, such as GWAS. Often, DNA contamination occurs when two or more samples become physically mixed while preparing the genotyping array, making certain contamination scenarios, like the mixture of two samples in adjacent wells on the array, more likely than others. Existing methods to detect the presence of contaminating DNA use deviations from the expected allele intensities in apparently homozygous reference, heterozygous, or homozygous alternate genotypes towards intensities that would be expected based on population allele frequencies at the locus. Here, we propose an improvedmodel that compares the allele specific intensity patterns between pairs of genotyped individuals to identify the source of DNA contamination. Maximizing the likelihood function over not only the proportion of contaminating DNA but also the contributing individual allows an estimate of the most probable contaminant. This approach has several advantages over existing methods and leads to more accurate estimation of the proportion of contaminating DNA. For example, our new approach should be less sensitive to the misspecification of population allele frequencies, which could lead to inaccurate estimates of contamination levels. Furthermore, if the mixed samples are related, then estimated contamination proportions derived using our method should remain accurate independent of the degree of genetic similarity between the mixed samples, whereas existing methods would underestimate this proportion. Importantly, our detecting the source of contamination provides useful information to guide improvements in sample processing and preparation protocols. Experimental results from an ongoing genotyping study of more than 15,000 individuals illustrate how estimating which samples on an array contributed to contamination helps draft effective laboratory protocols to prevent the mixture of DNA samples.

1369T

Genetic architecture and disease prediction. *E. Ziv, D. Hu.* UCSF, San Francisco, CA.

Genome wide association studies have identified dozens to hundreds of common single nucleotide polymorphisms (SNPs) for many complex traits. However, most of the heritabillity for most complex traits remains unexplained. Still, for many complex traits, combining all of the known SNPs and some high penetrance mutations which are known to cause disease, can provide some information on risk classification. Here we use some analytical approaches and simulations to address the question of what types of study designs would yield the best discoveries in terms of disease predictionWe use the receiver operator characteristic (ROC) area under the curve (AUC) as the measure of power and use parameters such as allele frequencies and relative risks from studies of breast cancer. Analyzing data from known common single nucleotide polymorphisms and mutations in susceptibility genes contribute to current predictive power. Then, we partition the remaining familial risk into common and rare variants. We show that assuming power for discovery is constrained by sample size only (ie full genome sequence is available), then the greatest boost to ROC AUC is from lowest frequency, highest penentrance variant that can be discovered by the given sample size. We then consider the possibility that some loci can only be discovered by combining evidence from many low frequency variants. We consider that not some risk variants may be misclassified at these loci and evaluate the loss of power for disease prediction under these circumstances. Overall, we find that additional discoveries of low frequency, intermediate penetrance risk variants should yield more predictive power than additional discoveries of common low penetrance risk variants.

1370F

Plan a large-scale genetic association study for predictive modelling. *C. Kuo.* Community Medicine and Health Care, Connecticut Institute for Translational Science, Institute for Systems Genomics, University of Connecticut Health Center, Farmington, CT.

With more and more susceptibility loci identified, the findings can be utilized for disease screening and diagnosis. Typically, a number of risk variants are selected to predict the outcome using a simple method such as logistic regression model for case-control studies. It has been known that a model that includes functional and/or (statistically) significant variants only doesn't lead to a high predictive power and the power significantly improves when using a moderate number of top hits from a large-scale genetic association study. Traditionally, a genetic association study is planned such that the true variants with effect sizes in a range will be detected or saying statistically significant. Motivated by the fact that a discovery study planned this way is not optimal for predictive modelling, I propose a method to calculate predictive power given a sample size, a rank or significance threshold, and other prior information. In other words, this method would allow one to choose a sample size and a number of top hits for a pre-specified predictive power. This method has been validated by simulations and implemented in R for practical use.

1371W

The Alzheimer Disease Sequencing Project: Design and Sample Selection. *G. Beecham on behalf of the Alzheimer Disease Sequencing Project.* Dr. John T. Macdonald Foundation Dept. of Human Genetics, Hussman Inst. for Human Genomics, University of Miami, Miami, FL.

The Alzheimer Disease Sequencing Project (ADSP) is an initiative to identify genetic variants influencing late-onset Alzheimer Disease (LOAD) risk and protection, with the goal of defining novel drug targets to treat AD. Given a limited budget and large number of samples available, we considered several study designs to maximize power. We chose a family design for whole genome sequencing (WGS), as linkage helps prioritize variants for follow-up. 1,400 multiplex LOAD families were evaluated. Families were of European (EA) and Caribbean Hispanic (CH) ancestry and lacked known mutations for AD. Families were ranked according to number of cases, generations/branches sampled, and lack of APOE e4 alleles. For whole exome sequencing (WES) we considered several case-control designs to select samples from over 30,000 samples available. These included: an unbalanced design with a greater number of cases (to maximize rare AD variants identified), a balanced design optimized to uncover protective variants (comparing high risk AD cases to high risk controls), and a balanced design focused on reducing heterogeneity and misclassification (by comparing low risk cases --younger, few APOE e4--to controls least likely to develop AD by age 85). Risk and rate of conversion to AD were estimated using age, sex and APOE status, and Braak score at autopsy if available. The utility of other AD risk alleles in evaluation was considered. Designs were compared for their power to detect risk and protective variants, allowing for genetic heterogeneity and misclassification. All cases from 111 Tier 1 families were sequenced (N=578). Array data were obtained on all family members to aid QC and linkage. Balanced designs had higher power than unbalanced and protective designs had lower power to identify risk variants. The third design (reduced misclassification and heterogeneity) was well-powered for risk and protective loci. Accordingly, we selected 5,107 cases (age-at-onset=76+/-9. 1;43% male) and 4,976 controls (age-at-exam=86+/-5. 7,41% male). Additional samples include 512 EA and 173 CH cases from lower tier families. This approach balances power and cost considerations while targeting a variety of important hypotheses (risk, protective, SNV, CNV, coding, non-coding, etc), and reflects the design of an important resource for AD genomics research. Additionally, the approach can be utilized by other late-onset, progressive disorders to better maximize power and sample availability.

1372T

PMAD: Precision Medicine Adaptive Designs for Validation of Genetic Biomarker Information for Targeted Therapeutics. *Q. Wu, M. V. Gwengi, L. B. Gillis.* BioStat Solutions, Inc 5280 Corporate Drive Suite C200 Frederick, MD, 21703.

Background: Precision medicine aims to deliver the 'right' drug to the 'right' patient at the 'right' dose at the 'right' time. An important application of precision medicine is in using genetic biomarker information to identify a subpopulation that may show enhanced drug performance. Once a biomarker model has been developed based on (retrospective) clinical trial data, it must be independently validated before the biomarker can be used as a companion diagnostic. Adaptive designs may be advantageous for this validation. Methods: The direct approach to validate the biomarker is to run a clinical trial to evaluate the identified biomarker model and its threshold. Since the genetic signature was derived from limited clinical data, it may not be truly replicable. However, refinements to the biomarker may yield more precise identification of the 'right' patient for the drug with additional study data. With an adaptive design, an interim look could provide information about whether or not the biomarker is replicable as well as how to improve the biomarker profile. An alternative approach is to employ Bayesian methods to update the biomarker model and/or threshold. Both the biomarker model and the threshold are available as prior information. In the adaptive design, the data available at the interim analysis will be used to create a posterior estimate of the genetic signature and corresponding threshold. This new biomarker profile will be independently validated in the second stage of the study. Results: Simulation studies are used to compare the performance of the two proposed validation designs under different scenarios. This newly proposed approach offers scientific innovation, potential cost savings, and increased likelihood of successfully targeting the 'right' patients based on their genetic signature.

1373F

Methodology for the analysis of multi-ethnic genome-wide association studies. *J. P. Cook, A. P. Morris.* Department of Biostatistics, Farr Institute, Liverpool, Liverpool, United Kingdom.

Traditional genome-wide association studies (GWAS) have primarily been conducted using collections of individuals from homogeneous population groups because: (i) geographical confounding between the trait and genetic variation can inflate type I error rates, if not accounted for; and (ii) there may be reduced power due to heterogeneity in allelic effects on the trait between ethnicities. Confounding is typically accounted for by including principal components (PCs) as covariates in a regression framework, and has been demonstrated to control type I error rates within ancestry groups. However, by including genetic data from diverse populations, the first two PCs have also been demonstrated to generate axes of genetic variation that are anchored by individuals of African, European and East Asian ancestry. In this study, we have undertaken simulations of binary outcomes and GWAS data from diverse ancestries to evaluate: (i) type I error and power to detect association in multi-ethnic analyses with adjustment for PCs; and (ii) power to detect heterogeneity in allelic effects between populations by including an interaction between a SNP and the first two PCs. Simulation results demonstrate that type I error rates are controlled by including 10 PCs as covariates in the regression analysis, even in the presence of extreme population structure, and that power to detect heterogeneity between populations is strong, even at modest effect sizes. We applied these methods to a multi-ethnic GWAS of type 2 diabetes (T2D) in the Resource for Genetic Epidemiology Research on Adult Health and Aging (GERA). After quality control, 71,604 unrelated participants, including 9,747 T2D cases, were retained for analysis and imputed up to the multi-ethnic reference panel (Phase 3, October 2014) from the 1000 Genomes Project. Lead SNPs at ten loci attained genome-wide significant evidence ($p<5x10-8$) of association with T2D, including a novel signal mapping to *TOMM40* ($p=2.8x10-9$), a gene previously implicated in Wolfram Syndrome, a neurodegenerative disorder characterised by diabetes. In conclusion, our study demonstrates that incorporation of PCs in multi-ethnic GWAS can fully account for population structure, and can be used to detect heterogeneity in allelic effects between ancestry groups. Furthermore, we provide further insight into the genetic architecture of T2D, and highlight the benefits of multi-ethnic analysis for the discovery of novel loci associated with complex diseases.

1374W

The LASER server: a web-based tool for tracing individual ancestry in a reference principal component space. *D. Taliun[1], S. Schönherr[2], L. Forer[2], M. Boehnke[1], G. Abecasis[1], C. Wang[3].* 1) Department of Biostatistics, University of Michigan, Ann Arbor, MI; 2) Division of Genetic Epidemiology, Department of Medical Genetics, Molecular and Clinical Pharmacology, Innsbruck Medical University, Innsbruck, Austria; 3) Department of Computational and Systems Biology, Genome Institute of Singapore, Singapore.

Recent genetic association studies focus on detecting association with rare variants, an analysis that typically requires large samples to achieve reasonable statistical power. Augmenting case samples with ancestry matched controls from other studies is a cost effective and attractive way to improve statistical power. A first step in identifying these ancestry matched controls is to define the ancestry of samples from different studies in the same coordinate space. Wang et al. (2014, 2015) introduced the LASER framework that unifies ancestry estimation for sequenced and genotyped samples in a predefined principal component space, providing a way to facilitate control of population stratification in joint analysis of multiple genetic datasets. Given a set of reference samples (for example, the 1000 Genomes Project samples), LASER can place sequenced or genotyped samples in the same ancestry coordinate space – even when samples are genotyped with different arrays, sequenced at only a few genes or the whole exome, or genome sequenced at different depths. To facilitate and simplify ancestry estimation using a shared coordinate space, we developed the LASER server, a web-based ancestry estimation tool with build-in ancestry reference panels for different geographic regions. The LASER server has a user-friendly web interface for job submission and monitoring and generates interactive web-based visualization of estimated ancestries. We parallelized the computations by adapting LASER's implementation to the MapReduce paradigm so that our server can deliver results in a timely fashion even for large datasets. As a result, the LASER server allows samples from different sources to be mapped to a shared ancestry coordinate space. Using these coordinates, samples can be precisely ancestry matched across studies to help identify opportunities to increase the study sample without introducing population stratification. The server can be accessed at http://laser. sph. umich. edu.

1375T

Unified tests for fine scale mapping and identifying sparse high-dimensional sequence associations. *S. Cao[1,2], H. Qin[2,3], A. Gossmann[2,4], H. Deng[2,3], Y. Wang[1,2,3].* 1) Department of Biomedical Engineering, Tulane University, New Orleans, LA; 2) Center for Bioinformatics and Genomics, Tulane University, New Orleans, LA; 3) Department of Biostatistics and Bioinformatics, Tulane University, New Orleans, LA; 4) Department of Mathematics, Tulane University, New Orleans, LA.

Motivation: In searching for genetic variants for complex diseases with deep sequencing data, genomic marker sets of high-dimensional genotypic data and sparse functional variants are quite common. Existing sequence association tests are incapable to identify such marker sets and individual causal loci, although they appeared powerful to identify small marker sets with dense functional variants. In sequence association studies of admixed individuals, cryptic relatedness and population structure are known to confound the association analyses. **Method:** We here propose a unified test (uFineMap) to accurately localize causal loci and a unified test (uHDSet) for identifying high-dimensional sparse associations in deep sequencing genomic data of multi-ethnic individuals. These novel tests are based on scaled sparse linear mixed regressions with Lp ($0<p<1$) norm regularization. They jointly adjust for cryptic relatedness, population structure and other confounders to prevent false discoveries and improve statistical power for identifying promising individual markers and marker sets that harbor functional genetic variants of a complex trait. **Results:** Under a wide range of simulated scenarios, the proposed tests appropriately controlled Type I error rate and appeared more powerful than several existing prominent methods. We illustrated their practical utilities by the applications to DNA sequence data of Framingham Heart Study for osteoporosis. The proposed tests identified 11 novel significant genes that were missed by the prominent famSKAT and GEMMA. In partiicular, four out of six most significant pathways identified by the uHDSet but failed by famSKAT have been reported to be related to BMD or osteoporosis in the literature.

1376F

Evidence for Further Colorectal Cancer Susceptibility Genes in Addition to the Mismatch Repair Genes and *MUTYH*. *M. Jenkins[1], A. K. Win[1], J. G. Dowty[1], A. C. Antoniou[2], A. Lee[2], G. G. Giles[1,3], D. D. Buchanan[1,4], D. J. Ahnen[5], S. N. Thibodeau[6], G. Casey[7], S. Gallinger[8], L. Le Marchand[9], R. W. Haile[10], J. D. Potter[11,12,13], Y. Zheng[11,12], N. M. Lindor[14], P. A. Newcomb[11,12], J. L. Hopper[1], R. J. MacInnis[1,3].* 1) Centre for Epidemiology and Biostatistics, The University of Melbourne, Carlton, Australia; 2) Centre for Cancer Genetic Epidemiology, Department of Public and Primary Care, University of Cambridge; 3) Cancer Epidemiology Centre, Cancer Council Victoria, Melbourne, Victoria, Australia; 4) Oncogenomics Group, Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Parkville, Victoria, Australia; 5) University of Colorado School of Medicine, Denver, Colorado, USA; 6) Molecular Genetics Laboratory, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA; 7) Department of Preventive Medicine, Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, California, USA; 8) Lunenfeld Tanenbaum Research Institute, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada; 9) University of Hawaii Cancer Center, Honolulu, Hawaii, USA; 10) Department of Medicine, Division of Oncology, Stanford Cancer Institute, Stanford University, California, USA; 11) School of Public Health, University of Washington, Seattle, Washington, USA; 12) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA; 13) Centre for Public Health Research, Massey University, Wellington, New Zealand; 14) Department of Health Science Research, Mayo Clinic Arizona, Scottsdale, Arizona, USA.

Background: Familial aggregation of colorectal cancer is probably due to multiple susceptibility genes, perhaps acting in conjunction with shared lifestyle risk factors. The aim of this paper was to investigate the genetic models that can best explain familial colorectal cancer not due to the DNA mismatch repair (MMR) genes or *MUTYH,* and to accurately estimate the mutation carrier frequency in the population*.* **Methods**: We studied 5,744 colorectal cancer cases (probands) who were recruited from population cancer registries from the US, Canada and Australia between 1997 and 2012. Blood samples taken from the probands were analysed for mutations in MMR genes and *MUTYH*. We investigated major gene models (dominant, recessive, general), polygenic models, and mixed models (both major gene and polygenic) by analysing information on cancer history in first-degree relatives and on the mutation status of the probands using the pedigree analysis software MENDEL. We estimated the simultaneous effects of MMR genes (*MLH1, MSH2, MSH6, PMS2*), *MUTYH*, a sixth hypothetical gene GENE6, and a polygenic effect. The models were assessed by likelihood comparisons and by comparison of the observed numbers of mutations and affected relatives with the predicted numbers. **Results**: The best-fitting model was mixed dominant model with an age-dependent polygenic standard deviation on top of the five known genes. Under this model, the estimated mutation frequency for GENE6 was 0. 10% (95% confidence interval [CI], 0. 02% – 0. 54%) and the relative risk for GENE6 mutation carriers was 31. 1 (95% CI, 11. 6 – 83. 4). The standard deviation of the polygenic component was 1. 81 (95% CI, 1. 05 – 3. 12) for age <40 years, 0. 96 (95% CI, 0. 51 – 1. 82) for age 40-49 years, 0. 68 (95% CI, 0. 34 – 1. 34) for age 50-59 years, 0. 89 (95% CI, 0. 52 – 1. 51) for age 60-69 years, and 0. 72 (95% CI, 0. 40 – 1. 28) for age ≥70 years. The estimated population allele frequency (%) was 0. 026 (95% CI, 0. 020 – 0. 034) for *MLH1*, 0. 018 (95% CI, 0. 013 – 0. 024) for *MSH2*, 0. 066 (95% CI, 0. 044 – 0. 098) for *MSH6*, 0. 070 (95% CI, 0. 047 – 0. 104) for *PMS2*, and 1. 11 (95% CI, 0. 95 – 1. 30) for *MUTYH*. **Conclusions**: These findings suggest that it is more likely that the remaining familial aggregation of colorectal cancer is due to a major gene on top of many low-risk variants than due to many low-risk variants alone. We have also provided accurate estimates of the carrier frequency of a MMR gene and *MUTYH* in the population.

1377W

Methods for detecting modifier genes responsible for phenotypic heterogeneity. *J. C. Carlson[1], E. J. Leslie[2], J. R. Shaffer[3], H. Schwender[4], M. A. Taub[5], I. Ruczinski[5], J. C. Murray[6], T. H. Beaty[7], M. L. Marazita[2], E. Feingold[3,1].* 1) Biostatistics, University of Pittsburgh, Pittsburgh, PA; 2) Oral Biology, University of Pittsburgh, Pittsburgh, PA; 3) Human Genetics, University of Pittsburgh, Pittsburgh, PA; 4) University of Dusseldorf; 5) Biostatistics, Johns Hopkins University, Baltimore MD; 6) Bill and Melinda Gates Foundation; 7) Epidemiology, Johns Hopkins University, Baltimore MD.

Phenotypic heterogeneity is a hallmark of common, complex diseases. Combining similar phenotypes for analysis is a common approach when it is hypothesized that they share a common genetic etiology. Alternatively, each of the phenotypes may have unique genetic susceptibilities and underlying pathophysiology. Or, phenotypes may share some genetic risk factors while others are modifier genes that define a particular subphenotype or a more extreme phenotype. We evaluated several different approaches for finding genes under these models, considering both population-based and family-based data. In particular, we compared a case-case analysis that compares subphenotypic groups to each other directly with more conventional heterogeneity tests. We applied these methods to a cohort of 1,409 case-parent trios with nonsyndromic orofacial clefts (OFCs) that were sequenced for 6. 3Mb surrounding GWAS regions. OFCs are known for their considerable phenotypic heterogeneity as there are prevalence biases relating to cleft type (i. e. , cleft lip, cleft lip and palate, cleft palate), gender, the side of the lip that is affected, and severity among affected individuals.

1378T

An effective approach to identify rare variants associated with common diseases utilizing large families. *J. Sul[1,2], S. Sunyaev[1,2].* 1) Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Harvard University, Boston, MA 02115, USA; 2) Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.

Majority of sequencing studies aiming at finding association of rare variants with complex traits use the case-control design. A number of new initiatives plan to sequence large numbers of well-phenotyped pedigrees. Family-based studies offer several advantages over case-control studies. First, genetic variants that are rare in the population may be enriched in certain large families, which allows family-based studies to achieve higher power in detecting associations of rare variants than case-control studies, as suggested in recent literature. Second, sequencing errors can be detected as Mendelian violations. This reduces erroneous calls made by NGS, and hence increases power of rare variant analysis by correctly calling rare variants. We present a new approach to detect rare variants associated with common diseases in large family collections. This method detects rare variants with the skewed segregation patterns with respect to the affected/unaffected status. To increase the statistical power, the proposed test combines multiple rare variants both by gene and across multiple families. The method can be applied to any set of pedigrees phenotyped for either a binary or a quantitative trait. It is equally applicable to affected-only pedigrees. This is in contrast with most of existing approaches that focus on small families and handle either binary or quantitative traits. Our method also handles pedigrees with missing founders. The proposed method achieves higher power than existing approaches, as shown in simulations. The power gain is especially profound for large families. We also demonstrate that many existing methods have inflated false-positive rates if founders of large families are not genotyped in the study. The new method handles missing founder genotypes without inflating false-positive rate.

1379F

A family-based rare variant association test for time-to-event data based on the frailty model. *W. Qi[1,2], W. Wang[3], A. Allen[1], Y. J. Li[1,2].* 1) Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, NC; 2) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC; 3) Institute of Statistics, National Chiao Tung University, Taiwan.

Since the rise of next generation sequencing technologies, many gene-based association methods have been developed for testing rare variants for disease susceptibility or modifiers. There is a robust literature on gene-based rare-variant association methods for binary or continuous phenotypes that can be applied to either unrelated case-control or familial data. On the other hand, the development of gene-based association tests for time-to-event data drew less attention, particularly for family data. Examples of the time-to-event data in genetics research include adverse events in pharmacogenetics research and age at onset (AAO) of the disease of interest in disease gene mapping research. For instance, we have previously shown that AAO of Alzheimer disease is genetically controlled (Li et al. 2002). Here, we propose a family-based rare variant association test for time-to-event data. We adopted the framework of orthogonal model used in the QTDT method (Abecasis et al. 2000) to decompose the genotype score into between- and within-family genotype scores for each rare variant. It can be applied to families of any size, with or without parental information. To model the time-to-event data in family, we included vectors of between- and within-family genotype scores for all rare variants of interest, as well as other covariates of interest in the frailty model. We assume that the between- and within-family effects are random effects following normal distributions. To differentiate the effect sizes between variants, we incorporated weights on rare variants in the partial likelihood function. Multiple weighting schemes for rare variants were evaluated. Both likelihood ratio and variance-component score tests were derived to test the within-family effects of rare variants, which provide a direct estimate of association of rare variants. The property and performance of both tests were then evaluated by extensive simulation studies using SeqSIMLA program (Chung and Shih 2014) to generate sequence data in families.

1380W

An Approach to High resolution IBD mapping under a linkage peak. *J. E. Hicks[1,2], P. An[2], H. J. Abel[2], M. A. Province[2].* 1) Division of Biostatistics, Washington University School of Medicine in Saint Louis, Clayton, MO; 2) Division of Statistical Genomics, Washington University School of Medicine in Saint Louis, Clayton, MO.

With the availability of high-throughput sequencing, rare variants as one of the causes of common phenotypes is a popular theory in genetic epidemiology. These rare-variants are more likely to be found on long haplotypes within the population, which are identifiable with common tools such as GERMLINE and BEAGLE. Conventional linkage analysis considers identify-by-descent (IBD) states within observed pedigrees to identify relationships between genotype and phenotype. By incorporating IBD states between pedigrees, further information can be useful for localizing these rare variants, since some putatively "independent" pedigrees may actually share recent common ancestry for the causative rare alleles. In order to better leverage ancestral information in combination with linkage information, we defined a statistic to quantify IBD in these pedigrees as the number of alleles shared IBD among individuals, summed over all pedigrees. We applied this approach to the Long Life Family Study, a family-based cohort used to identify the genetic determinants of aging traits, consisting of 615 families spanning 4,694 genotyped participants. These individuals were genotyped on the Illumina Human Omni 2. 5 v1 chip. In a genome-wide variance components linkage screen for hemoglobin A1c (HbA1c) levels, a region on chromosome 2 showed significant evidence of linkage in the region 36,919,293-39,982,534 bp. Using 309 individuals selected for membership in pedigrees displaying strong linkage to HbA1c, higher resolution analysis was used to narrow down the linkage region. IBD segments between these individuals were identified as genome spans of over 500 kilobase where at least 1000 genotypes were identical-by-state and there was a marker density of at least 100 markers per megabase. Of 2,378 within-pedigree pairs of individuals, the maximal amount of sharing using our approach was 1,258 allele pairs at rs7576549, a SNP within the region identified by standard linkage analysis. This demonstrates the utility of high-resolution IBD segment mapping in refining signals from linkage analyses.

1381T

Linkage Analyses Reveals Significant Association for Myopia. *A. Musolf[1], C. L. Simpson[1], F. Murgia[1,2], L. Portas[1,2], J. E. Bailey-Wilson[1], D. Stambolian[3].* 1) Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore, MD; 2) Institute of Population Genetics, CNR, Traversa La Crucca, 3 - 07040 Reg. Baldinca, Li Punti, Sassari, Italy; 3) Department of Ophthalmology, University of Pennsylvania, Philadelphia, PA 19104, USA.

Myopia (nearsightedness) is a condition where overgrowth of the eye causes light to focus in front of the retina and instead of directly on it, leading to blurring of distant images. It is one of the most common causes of reduced vision worldwide, affecting 1 in 4 Americans and has reached epidemic proportions in some parts of southeast Asia. Genotypes from the Illumina Exome array were merged with prior microsatellite data from extended pedigrees with multiple individuals affected with myopia. These families come from three discrete populations: African-Americans, Ashkenazi Jews from New Jersey, and Pennsylvania Amish. Two-point and multi-point linkage analyses were performed on each family using both a general inheritance model and an inheritance model specific to each population using MCLink. Familywise two-point LOD scores of greater than 2 were observed at chromosomes 3, 4, 6, 10, 13 and 14 in individual Ashkenazi Jewish families. The two highest LOD scores were equal to 2. 9 and appeared in a nonsynonymous SNP in an exon of the SYNE1 gene on chromosome 6 in one family and an intronic microsatellite in the antisense RNA gene ZMIZ1-AS1 on chromosome 10 in another family. We also observed familywise two-point LOD scores greater than 2 in the Pennsylvania Amish on chromosomes 2, 3, 9, 15, and 16. The two highest scores were 2. 3 at an intronic microsatellite located in the MTMR10 gene on chromosome 15 in one family and 2. 2 at a microsatellite in an intergenic region on chromosome 10 in another family. Although no two-point LOD scores higher than 2 were observed in the African-Americans, these families were smaller and therefore less informative. Nonetheless, familywise two-point LOD scores of greater than 1 were seen on chromosomes 1, 2, 5, 8, and 20. The highest two LOD scores were 1. 52 at a microsatellite in a noncoding RNA gene on chromosome 4 and 1. 47 at an exonic nonsynonymous SNP in GALNT15 on chromosome 3. Multipoint linkage analyses are ongoing and will be presented, including HLOD scores calculated for each population group and across all populations. Overall, this work identifies multiple interesting linkage signals for myopia across three discrete populations, using both two-point and multi-point analyses. Many of these signals are located within genes, which may be promising candidates for further study.

1382F

Power of Single Extended Pedigrees to Classify Rare Variants of Uncertain Significance in BRCA1 and BRCA2. *B. H. Shirts[1], E. A. Rosenthal[2].* 1) Laboratory Medicine, University of Washington, Seattle, WA; 2) Genetics, Medicine, University of Washington, Seattle, WA.

Family based analysis of trait and locus cosegregation can be used to facilitate classification of rare variants of uncertain significance (VUS). Many VUS will only be present in a single extended pedigree. Genealogy and social networking resources have made identifying and collecting distant relatives more feasible, potentially increasing the power to classify a VUS as benign or pathogenic. Using simulated family data we sought to identify the family sizes and structures necessary for adequate power. We simulated pedigrees based on historical demographic data from the United States (USA) and China (CHI), which differ in family size and structure. Pedigrees were simulated to include probands similar in age and sex to those currently tested in genetics clinics. Each simulated pedigree contained all descendants of the proband's carrier great-grandparent. Phenotypes for each individual were sampled based on age, sex and carrier status. We simulated >1000 families for all eight scenarios defined by three categories: pathogenic or benign variant; BRCA1 or BRCA2; USA or CHI. We calculated proband adjusted likelihood ratios (LRadj) using linkage analysis, assuming gene specific age-based penetrance and an allele frequency of 0. 001. A VUS was considered likely pathogenic when LRadj > 20, and likely benign when LRadj < 0. 05, per International Association for Research on Cancer (IARC) recommendations. The power and type I error rates for correctly classifying a pathogenic variant were: USA BRCA1: 26%, 3%; CHI BRCA1: 41%, 4%; USA BRCA2: 22%, 2%; CHI BRCA2: 34%, 3%. The power and type I error rates for correctly classifying a benign variant were: USA BRCA1: 70%, 0%; CHI BRCA1: 79%, 0%; USA BRCA2: 60%, 0%; CHI BRCA2: 69%, 0%. The number of breast/ovarian cases in families with power to classify pathogenic BRCA1/2 variants in US families ranged from 3 to 23 (median 7). As expected, power was higher in larger families. Despite lower recent fertility, simulated families with CHI structure had higher power to correctly classify both pathogenic and benign variants in both genes, most likely due to rapid family growth in the early 20th century and the presence of a larger number of older relatives. Given the relative prevalence of pathogenic and benign variants, we show that VUS classification in BRCA1/2 will likely be possible in >50% of families with either CHI or USA ancestry when all descendants of a common carrier great-grandparent are included in the analysis.

1383W

Large-scale phenome-wide scan in twins using electronic health records. *S. Hebbring[1], J. Pathak[2], J. Mayer[3], Z. Ye[3], S. Schrodi[1].* 1) Center for Human Genetics, Marshfield Clinic, Marshfield, WI; 2) Health Sciences research, Mayo Clinic, Rochester, MN; 3) Bioinformatics Research Center, Marshfield Clinic, Marshfield, WI.

Challenges in modern population-based genomic research have resulted in a re-awakening of family-based studies. However, difficulties arise with identifying the most interesting diseases and families for family-based research. Use of large patient populations linked to an electronic health record (EHR) may alleviate such challenges. Using readily available EHR data, we identified two independent twin cohorts from Marshfield (MCTC) and Mayo Clinic (Mayo-TC) (total=28,888 twins). In both cohorts, we measured familial aggregation across the phenome defined by 5,598 phenotypes/ICD9 codes. Not surprisingly, the top associations in both cohorts included the phenotype defining twin births (V31; MCTC, Pχ1. 5E-587; Mayo-TC, Pχ2. 7E-359). Additional top associations were comprised of numerous perinatal phenotypes including those related to birth weight (e. g. , ICD9 765: MCTC, Pχ2. 2E-501; Mayo-TC, Pχ1. 5E-161) and jaundice (e. g. , ICD9 774: MCTC, Pχ1. 3E-338; Mayo-TC, Pχ5. 6E-133). It is not unexpected that infants born together will likely have the same gestation period, comparable birth weights, and similar infant-related co-morbidities. There were also numerous common phenotypes previously characterized with high heritability such as myopia (ICD9 367. 1: MCTC; Pχ9. 5E-101; Mayo-TC, Pχ6. 0E-61) and developmental delays (ICD9 315: MCTC, P=1. 8E-134; Mayo-TC, Pχ2. 9E-57). Likewise, there were numerous rare Mendelian diseases (e. g. , hereditary spherocytosis, sickle cell anemia, and thalassemia) that were associated in both cohorts. Based on a combined analysis, 1,406 of the 5,598 phenotypes were statistically enriched for concordance (P<8. 9E-6), many with unknown genetic etiologies. With our novel phenome-wide methodologies highly translatable to other EHR systems, and potentially to other familial relationships, this study may pave the way to biotechnologically smart EHR systems that integrate family data to predict, prevent, and treat many diseases for the advancement of “precision medicine. ” Lastly, this study provides an intriguing perspective for the future of genomic research. Specifically, the future when large patient populations with sequenced genomes are unified by familial relationships in an integrated EHR system with extensive phenotypic data.

1384T

Detection of Gene-Gene Interaction and Genomic Imprinting in Affected Sib Pairs. *C. C. Wu[1], S. Shete[2].* 1) Environmental/Occupational Health, National Cheng Kung University, Tainan, Taiwan; 2) Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA.

Genomic imprinting is an epigenetic phenomenon, which results in 2 alleles at a locus being expressed differentially depending on the parental origin of the chromosome. Recent studies show that imprinted genes not only play important roles in development and placental biology before birth but also have major effects on postnatal stages on common diseases, including obesity, psychiatric disorders, and cancer. Genome-wide association studies have discovered several hundred genetic variants associated with common diseases; the efforts to map interacting genes have been less successful. In response, we proposed a method to detect gene-gene interaction effects in the presence of genomic imprinting. We extended our previous allele-sharing method and presented 3 mathematical two-locus models incorporating imprinting effects: additive, multiplicative, and general models. Our methods are model-free based on allelic identity-by-descent sharing by affected sib pairs. We used two-locus score method to assess the gene-gene interaction effects using affected sib pairs in the presence of imprinting effects. We further proposed to use a novel test to assess the gene-gene interaction and imprinting effects individually and jointly.

1385F

Association and Meta-analysis Methods for X Chromosome. *S. Feng[1,2], G. Pistis[3], C. Sidore[3], A. Mulas[3], M. Zoledziewska[3], F. Busonero[3], S. Sanna[3], F. Cucca[3], D. Schlessinger[4], G. Abecasis[1].* 1) Department of Biostatistics, Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, MI, USA; 2) 23andMe, 899 West Evelyn Avenue, Mountain View, CA 94041; 3) IRGB, Cittadella Universitaria di Monserrato, Monserrato, Cagliari, Italy; 4) Laboratory of Genetics, National Institute of Aging, Baltimore, MD, USA.

Variants and genes on the X chromosome account for 5% of the human genome sequence. Following our previous work on association and meta-analysis methods for autosomal genes, we propose a series of gene-level association tests including burden, variable threshold, and sequence kernel association tests for analysis of rare variants on chromosome X. We use a linear model with three variance components to take account autosomal, X chromosomal, and individual environment contributions to total variance, and allow estimation of empirical kinship matrix for both autosome and X chromosome when pedigree structure is not known or cryptic relatedness and population structure are a concern. By simulation, we show that all of our gene-level tests have type I error well controlled and that, in models that assume complete X-inactivation, samples of males provide more power to detect association with X-linked genes. Furthermore, for any given per allele effect size, power to detect X-linked alleles was larger than for autosomal alleles with the same effect sizes. Using the SardiNIA study, we examined association at X-linked variants on exome chip, showing that variants near *G6PD* are associated with multiple quantitative traits.

1386W

PSEUDOMARKER-WGS: Family-based association tests for whole genome sequence data. *T. Hiekkalinna[1,2], M. Perola[1,2,3], J. D. Terwilliger[1,4,5].* 1) Genomics and Biomarkers Unit, National Institute for Health and Welfare (THL), Helsinki, Finland; 2) Institute for Molecular Medicine Finland FIMM; 3) University of Tartu, Estonian Genome Center, Tartu, Estonia; 4) Department of Psychiatry, Department of Genetics and Development, Gertrude H. Sergievsky Center, Columbia University, New York, NY, USA; 5) Division of Medical Genetics, New York State Psychiatric Institute, New York, NY, USA.

In the past decade, the human genetics community has spent a huge amount of resources on attempting to identify genetic variants underlying common disorders in genome-wide association studies (GWAs) on large population based cohorts. The common-disease common-variant hypothesis was the motivation for this quest, suggesting that if we genotyped a sufficiently large number of SNP markers across the genome we would be able to identify allelic associations with any common variant influencing a given disease. However, the results to date have been in a way disappointing, with lots of SNPs showing correlations with such disease outcomes, but with miniscule effect sizes, even when combined. Researchers are now returning to family-based designs, noting that under every model considered, they provided higher power for gene mapping, with the reason for the shift to studying unrelated individuals having been motivated almost exclusively by cost and convenience of acquiring large samples. However, as the world moves closer to routine sequencing as part of normal health care to identify any of the hundred thousand or so variants known to outright cause Mendelian disorders, we will soon have data available from relatives sharing some disease once again, motivating a resurgence of interest in classical position cloning strategies based on family data, but adapted to the situation where we have full sequence data, rather than a sparse map of markers as inputs. Family data will be much more powerful, for obvious reasons, as it has been well known fact for several decades that it affected relatives are much more likely to share the same disease predisposing genetic variants than any pair of unrelated affected individuals from the population. In this post-GWAs era, whole genome sequencing (WGS) techniques will be used to generating a vast amount of data, including relatives, which will raise enormous computational challenges for human genetics researchers. We have developed a set of user friendly programs, PSEUDOMARKER-WGS, for analysis of massive amounts of sequence based data in family-based joint linkage and association mapping. Here, we will show the superior power of our PSEUDOMARKER-WGS approach to joint linkage and/or linkage disequilibrium analysis as compared with recently proposed and classical methods. Our PSEUDOMARKER method can analyze jointly extended pedigrees, sib ships, trios and unrelated individuals and it is therefore an optimal algorithm.

1387T

Statistical testing for rare variant associations using affected family members. *K. Lin[1], S. Zöllner[1,2].* 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Psychiatry, University of Michigan, Ann Arbor, MI.

Modern sequencing technology allows investigating the impact of rare variant on complex disease. However, in a conventional case-control design, large samples are needed to achieve sufficient power for testing associations of suspected rare variants to complex diseases. In such large samples, population structure can cause spurious signals. One approach to overcome low power and population stratification is family-based study design since the count of risk variants can be enriched in a family containing multiple affected members with the same genetic background. However, existing family-based methods do not efficiently utilize the fact that rare risk variants are more likely to reside on identical-by-descent chromosomes shared among affected family members. Given many family studies have been generated and new studies are ongoing, there is need to develop powerful approaches to discover associations between rare variants and disease phenotypes in families. We propose a novel test for a sample of families. In each family, we first determine the chromosome transmission at the locus of interest for all family members. Since affected family members are more likely to share risk variants identical-by-descent, we propose to test if rare variants are shared more than expected given the known haplotype transmission and the founder genotypes. In addition, we develop imputation algorithm for families that founders are missing, so that this test can adopt a sample of families with arbitrary family structures. The proposed test is generally robust to population stratification since each family is a matched unit. Using simulations, we compare our method with existing methods for family-based data as well as the conventional case-control design. Considering different models from rare to common variants, we show that the proposed approach is especially more powerful than the conventional case-control design for rare variants; in the presence of allelic heterogeneity, the proposed test further increases the power. In sum, we propose a family-based test that efficiently uses sharing of rare risk variants to maximize the power for detecting rare variant associations.

1388F

A Quasi-likelihood Approach To Detect Transmission Bias Using Sibship Data. *H. S. Kulkarni, S. Ghosh.* Human Genetics Unit, Indian Statistical Institute, Kolkata, West Bengal, India.

The Transmission Disequilibrium Test (TDT) is a popular family based alternative to the population based case-control design as it protects against population stratification. The study of quantitative traits has gained prominence since most complex disorders are governed by quantitative precursors which contain more information on within genotype variability. There have been several extensions of the classical TDT procedure (Speilman et al. 1993) for quantitative traits based on logistic regression (Waldman 1999) and center adjusted logistic regression model (Haldar and Ghosh 2015). However, the TDT procedure is not a valid test for association in the presence of multiple sibs in a family as the marginal effect of linkage can result in transmission bias of an allele. We propose a modification of the classical TDT procedure based on the quasi likelihood approach (Wedderbur 1974) which is also equivalent to the generalized linear regression model (McCullagh and Nelder 1989) using a logit link function. The major advantage of the proposed method over the classical TDT is that it includes the covariance of the transmission status of alleles within a sibship induced by linkage. The regression parameter is estimated using generalized estimating equation (Gourieroux et al 1984; Zeger 1986). We perform extensive simulations under a wide spectrum of the genetic models and different distributions of the quantitative trait (normal, chi-square and Poisson). We compare the empirically power of our proposed test with the FBAT test for sibship data (Lake et al 2000). We observe that the proposed test yields comparable powers across different probability distributions. We also find that the quasi likelihood approach is more powerful than the FBAT test for sibship data.

1389W

Increased power for detection of parent-of-origin effects via the use of haplotype estimation. *R. A. J. Howey[1], C. Mamasoula[1,2], A. Töpf[1], R. Nudel[3], D. F. Newbury[3], S. E. Fisher[4,5], J. A. Goodship[1], B. D. Keavney[1,6], H. J. Cordell[1].* 1) Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, NE1 3BZ, United Kingdom; 2) Institute of Health and Society, Newcastle University, Newcastle upon Tyne, NE2 4AX, United Kingdom; 3) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, United Kingdom; 4) Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands; 5) Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands; 6) Institute of Cardiovascular Sciences, University of Manchester, Manchester, M13 9NT, United Kingdom.

Parent-of-origin (or imprinting) effects relate to the situation where traits are influenced by the allele inherited from only one parent, with the allele from the other parent having little or no effect. Given SNP genotype data from case/parent trios, the parent-of-origin of each allele in the offspring can often be deduced unambiguously; however this is not true when all three individuals are heterozygous. Most existing methods for investigating parent-of-origin effects operate on a SNP by SNP basis and either perform some sort of "averaging" over the possible parental transmissions or else discard ambiguous trios. If the correct parent-of-origin at a SNP could be determined, this would provide extra information and increase the power to detect effects of imprinting. We propose making use of the surrounding SNP information, via haplotype estimation, to improve estimation of parent-of-origin at a test SNP for case/parent trios, case/mother duos and case/father duos. This extra information is then used in a multinomial modelling approach to estimate parent-of-origin effects at the test SNP. We show through computer simulations that our approach has increased power over previous approaches, particularly when the data consist only of duos. We apply our method to two real data sets and find a decrease in significance of p-values in genomic regions previously thought to possibly harbour imprinting effects, thus weakening the evidence that such effects actually exist in these regions, although some regions remained more significant than expected. Softwareis available at www. staff. ncl. ac. uk/richard. howey/emim.

1390T

The *LRRK2* gene is a shared genetic regulator among three clinically divergent common inflammatory disorders. *V. M. Fava[1,2], J. Manry[1,2], A. Cobat[1,2], M. Orlova[1,2], N. Van Thuc[3], N. Ngoc Ba[3], V. H. Thai[3], L. Abel[4,5,6], A. Alcaïs[4,5,6,7], E. Schurr[1,2], The Canadian Lrrk2 in Inflammation Team (CLINT) investigators.* 1) Program in Infectious Diseases and Immunity in Global Health, Research Institute of the McGill University Health Centre, Montreal, Canada; 2) The McGill International TB Centre, Human Genetics Department, McGill University, Montreal, Canada; 3) Hospital for Dermato-Venerology, Ho Chi Minh City, Vietnam; 4) Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale, Paris, France; 5) University Paris Descartes, Imagine Institute, Paris, France; 6) St Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, USA; 7) Centre d'Investigation Clinique, Unité de Recherche Clinique, Necker and Cochin Hospitals, Paris, France.

Coding variants of *LRRK2* have been associated with Parkinson's Disease (PD) and Crohn's Disease (CD). Moreover, non-coding variants of *LRRK2* were associated with leprosy. Recently, we demonstrated that a previously reported association between leprosy and the *TNFSF15*/*TNFSF8* genes (also a CD susceptibility locus) was most likely due to patients suffering from type-1 reactions (T1R). This prompted us to investigate if *LRRK2* variants were primarily associated with T1R rather than leprosy. We tested the preferential *LRRK2* association with T1R over leprosy in a family-based design. For this, we matched two samples of 229 families each with divergent T1R affection status in their offspring. In set 1, patients suffered from leprosy and T1R while in set 2 patients only suffered from leprosy. By formally comparing the T1R-affected versus the T1R-free subset we were able to address the association of genetic markers with the T1R endophenotype. We genotyped 156 SNPs in a 500kb window overlapping the *LRRK2* and *MUC19* genes. Of the 156 variants, 36 were nominally significantly associated in the T1R-affected subsetand 20 of these showed significant evidence for heterogeneity between T1R and leprosy. The "M" allele of the M2397T coding variant was significantly associated with T1R only (P = 0. 008, $P\ het$ = 0. 01). Thesame allele is a risk factor for CD (P = 2. 6 e-15). The M2397T impacts LRRK2 turn over with the methionine variant showing a lower half-life in compare to the threonine. A second SNP, rs7970326, significantly associated with T1R (P = 0. 03, $P\ het$ = 0. 006) is a risk factor for PD (P = 1. 1 e-6). Next, we quantified *LRRK2* mRNA expression in a whole blood assay of 54 leprosy cases in presence and absence of stimulation with *Mycobacterium leprae* antigen. Among the SNPs associated with T1R, we identified a series of eQTLs in unstimulated but not in stimulated blood. In that instance, the eQTL rs10784470 G-allele is a risk factor for T1R (P = 0. 02, $P\ het$ = 0. 02) and Inflammatory bowel disease (P = 7. 9 e-6). The *LRRK2* expression for this SNP bin is altered in response to mycobacterium antigen. Our results link the *LRRK2* gene to the pathogenesis of three clinically distinct common inflammatory disorders of medical importance.

1391F

The use of composite bi-allelic markers revisited: Creating highly informative multi-allelic markers from SNP and next generation sequence data. *A. F. Wilson, T. -H. Schwantes-An, C. M. Justice, A. J. M. Sorant.* Genometrics Section, CSGB, NIH/NHGRI, Baltimore, MD.

The renewed interest in using linkage analysis to investigate the role of rare variants in family data is hampered by the fact that the bi-allelic single nucleotide variants (SNVs) available from SNP arrays and next generation sequence data are not, by themselves, very informative. In the era of short tandem repeat polymorphisms (STRPs), the informativeness of these multi-allelic markers, as measured by the Polymorphic Information Content (PIC) was substantially higher than that of even the most common SNPs, resulting in a high proportion of matings that were informative for linkage analysis. The lack of informativeness for SNPs can be ameliorated with the use of multipoint linkage analysis, although this can be computationally intensive. Before the STRP era, bi-allelic markers in multi-locus systems like the MN and Ss blood group systems were used to create composite markers with relatively high PIC levels, increasing the proportion of informative matings for linkage analysis. Wilson and Sorant [2000] showed that the use of composite bi-allelic markers can be nearly as powerful in detecting linkage compared to the use of multi-allelic loci or multipoint methods. For example, a two or three locus composite bi-allelic marker was nearly as powerful to detect linkage with sib-pair linkage analysis as a multi-allelic marker and the power to detect linkage with a four-locus composite marker was similar to that using a multipoint approach. In this study, the bi-allelic variants for the MN and Ss loci, glycophorin A (*GYPA*) and glycophorin B (*GYPB*), respectively, are used as a model for the selection of bi-allelic variants that can be used to create multi-allelic composite markers. *GYPA* and *GYPB* are known to be about 90 kb apart and separated by a recombination hotspot (recombination rate of about 25 cM/Mb). HapMap markers were used to identify recombination hot and cold spot regions over the entire genome. Pairs of common SNPs with a minor allele frequency > 0. 4 and separated by a recombination hot spot were selected as composite marker candidates. Four sets of composite markers were generated, one for each of the HapMap reference populations (CEU, YRI and CHB) and one over all three populations. On average, 2144 two-locus composite markers, separated by approximately 0. 5 Mb were selected. Lists of composite markers will be posted on the NHGRI web-site. These markers will be useful for linkage analysis.

1392W

Regional IBD Analysis (RIA): a new method for linkage analysis in extended pedigrees using genome-wide SNP data. *H. J. Cordell[1], J. Eu-ahsunthornwattana[1,2], R. A. J. Howey[1].* 1) Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, NE1 3BZ, United Kingdom; 2) Division of Medical Genetics, Department of Internal Medicine, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Ratchathevi, Bangkok 10400, Thailand.

The study of rare variants has revived interest in linkage analysis. However, exact calculations for traditional linkage analysis are computationally impractical in large, extended pedigrees. Although simulation-based methods can be used, they require significant computation and are not exact. We propose Regional IBD Analysis (RIA), a non-parametric linkage method based on comparison of locally and globally estimated identity by descent (IBD) sharing in affected relative pairs (ARPs). In RIA, genome-wide SNP data are used to calculate the "global" expected IBD sharing probabilities for each ARP, against which a "local" set of IBD probabilities, estimated using SNP data within windows of pre-specified width, are compared. The global and local IBD probabilities are used to construct a non-parametric maximum likelihood statistic (MLS)-like test of linkage in each window. We illustrate our method with real nuclear-family data from a study of vesicoureteral reflux and simulated data based on large extended pedigrees from a study of visceral leishmaniasis. RIA successfully detected the linkage signals with significant reduction in computational time (e. g. 2 hours vs 66 hours on 3,626 individuals from 308 extended families, genotyped at 545,433 SNPs) compared to traditional methods. RIA should be useful in studies involving large extended families, in which traditional linkage analysis is not feasible. Additionally, because it does not rely on prior knowledge about familial relatedness, RIA has an additional advantage of being robust to pedigree misspecification and can be used even in the absence of pedigree information.

1393T

Using genealogy clusters to find high-penetrant disease variants in the Danish population. *A. Rosengren[1], A. Buil[1], M. Bertalan[1], JH. Thygesen[1], PB. Mortensen[2], C. Bøcker Pedersen[2], T. Werge[1], SSI, BROAD & iPSYCH investigators.* 1) Institute of Biological Psychiatry; Mental Health Services of Copenhagen, Copenhagen, Denmark; 2) National Centre for Register-based Research, Aarhus University, Aarhus, Denmark.

The Danish genealogy of recent times is established using information from the national Danish civil registry system. The registry includes approximately 9 million people alive in Denmark from 1968 onwards, their birth year and place as well as legal links between parent and offspring. The connection of siblings of mothers born since 1935 is nearly complete. This allows us to construct meiotic clusters of up to 6 generations, and to define an index pedigrees as all individuals with whom an index person share a common ancestor in the genealogy. We identify index pedigrees for all subjects recorded with a diagnosis of mental disorders in near complete, nationwide Danish Psychiatric Central Research Register*. We estimate heritability across all index pedigree for mental disorders, separately and joined using the genealogy information as co-variance matrix. We use the heritability estimates to examine whether heritability estimates correlate with time and place of birth of the index person, structure of the index pedigree, and with sub-diagnostic phenotypes, e. g. age-at-onset, hospitalization, treatment outcome and comorbidity. Also, we use the complete dataset to derived powerful estimates of genetic correlations between disease phenotypes to obtain measurements of the shared genetic effects among mental disorders. In particular, we will use genotype data from the approx. 80,000 index persons genotyped in iPSYCH initiative^ to correlate local heritability estimates to the underlying genetic architecture (e. g. polygenetic inheritance, high-penetrant variants etc.) and to guide IBD and NGS initiatives for the discovery of high-penetrant disease variants. The study offers the first compilation of the 9 mio. individuals in the know Danish genealogy linking individuals affected with severe mental disorders. The results provide a unique resource for population-bases disease studies. ** The near complete, nationwide Danish Psychiatric Central Research Register contains more than 4mio. hospital contacts and 15mio observations of psychiatric nature divided on roughly 900,000 unique individuals in the period 1969-2013. ^ The iPSYCH initiative includes all individuals in the Danish population born 1981-2006 with a diagnosis of AUT, SCZ, ADHD, BPD, MDD or anorexia as well as a randomly selected population sample, in total more than 80,000 individuals.*

1394F

A two-stage rare variant association test for family data with quantitative traits. *Y. Jiang[1], Y. Ji[1], A. Sibley[1], Y. Li[1,2], A. Allen[1].* 1) Biostatistics and Bioinformatics, Duke University, Durham, NC; 2) Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC.

Confounding due to population substructure is always a concern in genetic association studies. While methods have been proposed to adjust for population stratification in the context of common variation, it is unclear how well these approaches will work when interrogating rare variation. Family-based association tests are known to be robust to population stratification and have been well-developed, particularly for common variants. Here, we adopted the concept of the orthogonal model described in Abecasis et al. (2000) to decompose genetic effects into between and within-family components. While atest of the within-family component is robust to population stratification, this within-family test ignores between-family information potentially leading to a loss of power. We, therefore, propose a family-based two-stage rare-variant test for quantitative traits. We first construct a weight for each variant within a gene, or other genetic unit, based on score tests of between-family effect parameters. These weights are then used to weight the total within-family efficient score of all variants to form a score test of within-family effect parameters. Since the between-family and within-family tests are orthogonal under the null hypothesis, this two-stage approach can increase power while still maintaining validity. Using simulation, we show that this two-stage test can significantly improve power while correctly maintaining type I error. We further show that the two-stage approach maintains the robustness to population stratification of the within-family test and we illustrate this using simulations reflecting samples comprised of continental and closely related subpopulations.

1395W

Three networks comprised of both multiple sclerosis and obesity-associated genes demonstrate strong evidence of protein-protein interaction and significantly increase disease susceptibility: Results from the Kaiser Permanente MS Research Program. *M. Gianfrancesco[1], X. Shao[1], B. Rhead[1], L. Shen[2], H. Quach[1], A. Bernstein[3], C. Schaefer[2,4], L. F. Barcellos[1,2].* 1) Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, CA; 2) Kaiser Permanente Division of Research, Oakland, CA; 3) Palm Drive Hospital, Sebastopol, CA; 4) Research Program on Genes, Environment and Health, Kaiser Permanente, Oakland, CA.

Multiple sclerosis (MS) [MIM 126200] is an autoimmune disease that involves both genetic and environmental risk factors, such as early life obesity. However, the underlying biological mechanism mediating the association between obesity and MS has not been determined. The Disease Association Protein-Protein Link Evaluator (DAPPLE, v2. 0, Broad Institute) was utilized for bioinformatics analysis to examine whether protein-protein interactions exist between established genome-wide significant gene regions associated with both body mass index (n=97) and MS disease susceptibility (n=110). Gene sub-scores based on resulting protein networks were derived by summing the number of risk alleles across each representative gene within each network. Logistic regression was used to test for association between each network sub-score and MS susceptibility, controlling for smoking, education, year of birth, ancestry based on principal components analysis (EIGENSTRAT), and carrier status for *HLA-DRB1*15:01*, the strongest genetic risk factor for MS. Participants included non-Hispanic White members of Kaiser Permanente, Northern California (KPNC; 1,200 MS cases, 10,000 controls). Common and rare variant data for each candidate gene were obtained through whole-genome profiling and imputation. Results revealed significant evidence for direct and indirect protein-protein interactions (p<0. 001 and p=0. 01, respectively); of 99 total direct interactions observed, 22 were between obesity and MS genes. The interactions clustered into 13 networks; five were significantly associated with MS susceptibility in the KPNC dataset after controlling for covariates (p<0. 05). Three significant networks included both obesity and MS genes, with evidence of protein-protein interaction between *DMXL2* and *RAB3A* (odds ratio [OR] =1. 18, 95% CI 1. 10, 1. 27; p<0. 001), *NPEPPS* and *GBE1* (OR= 1. 07, 95% CI 1. 01, 1. 15; p=0. 04), as well as a network of 38 genes including *ERBB4*, *NEXN*, *NFKB1*, and *IL7R* (OR=1. 08, 95% CI 1. 06, 1. 10; p<0. 001). This study is the first to characterize complex gene contributions to MS susceptibility while accounting for other established genetic and environmental risk factors, and establishes the first evidence for potential underlying biological mechanisms between obesity and MS through bioinformatics analysis of protein-protein networks.

1396T

Sovereignty Hypothesis: Strong effect of HLA shared epitope alleles in developing risk for rheumatoid arthritis is due to massive gene-gene interactions. *L. Padyukov[1], L. Folkersen[2], K. Shchetynsky[1], S. Uebe[3], L. -M. Diaz-Gallo[1].* 1) Rheumatology Unit, Karolinska Inst, Stockholm, Sweden; 2) Department of Systems Biology, Technical University of Denmark, Kongens Lyngby, Denmark; 3) 3. Human Genetics Institute, Universitétsklinikum Erlangen, Erlangren, Germany.

Introduction. Rheumatoid arthritis (RA) is relatively common disabling systemic autoimmune disease characterised by join inflammation and destruction. A particular subset of *HLA-DRB1* alleles (*01, *04 and *10) with shared amino acid sequence, known as share epitope (SE), is the most important genetic contributor to the developing of anti-citrullinated protein antibody (ACPA) positive RA although the mechanism impelling such high effect is not clear. We hypothesize that the fraction of non-HLA SNPs with individually low contribution to the risk of RA are in interaction with HLA SE alleles and this "sovereign" is representing a sum effect, reflected in the high risk value of HLA in RA. **Methods.** We performed analysis of two independent RA cohorts from EIRA and NARAC studies with 2018 ACPA positive RA patients and 2273 healthy control individuals. Genome wide genotyping was previously performed on Illumina platform and imputation was based on 1000G reference. We compared group of SNPs with significant association with disease to non-associated SNPs. The interaction between SNPs and the SE alleles was calculated using the attributable proportion (AP) measure in GEISA and the distribution of the observed AP p-values was compared by the Kolmogorov-Smirnov (KS) test. **Results.** We found an enrichment of significant interactions between the SE and the group of SNPs associated with RA both in EIRA and NARAC cohorts. Statistical evaluation of distribution of AP p-values reveals significant difference between associated and not associated groups of SNPs (KS test: D=0. 29 for the EIRA cohort and D=0. 48 for the NARAC cohort, p<2. 2e-16 for both). The analysis of RNA expression in PBMC for *LINC00355* and *YTHDC2* genes (marked by proximal ~200Kbp SNPs), that are overlapping for interaction in both cohorts, indicates SE alleles dependent eQTLs, thus suggesting new candidate genes for functional study of RA in context of HLA. **Conclusion.** Our data is in favour of Sovereignty hypothesis suggesting that the high SE risk effect in ACPA positive RA could be at least in part attributed to its interaction with non-HLA genetic risk factors. A number of non-HLA genetic risk factors does not have visible effect on development of RA in absence of SE alleles and it could explain the difficulty in detection of the risk alleles and incompleteness of heritability picture of RA based on the univariate analysis of genetic variations.

1397F

Construction and inference of large-scale biological network using Gaussian graphical model. *W. Chen[1,3], Z. Ren[2], T. Wang[1], Y. Ding[3].* 1) Division of Pulmonary Medicine, Allergy and Immunology Children's Hospital of Pittsburgh of UPMC, Department of Pediatrics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Statistics, University of Pittsburgh, Pittsburgh, PA; 3) Department of Biostatistics, University of Pittsburgh, Pittsburgh, PA.

The analysis of biological networks provides additional information to understand human diseases beyond the traditional analysis that focuses on single SNP, gene, or protein. Gaussian graphical model (GGM), a probability model that characterizes the conditional independence structure of a set of random variables by a graph, has wide applications in the analysis of biological network, such as inferring gene-gene interaction or comparing differential networks. However, existing approaches cannot perform formal statistical testing and are usually inefficient for high-dimensional data that include tens of thousands of variables. Recently, we proposed a novel regression approach to obtain asymptotically efficient estimation of conditional dependences between each pair of variables (Ann. Stat. 2015, 43(3)). To our knowledge, It is the first theoretical work that produces p-value and confidence interval of each edge simultaneously in the graph. In this project, we illustrate the application of this method on network analysis with a novel and fast algorithm to perform the asymptotic estimation of GGM without any approximation. In simulation study, we demonstrate that the algorithm is faster in several orders of magnitude than existing methods without losing any accuracy or efficiency. Then, we apply our method to two real data sets: the proteomic data from a study of Alzheimer's disease and the transcriptomic data from a study of childhood asthma. We estimate the global molecular interaction networks for disease and health samples. The results from interaction networks identify interesting protein-protein interactions that would be overlooked by marginal correlation analysis. The differential networks between cases and controls in asthma study show functional relevance to the diseases. In conclusion, we develop a statistically sound and computationally efficient approach for constructing Gaussian graphical model and making inference for high-dimensional biological data. The algorithm has been implemented in an R package.

1398W

Data-Driven Genetic Encoding: A robust approach for detecting diverse action in main effect models and genetic interactions. *M. A. Hall[1], J. R Wallace[1], S. S. Verma[1], Y. Bradford[1], S. Pendergrass[2], M. Brantley[3,4], M. Brilliant[5], L. Chen[6], J. Denny[7,8], R. Frost[9], M. G. Hayes[10], T. Kitchner[5], C. McCarty[11], B. Muller-Myhsok[12,13,14], P. Peissig[5], D. M. Roden[15], K. Van Steen[16], J. H. Moore[17], M. D. Ritchie[1,2].* 1) Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA; 2) Geisinger Health System, Danville, PA; 3) Department of Ophthalmology & Visual Sciences, Vanderbilt University School of Medicine, Nashville, TN; 4) Department of Molecular Physiology and Biophysics, Vanderbilt University School of Medicine, Nashville, TN; 5) Marshfield Clinic, Marshfield, WI; 6) Vanderbilt Institute for Clinical Translational Research, Vanderbilt University, TN; 7) Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN; 8) Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN; 9) Department of Genetics, Institute for Quantitative Biomedical Sciences, Geisel School of Medicine, Dartmouth College, Hanover, NH; 10) Division of Endocrinology, Metabolism, and Molecular Medicine, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 11) Essentia Rural Health, Duluth, MN; 12) Max Planck Institute of Psychiatry, Munich, Germany; 13) Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; 14) University of Liverpool, Institute of Translational Medicine, Liverpool, UK; 15) Department of Medicine and Pharmacology, Vanderbilt University School of Medicine, Nashville, TN; 16) Department of Electrical Engineering and Computer Science, Montefiore, Belgium; 17) Institute for Biomedical Informatics, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA.

Certain assumptions are made about the biological action of a SNP when choosing a traditional genetic encoding. For each encoding, risk incurred by one copy of the alternate allele in relation to two copies varies: the heterozygous genotype (HET) is coded to bear 0%, 50%, and 100% the risk of homozygous alternate (HA) for recessive (REC), additive (ADD), and dominant (DOM) encodings, respectively. However, the actual HET action for a given SNP may incur any portion of the risk of HA. Further, choosing just one encoding is not flexible to the diversity of action in biology; while running every encoding raises the multiple testing burden. We present a novel, robust alternative: the data-driven genetic encoding. Here, a HET value is assigned based on the action a SNP demonstrates in the data, derived by its DOM and REC effects, thereby providing an individualized encoding for every SNP. We compared power of this method to detect main effect and genetic interactions using a comprehensive combination of simulated genetic models and found the novel method to outperform the traditional methods for the highest number of underlying models across varying minor allele frequencies (MAFs). We additionally compared the method to the codominant (CODOM) encoding and found that in situations where CODOM loses power (for low MAF) our method retained sufficient power. We further tested our method with null and main effect-only simulated datasets and found that our method maintained a low false positive rate (FPR) for likelihood ratio test (LRT) p-value, while the ADD and DOM encodings demonstrated inflation (as high as 16% FPR). Finally, we applied our method using age-related macular degeneration (AMD) cases (2,167) and controls (10,986) derived from the Electronic Medical Records and Genomics (eMERGE) Network. We identified 438 SNP-SNP models that were significant when allowing for a 5% false discovery rate. One top model included genes that encode *CFH* and *C2*, both with previously reported involvement in AMD. Our results reveal that caution is needed when using the ADD and DOM encodings for interactions due to the inflation observed. Further, we demonstrate the utility of our novel encoding method at flexibly assigning action to individual SNPs, identifying interactions between SNPs with diverse action, and uncovering examples of SNP-SNP interaction in natural data.

1399T

Sparse Functional Structural Equations for Causal Genotype-Phenotype Network Analysis with Next-Generation Sequencing Data. *M. Rahman[1], P. Wang[2], L. Jin[2], M. Xiong[1].* 1) University of Texas School of Public Health, Houston, TX; 2) Fudan University.

Despite their differences in selection of specific methods for estimation, the widely used methods for genetic analysis of complex traits do not detect, distinguish and characterize the true biological, mediated and spurious pleiotropic effects and are unable to unravel causal structures among multiple phenotype and genotype variants. To overcome these limitations, we develop sparse structural equation models (SEMs) as a general framework for a new paradigm of genetic analysis of multiple phenotypes. To incorporate both common and rare variants into the analysis, we further extend the sparse multivariate SEMs to sparse functional SEMs. To improve computational efficiency and reduce data dimension, functional data analysis techniques and the alternative direction methods of multiplier (ADMM) are used to develop a novel sparse two-stage least square estimation method for the structure and parameter estimation of large SEMs. Borrowing causal information from the SEMs and maximizing the power of marginal association analysis, we developed a novel statistic for testing association of genetic variants with multiple variants. Using large scale simulations we showed that the true network structure can be accurately recovered by our models and the new statistics have higher power than the PCA-based statistics. The proposed method is applied to exome sequence data from the NHLBI's Exome Sequencing Project (ESP) with 11 phenotypes, which identifies a network with 140 genes connected to 11 phenotypes and 15 genes with pleiotropic genetic effects and demonstrates that the proposed statistic has smaller P-values than the PCA-based statistics for testing marginal associations.

1400F

Modelling the IGF-Axis as a gene-network to identify early predictors of elevated blood pressure in children and adolescents. *P. G. Parmar[1], C. E. Pennell[2], L. J. Palmer[3], L. Briollais[4].* 1) Department of Biostatistics and Epidemiology, Auckland University of Technology, Auckland, Auckland, New Zealand; 2) School of Women's and Infants' Health, The University of Western Australia, Perth, Western Australia, Australia; 3) School of Public Health, University of Adelaide, Adelaide, Australia; 4) Lunenfeld-Tanenbaum Research Institute, University of Toronto, Ontario, Canada.

Genes of the Insulin-like Growth Factor (IGF) network have been shown to be associated with blood pressure and cardiovascular health. We aim to further current knowledge relating to the molecular mechanisms underlying the developmental origins of hypertension by accounting for some genetic complexity of Systolic Blood Pressure (SBP) by modelling the IGF-Axis as a gene-network and provide insight for the early onset of hypertension in growing children. We applied the gene-network methodology from Tsonaka *et al* using longitudinal data on growth and SBP across childhood and adolescence (5-18 years) from the Western Australian Pregnancy and Birth Cohort (Raine) Study. We found 27 gene-gene interactions were significant at p<0. 05, three were present in children aged ≤15 years (males only), two of which did carry over into adolescence (IGF1-IGFBP4 and IGF2-IGFBP3, p<0. 0001). The remaining 24 interactions were present in children aged >15 years (15 in males and nine in females). Of those, nine were significant to p<0. 0001 (six in males and three in females). We then investigated the gene-network profile of those who were classified within the highest 5% of SBP measures recorded across childhood and adolescence. We identified again, noticeable sex- and time-specific effects. These extreme (high) SBP profiles were associated with the following interactions; for females (under the age of 15 years) IGFBP3-IGFBP4 (p=0. 0076), for females (aged over 15 years) IGF1-IGF2 (p=0. 0014), IGF1-IGFBP3 (p=0. 0039), IGF1R-IGFBP5 (p=0. 0036) and for males (aged over 15 years) IGF2-IGFBP1 (p=0. 0080). The results attained are reasonable and align logically with current literature, particularly as IGF1 has been previously linked to hypertension and IGFBP1 has been linked with both positive and negative associations to SBP and to well-known risk factors of cardiovascular diseases. We found that this gene-network is modified with age; this in itself may be due to many reasons such as diet, hormones and developmental growth over time, particularly post-puberty. Further investigation using this method in cohort and consortium data would be ideal to validate our findings, improve accuracy surrounding the estimates produced here, and further improve the power to detect complex interactions. Through characterizing the association between multiple genes and disease outcomes we will offer new insight into disease aetiology whilst providing tools for making individualized treatment decisions.

1401W

Genome-wide search for gene-gene interactions in asthma susceptibility. *W. Murk, A. DeWan.* Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, CT 06510, USA.

PURPOSE: Asthma is a highly heritable, common complex disease. Although many studies have attempted to identify gene-gene interactions in this disease, such studies have typically been conducted by only considering a small number of candidate genes, and there have been few genome-wide searches for interactions. We aimed to conduct a genome-wide gene-gene interaction study for asthma susceptibility using data generated from the GABRIEL Consortium [1]. **METHODS**: Three sets of asthma case-control data were used: a screening dataset (N=1,625 subjects), a follow-up dataset (N=5,264 subjects), and a replication dataset (N=230 subjects). All possible pairwise interactions among 301,547 SNPs were evaluated in the screening dataset; those with a screening interaction p < 10^-5 were evaluated in the follow-up dataset. The most significant interaction in the follow-up was subjected to further analysis in the replication dataset. Interactions were assessed via logistic regression modeling. **RESULTS**: From N = 4. 6 × 10^10 interactions tested in the screening dataset, N = 3. 5 × 10^5 met criteria for follow-up. None of these interactions passed a strict significance threshold of p < 1. 4 × 10^-7 in the follow-up analysis. The most significant interaction in the follow-up (involving rs910652 and rs11684871) had an interaction p-value of 1. 7 × 10^-6. Interestingly, rs910652 is located within 78 kilobases of *ADAM33*, which is one of the most well studied asthma susceptibility genes. The significance of this interaction was confirmed (p = 0. 025) in the replication dataset. **CONCLUSIONS**: Using a genome-wide approach, we identified and replicated a gene-gene interaction in asthma that potentially involves a highly plausible risk locus. This study provides evidence that non-additive effects among SNPs contribute to asthma susceptibility. **REFERENCE**: [1] Moffatt et al. N Engl J Med 2010. 363:1211-21.

1402T

A novel gene-level test of association using common variants. *P. Nakka, B. J. Raphael, S. Ramachandran.* Brown University, Providence, RI.

Genome wide association (GWA) studies have been used widely to identify susceptibility loci for complex phenotypes under the hypothesis that single mutations of large effect generate a phenotype of interest. However, complex diseases and traits are known to exhibit genetic heterogeneity on multiple levels: 1) within a causal gene, the phenotype may be generated by multiple mutations; and 2) mutations in distinct genes within a pathway may interact to cause the disease state. In both cases, gene-level tests of association and pathway analysis are necessary to identify candidate variants for experimental validation. However, in order to carry out gene-level tests using common variants, the SNP-level *p*-values need to be combined into a gene-level score. Here we introduce a new method to calculate gene-level scores that are*p*-values, calculated analytically from a null chi-square distribution that captures linkage disequilibrium (LD) between SNPs in a gene. We compare our new approach to three commonly used methods: minSNP, in which the lowest SNP *p*-value in a gene is chosen to be the gene-level *p*-value; permSNP, which calculates an empirical *p*-value using case-control permutations; and VEGAS, which uses simulations from a multivariate normal distribution whose variance is the empirical LD among SNPs in each gene. These methods have substantial drawbacks that can affect downstream pathway analysis including gene length bias, imprecision and computational inefficiency. We find that our method is unbiased by gene length, produces *p*-values with higher numerical precision by as much as 10 orders of magnitude and runs twice as fast as VEGAS. We conduct simulations to assess the sensitivity and specificity of our method, and find that our method has 30% higher true positive rate than minSNP when the false positive rate is fixed at 1%. Pathway analysis using HotNet2 uncovers gene interaction subnetworks containing a significant number of associated variants in cases (*p*-value <0. 05) for three traits: Attention-Deficit/Hyperactivity Disorder (ADHD), Ulcerative Colitis (UC) and Waist-Hip Ratio (WHR). We find gene networks associated with UC and WHR that have been functionally validated and novel gene associations for ADHD, UC and WHR, which may point to the underlying biological mechanisms and present potential new drug targets. Using gene scores from our method in HotNet2 analysis uncovers significantly associatedgene networks that are missed by minSNP and VEGAS.

1403F

Causal gene-gene and gene-environment interaction network analysis- A new generation of genetic interaction studies. *M. Xiong[1], P. Wang[3,1], Y. Zhu[2].* 1) Dept Biostatistics, Univ Texas Hlth Sci, Houston, TX; 2) tulane university; 3) Fudan University, China.

Most genetic analyses of phenotypes have focused on analyzing single traits or, analyzing each phenotype independently. However, multiple phenotypes are highly correlated. It has been reported that genetic pleiotropic effects likely play a crucial role in the molecular basis of correlated phenotypes. Joint gene-gene (GxG) and gene-environment (GxE) interaction analysis of multiple complementary traits will increase statistical power to identify GxG and GxE interactions, and hold the key to understanding the complicated genetic structure of the complex diseases. Despite their importance in uncovering the genetic structure of complex traits, the current marginal approaches for interaction analysis cannot distinguish between direct and indirect interactions, which will decrease our power to unravel mechanisms underlying complex traits. To overcome these limitations, we propose to use directed acyclic graphs (DAGs) as a major concept and a general framework for interaction analysis of multiple phenotypes to change the paradigm of traditional association guided interaction analysis. An essential issue for using DAGs to study interaction of multiple phenotypes is how to accurately and efficiently estimate the structure of DAGs from observational data. Structure learning of DAGs has been shown to be NP-hard. To obtain the causal graph from observation data as close to the biological causal graph as possible, we propose "score and search"-based methods for exact learning causal graphs of interaction to find the best-scoring structures for a given dataset. We develop novel functional structural equation models as a tool for defining score function and formulate the exact learning causal graphs of interaction as an integer programming. We develop efficient branch and bound algorithm to solve the formulated integer programming. To evaluate the feasibility of the proposed methods for causal GxG interaction analysis of multiple phenotypes, the developed methods are applied to a sample of 1,011 European-Americans (EA) with complete exome sequencing and 11 phenotypes. We identify a causal network with 22 nodes and 62 directed edges in which 39 gene-gene interactions cause the variation of 11 phenotypes. We also identify a hub gene CSMD1 formed interactions with 6 other genes which affect 5 phenotypes. The developed concept of causal inference for GxG and GxE interactions of multiple phenotypes is novel and powerful.

1404W

Multivariate Functional Regression Models for Epistasis Analysis. *F. Zhang[1], D. Xie[3], M. Xiong[2].* 1) Hohai University, Changzhou, China; 2) University of Texas School of Public Health; 3) Hubei University of Chinese Medicine.

To date, most genetic analyses of phenotypes have focused on analyzing single traits or, analyzing each phenotype independently. However, joint epistasis analysis of multiple complementary traits will increase statistical power, and hold the key to understanding the complicated genetic structure of the complex diseases. Despite their importance in uncovering the genetic structure of complex traits, the statistical methods for identifying epistasis in multiple phenotypes remains "fundamentally unexplored ". To fill this gap, we formulate a test for interaction between two gens in multiple quantitative trait analysis as a multiple functional regression (MFRG) in which the genotype functions (genetic variant profiles) are defined as a function of the genomic position of the genetic variants. We use large scale simulations to calculate its type I error rates for testing interaction between two genes with multiple phenotypes and to compare its power with multivariate pair-wise interaction analysis and single trait interaction analysis by a single variate functional regression model. To further evaluate its performance, the MFRG for epistasis analysis is applied to five phenotypes and exome sequence data from the NHLBI's Exome Sequencing Project (ESP) to detect pleiotropic epistasis. A total of 136 pairs of genes that formed a genetic interaction network showed significant evidence of epistasis influencing five traits. The results demonstrate that the joint interaction analysis of multiple phenotypes has much higher power to detect interaction than the interaction analysis of single trait and may open a new direction to fully uncovering the genetic structure of multiple phenotypes.

1405T

Genetic interaction networks reveal novel Type 2 Diabetes networks in the Marshfield PMRP. *R. Li[1], L. Bao[1], S. Dudek[1], D. Kim[1], M. Hall[1], Y. Bradford[1], P. Peissig[2], M. Brilliant[2], J. Linneman[2], C. McCarty[3], M. Ritchie[1,4].* 1) The Pennsylvania State University, University Park, PA; 2) Marshfield Clinic, Marshfield, WI; 3) Essentia Rural Health, Duluth, MN; 4) Geisinger Health System, Danville, PA.

Type 2 diabetes (T2D) accounts for over 90% of all cases of diabetes and is rapidly increasing around the world. Previous efforts to identify the genetic architecture of T2D led by Genome-Wide Association Studies (GWAS) have identified numerous genetic variants associated with the disease; however, the marginal effects of detected variants only explain a small portion of the disease risk. Typically, association studies only consider one model at a time to discriminate disease and healthy individuals; in addition, genetic interactions are often unexplored. Here, we propose a new analysis approach that can identify different genetic interaction architectures for T2D and non-T2D individuals simultaneously. We first performed a GWAS on 3265 T2D samples obtained from the Marshfield Personalized Medicine Research Project (PMRP) to identify the most strongly associated genetic variants. Our GWAS results replicated many previously known associations including the *TCF7L2* gene and *Hippo signaling pathway*. Next, we used SNPs that are marginally associated with T2D ($p < 0.001$) to construct genetic interaction networks in T2D and non-T2D individuals simultaneously. To search for the optimal interaction networks of genetic variants, we used Grammatical Evolution Bayesian Network (GEBN) that uses an evolutionary algorithm to significantly reduce the computational burden associated with structure estimation of the network. As a result, GEBN can be easily scaled up to thousands of input variables. We separately applied GEBN to T2D and non-T2D individuals and discovered different interaction networks for the two subgroups. Functional analysis revealed that T2D and non-T2D networks are enriched for different genes, GO terms, and pathways. A few notable examples include *CYP7A1, a* gene that controls synthesis of cholesterol, enriched only in the non-T2D individuals and *biosynthesis of unsaturated fatty acids* pathway only identified in the T2D individuals. We then combined the interaction networks and top GWAS associations to perform discriminant analysis for the T2D and non-T2D samples. We achieved an average AUC of 81% using cross validation, which presents a significant improvement over methods that only consider marginal effects. By combining the distinct genetic interaction networks in T2D and non-T2D and GWAS associations through discriminant analysis, we were able to better model the disease risk and gain novel insights in understanding the genetic architecture of T2D.

1406F

Testing for gene-by-environment interaction using a kernel-based score test. *F. Chen, S. Wang, L. Li.* BioStat Solutions, Inc. , Frederick, MD.

Background: Gene-by-environment interactions play an important role in both observational and interventional studies for precision medicine. Environmental factors such as smoking, radiation exposure and drug treatment may possibly have an interactive effect with genes on disease predisposition or drug response. Kernel-based approaches have been developed to test for the aggregate effects of genetic variants, but few methods are available to analyze interaction effects of a genomic region with environment factors. **Methods:** A novel kernel-based linear score test has been developed using a variance component approach to test for the interaction between the aggregate effects of genetic variants and environment. Specifically, both main effect of a gene and its interaction with an environment risk factor were modeled as variance components. A score test was employed to test for the interaction effect. Different types of kernels (e. g. Gaussian, identity-by-state (IBS), weighted IBS) have been considered in this approach. Type I error and power related to each type of the kernels were assessed and compared against existing methods in a simulation study. The approach has been applied to data from a genome-wide association study. **Results:** The proposed approach has preserved type I error and could achieve reasonable power, especially in the case of small sample size. This approach also showed higher computational efficiency. **Conclusions:** The novel kernel-based approach tests for the interaction between aggregate effects of genetic variants residing in a genomic region and environmental risk factors. The proposed approach provides a powerful tool to understand gene-by-environment interaction in genome-wide association studies or sequencing studies.

1407W

Detecting Interactions By Leveraging Genetic Ancestry. *D. S. Park[1], E. Y. Kang[2], I. Eskin[3], C. J. Ye[1], H. Aschard[4], E. Eskin[2], E. Halperin[3], N. Zaitlen[1].* 1) UC San Francisco, San Francisco, CA; 2) UC Los Angeles, Los Angeles, CA; 3) Tel Aviv University, Tel Aviv, Israel; 4) Harvard School of Public Health, Boston, CA.

Genetic association studies in humans have focused primarily on the identification of SNP effects contributing to the additive component of heritability. There is growing evidence that both gene-gene and gene-environment interactions contribute significantly to phenotypic variation in humans and model organisms. In addition to explaining components of missing heritability, interactions lend insights into the biological pathways regulating phenotypes and improve our understanding of their genetic architectures. Identification of interactions in human studies has been complicated by the multiple testing burden in the case of gene-gene interactions, and the lack of consistently measured environmental covariates in the case of gene-environment interactions. To overcome these issues we conduct SNP-by-ancestry tests of interaction in admixed populations. We use genetic ancestry (χ) as a proxy for unmeasured environmental covariates as well as SNPs with different frequencies in the ancestral populations of the admixed individuals. Our method is best illustrated with an example. Consider the case of a gene-environment interaction for an environmental covariate E. If E is correlated with χ ancestry can serve as a proxy for E (even if E is unknown or unmeasured), and can be analyzed for interaction in lieu of E. Many environmental covariates, such as socio-economic status, have been shown to be correlated with genetic ancestry. Ancestry can also serve as a proxy in the identification of gene-gene interactions for SNPs highly differentiated between ancestral populations. These high Fst SNPs will be correlated with χ and the resulting interaction test will maintain a 5x10-8 burden. To prevent false positive interactions driven by differential LD between ancestral populations we include local ancestry in our model. We applied our method to gene expression data from African American individuals in Corriell and ImmVar (Ye *et al.* 2014), which use LCLs and CD14 cells, respectively. In Corriell we found 2 genes (ERBB4, SGCG) that had a genome-wide significant ($p < 5x10-8$) interaction signal and 3 genes (MICU2, HBA1, HSF2BP) that displayed a moderate interaction signal ($p < 5x10-6$). For ImmVar there were no genes passing genome-wide significance but MAP3K7 showed a moderate interaction signal ($p < 5x10-6$), which replicated in the Corriell cohort.

1408T

Extracting socioeconomic data from electronic health records for gene-environment studies of blood pressure. *B. Hollister[1], E. Farber-Eger[1], D. Crawford[3], M. Aldrich[1,4,5], A. Non[1,2].* 1) Vanderbilt Genetics Institue, Vanderbilt University Medical Center, Nashville, TN; 2) Department of Anthropology, Vanderbilt University, Nashville, TN; 3) Department of Epidemiology & Biostatistics, Institute for Computational Biology,Case Western Reserve University, Cleveland, OH; 4) Department of Thoracic Surgery, Vanderbilt University Medical Center, Nashville, TN; 5) Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN.

Socioeconomic status (SES) is a fundamental variable contributing to health, particularly when analyzing racial disparities in health. SES data are rarely included in genetic studies, due in part to the difficultly of collecting these data when studies were not designed for that purpose. The emergence of large clinic-based biobanks linked to electronic health records (EHRs) provides research access to large populations with longitudinal phenotypic and exposure data captured in structured fields as billing codes, procedure codes, and prescriptions. SES in the EHRs, however, is often not explicitly recorded in structured fields. Rather, SES data are recorded in the free text of clinical notes and the content and completeness of these data vary widely by practitioner. To enable gene-environment studies that consider SES as an exposure, we sought to extract SES variables from BioVU, the Vanderbilt University Medical Center biorepository linked to de-identified EHRs. We developed an index of SES using information available within the de-identified EHR, including broad categories of occupation, education, insurance status, and homelessness. We performed free-text search across 15,865 individuals (11,521 African Americans; 1,714 Hispanics, 1,122 Asians, and 1,412 others) in BioVU for 22 different categories (534 unique terms) to extract relevant SES data. We identified 14,186 individuals with education information (89%) and 14,523 individuals with occupation information (92%). Examples of information collected include years of education an individual received, degrees earned, and latest recorded occupation. Insurance status was found for 75% of the individuals, and the term homelessness was found in the EHR for 43% of the individuals. We are in the process of investigating how the extracted SES data contribute to hypertension through interactions with genetic variants previously associated with hypertension from the Illumina Metabochip in African Americans. The SES data extraction approach and index developed here will enable future EHR-based genetic studies to feasibly integrate SES data into statistical analyses. Ultimately, increased incorporation of SES measures into genetic studies and examination of gene-SES interactions will help elucidate the impact of the social environment on common diseases.

1409F

Using Bayes model averaging to identify GxE interactions in genome-wide association studies. *L. C. Moss, D. V. Conti.* Preventive Medicine, University of Southern California, Los Angeles, CA.

Genome-wide association studies (GWAS) typically search for marginal associations between a single nucleotide polymorphism (SNP) and a disease trait while gene-environment (GxE) interactions remain generally unexplored. Numerous approaches exist to test for GxE interaction for either a single SNP or in a GWAS context. More powerful methods beyond the simple case-control approach leverage either SNP marginal effects or case-control ascertainment to increase power. However, these potential gains are accomplished under certain assumptions and it is often unclear if these assumptions are applicable *a priori*. Here, we review GxE methods, such as the case-only approach and more recently proposed two-step approaches, and use simulations to highlight performance as a function of marginal and interaction effect sizes, direction of effects, and the correlation of the two factors in the source population. Substantial variation in performance across methods leads to uncertainty as to which approach is most appropriate for any given analysis. Bayes model averaging offers a statistical foundation for incorporating model uncertainty and we present a framework that: (1) balances the robustness of a case-control approach with the power of the case-only approach; (2) leverages marginal SNP effects; (3) allows for the incorporation of prior information; and (4) allows the data to determine the most appropriate model. Within a log-linear model, we average over the inclusion of parameters corresponding to the marginal SNP and GxE interaction effects and the G-E association in controls. The resulting method exploits the joint evidence for marginal SNP and GxE interaction effects while gaining power from a case-only equivalent analysis. We use single-SNP and GWAS simulations to demonstrate that this method detects SNPs within a wide range of scenarios with the potential for increased power over current methods. We show the potential for increased power particularly for small marginal and interaction effect size combinations in various directions using this method. We demonstrate the approach on a genomewide gene-environment scan for asthma and lung function in the USC Children's Health Study.

1410W

Integration of text mining and epistasis analysis identifies new genes underlying atopy. *P-E. Sugier[1,2], M. Brossard[1,3], A. Vaysse[1,4], C. Sarnowski[1,3], M-H. Dizier[1,4], M. Lathrop[5], C. Laprise[6], F. Demenais[1,4], E. Bouzigon[1,4].* 1) INSERM UMR946, PARIS, France; 2) Université Pierre et Marie Curie, Paris, France; 3) Université Paris Sud, Paris, France; 4) Université Paris Diderot, Paris, France; 5) McGill University and Genome Quebec Innovation Centre, Montréal, Canada; 6) Université du Québec à Chicoutimi, Chicoutimi, Canada.

A few genome-wide association studies (GWAS) of atopy defined by skin prick test reactivity to allergens have been conducted in population-based cohorts and have led to the identification of at most 10 loci. Investigation of complex mechanisms such as gene-gene interactions and integration of prior-knowledge based on PubMed literature, in order to limit the multiple testing burden, may contribute to uncover new susceptibility genes. To identify new genetic determinants of atopy, we conducted in the French EGEA discovery dataset with validation in the French-Canadian SLSJ dataset a GWAS followed by epistasis analysis between the gene containing the most significant GWAS signal and a set of genes filtered according to the following criteria: (1) genes harbouring at least one SNP reaching a threshold of 10-4 in the GWAS; (2) genes selected in (1) and showing significant relationship through text mining using GRAIL (Gene Relationships Among Implicated Loci) applied to PubMed abstracts. The GWAS of atopy conducted in 1,660 EGEA subjects (including 925 atopics) with replication in 1,194 SLSJ subjects (with 675 atopics) identified a single genome-wide significant signal (Pmeta=9. 4x10-9) within adhesion G protein-coupled receptor V1 gene (*ADGRV1*). Based on the EGEA discovery GWAS results, 32 genes with at least one SNP with P≤10-4 were selected for cross-gene SNPx-SNP interaction analyses with *ADGRV1*. When using this first filtering criterion, 81,730 SNPxSNP interaction tests were performed in EGEA of which 403 reached the threshold of 5x10-3 and were taken forward for replication in SLSJ. Meta-analysis of the outcomes from these two datasets showed three SNP pairs with PmetaINT≤10-4. The GRAIL approach applied to the selected genes identified a relationship (PGRAIL<0. 1) between *ADGRV1* and three genes (*DNAH5*, *CHD7* and *ATP8B1).* This allowed reducing the number of SNPxSNP interaction tests by 9-fold. The same three SNP pairs previously identified had PmetaINT≤10-4. One of these pairs at *ADGRV1* and *CHD7* (chromodomain helicase DNA binding protein 7) reached the threshold of 2x10-5, which corresponds to the multiple-testing corrected threshold based on the text-mining filtering approach. This study highlights that integrating text mining and epistasis analysis facilitates the identification of new susceptibility genes. Funding: ANR-11-BSV1-027, ANR-USPC-2013-EDAGWAS.

1411T

Pathway analysis of next generation sequencing data based on a linear score test. *S. Wang, L. Li.* BioStat Solutions, Inc. , Frederick, MD 21703, MD.

Background: Over the past decade, there has been much focus on pathway analysis, which utilizes prior biological knowledge on gene function and relationship between genes. Such information is potentially useful for a better understanding of etiology of human diseases or mechanism of action in drug response. While gene expression microarray data have often been used, recent techniques have been developed to make use of genotype data, like Genome-wide association (GWA) and next generation sequencing (NGS) analysis. Methods: Our pathway analysis approach is tailored to NGS data. This approach is based on kernel machines which are often used in sequencing analysis. The method considers a kernel that defines genetic similarity of two individuals based on both, sequencing data and biological pathway information, and then uses a linear score test to assess statistical significance of the pathway effect on the outcome of interest. Based on a similar kernel setup, the application was also extended to test gene-gene (GxG) interaction effects while accounting for the gene main effect. Sample size and power of the proposed methods were assessed through multiple simulation scenarios, and we also applied the approach to sequencing data to demonstrate its application. Results: Based on the simulation, type I error was well controlled and the relative power has been demonstrated through different simulation scenarios. Conclusion: Different kernel strategies were implemented to identify either GxG interactions or profound pathway effects, and the proposed approaches have been built into a high-throughput analytics platform for NGS data analysis.

1412F

A stochastic search algorithm for finding multi-SNP effects using nuclear families. *C. R. Weinberg, M. Shi, A. Wise, D. M. Umbach, L. Li.* Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC.

Given that biologic systems typically involve failure-resistant redundancy, phenotypes such as birth defects may occur only through the joint effects of exposures and several genetic variants. Such joint effects tend to produce a weak signal in typical GWAS analyses that assess only single-SNP associations. We describe an approach that uses case-parent triads and applies an “evolutionary algorithm” to stochastically search the large sample space that includes all sets of size S of potentially-interacting SNPs. We assess the performance of our algorithm using simulated but realistic data from the dbGaP GWAS data on families with the birth defect oral cleft. We simulate specific multi-SNP causal effects and then try to recover those causative complexes *de novo*. Initial simulations using our method show promising results in scenarios that involve two sets of four interacting SNPs, against a background of random cases. We are applying the refined method, using the original cleft data to explore a large set of candidate SNPs for epistatic effects related to risk of oral cleft.

1413W

Detecting 3-way G x AGE x SEX interaction effects for obesity traits: Methodological evaluation of approaches and results from the GIANT consortium. *T. W. Winkler[1], A. E. Justice[2], M. Graff[2], L. Barata[3], M. F. Feitosa[3], I. M. Heid[1], K. E. North[2], I. Borecki[3], Z. Kutalik[4,5,6], R. J. F. Loos[7].* 1) Genetic Epidemiology, University of Regensburg, Regensburg, Germany; 2) Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514, USA; 3) Department of Genetics, Washington University School of Medicine, St Louis, Missouri 63110, USA; 4) Department of Medical Genetics, University of Lausanne, 1005 Lausanne, Switzerland; 5) Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland; 6) Institute of Social and Preventive Medicine, CHUV-UNIL, 1010 Lausanne, Switzerland; 7) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.

Overall and central obesity – as measured by body mass index (BMI) and waist-hip ratio adjusted for BMI (WHRa) - differ between men and women and change over time (e. g. following menopause in women). Previously, genome-wide association meta-analyses (GWAMAs) revealed hundreds of associated loci, but little is known about the genetic impact on the sex-specific change in obesity traits around menopause. Given two potential interaction factors (sex and age), it can be tempting to test 3-way G x AGE x SEX interactions genome-wide. However, we are lacking methods and software to identify such 3-way interactions. Therefore, we introduce a difference-of-difference (d-o-d) test that can be utilized to infer 3-way G x AGE x SEX interactions from GWAMAs stratified by AGE (cut-off at 50 years, i. e. average age of menopause onset) and by SEX. First, we performed exhaustive simulations and analytical computations to evaluate type 1 error and power of different approaches to identify d-o-d. We found that the genome-wide screen for d-o-d showed highest power for extreme 3-way interactions that involve opposite effects across AGE and SEX. We also found that filtering for a marginal joint two-degree of freedom association, i. e. , allowing for heterogeneity in genetic effects for one, but not for the other factor (e. g. same effect in men and women, but different across the age groups), followed by a test for d-o-d was best powered to identify biologically more plausible 3-way interactions. Examples of these are the 1-stratum (effect in one stratum, lacking in others) or the 3-strata (effect in three strata, lacking in one) interaction. Application of approaches to age- and sex-stratified GWAMA data (N > 320,000) for BMI and WHRa from the GIANT consortium did not identify any significant 3-way G x AGE x SEX interactions. Given the large power (> 80%) for extreme 3-way interactions, our analysis does not support existence of such loci. Power for biologically plausible 3-way interactions was limited (< 20%). Even larger sample sizes are required to identify such loci with sufficient power (e. g. , N > 750,000 to identify a medium-sized 1-strata interaction with 80% power). In summary, we recommend screening approaches to identify 3-way interactions and provide a software implementation in the freely available R package EasyStrata. We demonstrate that extremely large sample sizes are required to find biologically plausible 3-way G x AGE x SEX interactions for obesity traits.

1414T

Multivariate functional regression model for gene-gene interaction analysis of highly correlated gene expressions. *K. Xu[1], M. Xiong[1,2].* 1) 1. School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai, Shanghai, China; 2) 2. Human Genetics Center, Department of Biostatistics, the University of Texas School of Public Health, Houston, TX, USA.

Gene-Gene interactions are essential components of the genetic architecture of the gene expressions. But, gene-gene interaction analysis of gene expressions remains fundamentally unexplored due to great computational challenges and data availability. On the other hand, considering the fact that many biological synthetic process require the balanced expression of multiple genes, analyzing the regulation mechanism of highly correlated gene expressions is crucial. However, simultaneously analyzing gene-gene interactions for highly correlated gene expressions poses enormous challenges for methodologies and computations: 1) lack of methods for gene-gene interaction analysis of multiple gene expressions, 2) severe multiple testing, 3) heavy computation burden. To meet these challenges, we use multivariate functional regression method for testing gene-gene interaction and estimating its effect in regulating gene expressions. In this model, we take a gene as a basic unit of the interaction analysis which includes both common and rare variants. By thorough simulation, we demonstrate that this method can achieve the correct type 1 error and have a better strength to detect the interaction effect than the pairwise interaction analysis. The proposed methods are applied to the gene expression and next-generation sequencing data with the 360 European samples and 87 African Samples from the 1000 Genome Project. For the 242 genes in MAPK pathway which are shared in these datasets, we discovered 46 pairs of genes showing significant interaction effects (P-value < 1. 71×10-6) on the highly correlated genes in this pathway, in both European and African people. For further study, we are applying this method on the whole genome data.

1415F

A comparison of methods for inferring causal relationships between genotype and phenotype using multi-omics data. *H. F. Ainsworth[1], S. Y Shin[2], H. J. Cordell[1].* 1) Institute of Genetic Medicine, Newcastle University, Newcastle Upon Tyne, UK, United Kingdom; 2) MRC Integrative Epidemiology Unit (IEU), University of Bristol, Bristol, UK.

Many novel associations between common genetic variants and complex human disease have been successfully identified using genome-wide association studies (GWAS). However, a typical GWAS gives little insight into the biological function through which these associated genetic variants are implicated in disease. Indeed, rather than finding variants which directly influence disease risk, the variants implicated by GWAS are typically in linkage disequilibrium with the true causal variants. Understanding the causal role of the genetic variants in disease etiology and moving towards therapeutic interventions is not simple. Integration of additional data such as transcriptomic, proteomic and metabolomic data, measured in relevant tissue in the same individuals for whom we have genomic (i. e. GWAS) data, could potentially provide further insight into disease pathways. Yet, open questions remain on how to assess the causal direction of association between these variables.

We review currently available statistical methods for inferring causality between variables that use a genetic variant as a directional anchor. We consider Mendelian Randomisation, Structural Equation Modelling, a Causal Inference Test and several Bayesian methods. We present a simulation study assessing the performance of the methods under different conditions, assuming throughout that we have a single genetic variant and two phenotypic variants that are associated with one another, although the underlying causal relationship may vary. In particular, we consider how the causal inference is affected by the presence of common environmental factors influencing the observed traits.

1416W

Pathway Analysis of Genome-Wide Glaucoma Data. *J. N. Cooke Bailey[1], M. Butkiewicz[1], L. R. Pasquale[2, 3], M. A. Hauser[4, 5], R. R. Allingham[4], J. L. Wiggs[2], J. L. Haines[1], NEIGHBORHOOD Consortium.* 1) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, MA; 3) Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 4) Department of Ophthalmology Duke University Medical Center, Durham, NC; 5) Department of Medicine, Duke University Medical Center, Durham, NC.

Primary open-angle glaucoma (POAG) is a phenotypically and genetically complex neurodegenerative disease that is a leading cause of blindness worldwide. To better understand the disease mechanism and genetic architecture of POAG, we performed an intensive pathway analysis with the goal of elucidating potentially novel mechanisms of disease. We implemented an updated version of the Pathway Analysis by Randomization Incorporating Structure (PARIS) program, which bins variants based on linkage disequilibrium blocks of similar structure and size to assess cumulative effects across biological pathways. Utilizing genome-wide logistic regression results from the NEIGHBORHOOD Consortium meta-analysis of imputed GWAS data from 3,853 POAG cases and 33,480 controls, we interrogated four pathway databases: Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), Reactome, and NetPath. Because PARIS minimizes multiple testing and type I error concerns through the use of empirical genomic randomization to estimate significance, we are not obligated to correct for every variant evaluated and therefore set a significance threshold of P<0. 0001. We additionally set the number of permutations to 10,000. Our preliminary analysis which was restricted to pathways containing at least 3 genes identified several pathways of potential involvement in POAG beyond those previously reported; this includes 3/26 NetPath, 30/293 KEGG, 123/1748 Reactome, and 677/12765 GO pathways. From NetPath, the signal transduction pathways IL9, IL1, and TNFalpha show particularly intriguing results as 6 IL9, 21 IL1, and 28 TNFalpha genes were individually associated with P<0. 001 with very little overlap between the significant genes defining each pathway. From KEGG, the pathways with the highest percentage of associated features and with no genic overlap were the phenylalanine, tyrosine and tryptophan biosynthesis and the spliceosome pathways. Identifying novel POAG pathways will help to define the underlying mechanisms of POAG by identifying proteins and molecular pathways that influence pathogenesis; this information can then be utilized to identify biomarkers for early molecular diagnosis and treatment and improve prevention strategies. This work was supported by NIH/NEIEY022305.

1417T

A Highly Adaptive Test for Gene- or Pathway-Multivariate Trait Associationwith Application to Neuroimaging Data. *J. Kim, W. Pan.* Division of Biostatistics, University of Minnesota, Minneapolis, MN.

Testing for genetic association with multivariate traits has become increasinglyimportant, not only because of its potential to boost statistical power, but alsofor its direct relevance to some applications. For example, there is accumulatingevidence showing that some complex neurodegenerative and psychiatric diseaseslike Alzheimer's are due to disrupted brain networks, for which it would benatural to identify genetic variants associated with disrupted brain networks. In spite of its promise, testing for multivariate trait associations is challenging:if not appropriately used, its power can be much lower than testing on each univariatetrait separately (with a proper control for multiple testing). Furthermore, differing from most existing methods forsingle SNP--multiple trait association, we consider gene- and pathway-basedassociation testing for multiple traits, due to well known genetic heterogeneity andsmall effect sizes of individual SNPs. Because the power of a test critically dependson several unknown factors such as the proportions of associated SNPs, genes andtraits among those tested, we propose a highly adaptive test that data-adaptivelydetermines some optimal parameters in the test to yield high power across a widespectrum of situations. We compare the performance of the new test with severalexisting tests using both simulated and real data. We apply the proposed test tostructural MRI data drawn from the Alzheimer's Disease NeuroimagingInitiative (ADNI) project to identify genetic variants associated withthe human brain default mode network (DMN).

1418F

Covariate-selection approach in Genome Wide Association Studies for multivariate data. *Y. A. Tsepilov[1,2], G. Kastenmuller[3,4], R. Wang-Sattler[5,6], K. Strauch[7,8], C. Gieger[5,6,7], Y. S. Aulchenko[1,2].* 1) Institute of Cytology and Genetics SD RAS, Novosibirsk, Russia, 630090; 2) Novosibirsk State University, Novosibirsk, Russia, 630090; 3) Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 4) Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom; 5) Research Unit of Molecular Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 6) Institute of Epidemiology II, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 7) Institute of Genetic Epidemiology, Helmholtz Zentrum München - German Research Center for Environmental Health, Neuherberg, Germany; 8) Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig-Maximilians-Universitét, Munich, Germany.

Genome-wide association analysis (GWAS) is one of the most popular methods of identification of alleles that affect complex traits. Recent accumulation of functional genomics data ("ome" data, e. g. transcriptome, metabolome etc) could give a new insight into the functional role of specific changes in the genome. Such data require special statistical methods for their analysis, as they are characteristically high dimensional and contain statistical interactions and functional relationships between individual components, which usually reflect biological interactions. Development of such methods is of current importance as the progress of molecular biology techniques marches on. We introduce a group of approaches to evaluate the impact of genetic variants to "omics" phenotypes with relatively high power and rather modest computational intensities. It is based on the idea that if biological pathways and relationships in data could be reconstructed, we can use the knowledge about biochemical neighbors for chosen trait and include this information into analysis of this trait. The sources of biological relationship information can vary: biological pathway databases or reconstructed net from the data. We assessed our approach using real population study data from big German study KORA (n=1,784, 2M SNPs) with measured metabolomics data (151 metabolites). For reconstruction of the pathways we used Gaussian Graphical Models (GGM) which was shown as effective tool for the unbiased reconstruction of metabolic reaction. Previously it was demonstrated that using GGM-driven ratios instead of all ratios leads to comparable power with much less computational expenses. Our approach as well has shown the comparable power with all ratio approach with computational complexity similar to the analysis of original concentrations only (n instead of n2). Moreover we generalized this method using shrinkage regularization (RIDGE and LASSO) for phenotypic correction with different types of penalization (penalties proportional to GGM distances and penalties proportional to real biochemical distances extracted from databases). Using these types of models decreases the phenotype variances and the noise driven by network influences between metabolites with the dissolution of induced signals and the improvement of actual associations that helps with further functional annotation.

1419W

Identifying ethnic-group and SNP interaction in meta-analysis for sequencing association studies. *S. Wu[1], H. Wang[1], C. Lin[2], D. Chen[1], C. Chang[1], C. Fann[1].* 1) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan, Taipei, Taiwan; 2) Graduate Institute of Life Sciences, National Defense Medical Center, Taipei, Taiwan.

Meta-analysis is a routine approach in genome-wide association studies to analyze data across studies to increase power. Although large sample sizes can help guard against false-negative and false-positive results, a major potential problem is that allelic heterogeneity across populations could lead to dramatic reduction of power in detecting true positive signals. Therefore, the issue of population heterogeneity can not be ignored in meta-analysis. Currently, a widely used approach for adjusting population stratification (PS) is EIGENSTRAT method based on principal component analysis (PCA). The method identifies several top principal components (PCs) and uses them as covariates to adjust for PS. However, it is difficult to quantify interaction between ethnic-group and SNP. Very few studies have discussed the impact of ethnic-group and only coded them as a moderator variable in meta-analysis. In this study, we proposed an approach that incorporated ancestry informative markers (AIMs) information for investigating heterogeneity among populations and constructed a meta-regression model using ethnic-group as a moderator variable to detect interaction between ethnic-group and SNP in meta-analysis. In order to discern the population-heterogeneity markers among populations, 1000 genomes datasets were used as reference to create a panel of AIMs. The population differentiation measure, FST, was calculated for each SNP. In order to evaluate the efficacy of our proposed method, two simulation data structures were considered: (i) European and Asian populations; (ii) American and European populations. A few parameters such as minor allele frequency, the difference between two populations, the effect size for different ethnic groups, sample sizes and the numbers of studies were assumed to calculate type I error and power using simulation study. In our results, we found that fewer than 100 SNPs would be sufficient to determine the most representative population-heterogeneity markers among populations. In conclusion, our proposed method increase both sensitivity and specificity and hence reduce heterogeneity in meta-analysis.

1420T

New disease susceptibility loci identified by a multiple SNPs interaction search in GWAS. *Y. Yasui[1], N. Sharafeldin[1], Q. Liu[1], S. Jabbari[1], L. Wang[2], C. Franco-Villalobos[1], S. Mahasirimongkol[3], H. Yanai[3, 4], K. Tokunaga[3].* 1) School of Public Health, University of Alberta, Edmonton, Canada; 2) College of Agriculture & Biotechnology, Zhejiang University, China; 3) Department of Human Genetics, School of International Health, Graduate School of Medicine, University of Tokyo, Tokyo, Japan; 4) Fukujuji Hospital, Japan Anti-Tuberculosis Association, Kiyose, Japan.

A typical genome-wide association study (GWAS) explores SNP-disease associations following the common practice of analysis of a single SNP at a time in association with disease. We focused here on examination of *SNP-set* interaction effects within-genes on disease risk. We applied logic regression to search for SNP-set interactions using data from the Wellcome-Trust-Case-Control-Consortium (WTCCC). The genotype cluster intensity plots were generated for SNPs of statistically significant genes to exclude false-positive associations. Plots were visually inspected one by one: this was necessary to remove the genes discovered due to genotyping errors. The majority of the discovery had to be removed from the list of significant genes if one SNP or more in the logic trees had a genotype call that appeared erroneous. Genes showing strong evidence of association were: 13 for bipolar disorder (BD), 16 for coronary artery disease (CAD), 15 for hypertension (HT), 72 for rheumatoid arthritis (RA), 105 for type 1 diabetes (T1D) and 19 for type 2 diabetes (T2D). The identified loci included almost all strong single-SNP signals reported by WTCCC. On average 51% of recently reported GWAS meta-analyses signals across the six diseases overlapped with our findings. The top significant genes were: *NFIA* with BD, *CDKN2B* and *TPH1* with CAD, *COL4A4* with HT, *BTNL2* with RA, and *TCF7L2* and *TMEM155* with T2D. Our approach demonstrates the confirmatory and discovery value of considering higher order SNP-set interactions in addition to the standard single-SNP analysis in GWAS.

1421F

A Statistical Approach for Testing Gene by Microbiome Interactions. *N. Zhao, M. C. Wu.* Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA.

Although GWAS and next generation sequencing studies of human genetic profiles have helped elucidate the genetic component of complex traits and disease, a comprehensive understanding requires understanding the role of other risk factors, including the microbiome. Facilitated by sequencing technology, comprehensive profiling of the microbiome is now possible and has led to a range of traits being associated with microbiome composition. However, despite the importance of both microbiome and genotype as factors underlying many traits, there has been little work on assessing their joint effect. A problem of particular interest is identifying genetic variants that modify the effect of microbiome composition on traits and the discovery of genotypes for which the effect of microbiome composition is heterogeneous. Yet, how to test for statistical interactions between microbiome composition and genetic variants remains un-addressed. Therefore, we propose a new strategy for assessing the interaction between individual variants and microbiome using a regression based framework. In particular, following standard microbiome approaches, we encode microbiome as a matrix of pair-wise distances between individuals where the distance is defined based on a phylogenetically informed metric. The distance can then be transformed to be a kernel matrix of pairwise similarities between the individuals in the study. This, in turn, enables us to exploit the kernel machine framework (e. g. SKAT) to model the relationship between a quantitative trait and microbiome (modeled flexibly through the phylogenetically informed kernel), the main effect of a genetic variant, and their interaction. Modifications to the standard kernel machine framework are necessary to accommodate the dimensionality and complexity of microbiome composition data. Under this modified framework, we develop a score test for testing the interaction term. Simulations and real data applications show that our approach has reasonable power and correct type I error.

1422W

Efficient heritability partitioning via controlled down-sampling with the Haseman-Elston approximate regression. *X. Zhou.* Biostatistics, University of Michigan, ANN ARBOR, MI.

Although genome-wide association and sequencing studies have identified many SNPs associated with common diseases and complex traits, the biological mechanisms underlying these associations are largely unknown. Quantifying the proportion of phenotypic variance contributed by various SNP functional categories, or partitioning chip heritability, has thus become an initial important step towards understanding the molecular mechanisms of SNP associations. Here, we present a new method for such analysis. Our method builds upon the renowned Haseman-Elston regression for variance component estimation and requires only summary level statistics. More importantly, our method effectively uses only a small fraction of the sample (e. g. 1%), while still produces accurate results with only slightly increased confidence intervals (e. g. 10%). This allows our method to be efficiently applied to large scale association studies while producing calibrated test statistics. Compared with the competing approach (i. e. LD score regression), our method produces unbiased and more efficient estimator, has proper control for type I error in hypothesis test, and is faster. Using simulations and real data applications, we illustrate the benefits of our method and provide insights for the importance of several functional categories. The method is implemented in the GEMMA software package, freely available at www. xzlab. org/software. html.

1423T

Causal Genomic Network Analysis emerges as a New Generation of GeneticStudies of Complex Diseases. *Y. Zhu[1], P. Wang[2], J. Zhao[1], M. Xiong[2].* 1) Epidemiology, Tulane Univ Pub Health & Tropical Med, New Orleans, LA; 2) Division of Biostatistics, The University of Texas Health Science Center at Houston,School of Public Health, Houston, TX.

In the past decade, rapid advances in genomic technologies have dramatically changed the genetic studies of complex diseases. Genome-wide association studies (GWAS) have been widely used in dissecting genetic structure of complex diseases. However, numerous studies reported that the genetic loci identified by GWAS collectively explain only < 10% of genetic variation across the population. Although extremely large number of samples are collected and whole genome sequencing studies will be conducted very soon, which will lead to reducing he fraction of missing heritability, a large proportion of heritability will be still missing under the paradigm of single trait genetic analysis. The methods for heritability estimation and single trait genetic study paradigm are questionable. The heritability of individual phenotype cannot reveal complicated genotype-phenotype structure and is highly unlikely to fully capture the structure of heritability of multiple phenotypes. The concept of heritability should be extended from a single trait to multiple traits. The genetic studies of complex diseases also should be extended from a single trait analysis to multiple trait analysis. The popular methods for genetic studies of multiple traits are mainly based on correlation and association analysis. These methods cannot efficiently detect, distinguish and characterize the true biological, mediated and spurious pleiotropic effects. To overcome these limitations, we develop a directed acyclic graph (DAG)-based new framework and novel statistical methods for inferring causal networks of genotype-phenotypes with NGS data and detecting, distinguishing and characterizing the true biological pleiotropic, mediated pleiotropic and spurious pleiotropic effects of genetic variants. The propose methods for causal network analysis of genotype-phenotype were applied to two independent GWAS datasets: WHI (4382 samples and 742,996 SNPs) and ARIC (7756 samples and 780,062 SNPs). A total of 9,732 genes and 16 phenotypes shared in two datasets were used to construct causal genotype-phenotype networks. We identified 23 significantly connected to phenotypes in causal genotype-phenotype networks common in the WHI and ARIC studies. There were 13 edges directed from gene to phenotypes and 7 edges directed from phenotypes to phenotypes were common in the causal networks constructed by 23 genes and 16 phenotypes in the in the WHI and ARIC studies.

1424F

Genetic modifiers of immune reactivity to a human pathogenic mycobacterium. *J. Manry[1,2], Y. Nédélec[3,4], V. M. Fava[1,2], A. Cobat[5], M. Orlova[1,2], L. B. Barreiro[3,4,6], E. Schurr[1,2].* 1) Program in Infectious Diseases and Immunity in Global Health, the Research Institute of the McGill University Health Centre, Montreal, Quebec H4G 3J1, Canada; 2) McGill International TB Centre and Departments of Medicine and Human Genetics, McGill University, Montreal, Quebec H4G 3J1, Canada; 3) Sainte-Justine Hospital Research Centre, Montreal, Quebec H3T 1C5, Canada; 4) Department of Biochemistry, Faculty of Medicine, University of Montreal, Montreal, Quebec H3T 1J4, Canada; 5) Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale, U. 1163, Paris 75015, France; 6) Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, Quebec H3T 1J4, Canada.

Leprosy is a human infectious disease caused by *Mycobacterium leprae*. A strong host genetic contribution to leprosy susceptibility is well established. However, the modulation of the transcriptional response to infection and the mechanism of disease control are poorly understood. To address this gap in knowledge of leprosy pathogenicity, we conducted a genome-wide search for loci that are associated with transcript variation – expressed quantitative trait loci (eQTL) – before and after stimulation with *M. leprae* sonicate in whole blood cells from 54 unrelated leprosy patients from Vietnam. Among 12043 tested genes, we found 6787 as being significantly up- or down-regulated. Gene ontology (GO) enrichment analysis on differentially expressed genes (DEG) allowed us to confirm that the vast majority of significant enrichments were related to the immune response, including genes that had previously been associated with leprosy (e. g. *TNF* is in the top 15 of the most DEG). We then identified 36700 cis-eQTL for 818 genes, including 5784 for 33 genes associated with the “immune response” GO term. Interestingly, most of these “immune response” eQTL were eQTL for genes in the HLA region. In addition, we found 7814 eQTL as being specific to stimulated or to non-stimulated cells but not to both conditions. Of those, we found 336 response eQTL for genes associated with the “immune response” GO term including 265 for HLA genes. This finding is concordant with the results of our genome-wide association study, which had identified the HLA system as an important genetic risk factor of disease. The response eQTL correspond to regulatory variations that likely affect the interaction between human whole blood cells and *M. leprae* sonicate, and thus likely between the human host and *M. leprae* themselves. Remarkably, *ADCY3* – previously identified as a leprosy type-I reaction signature gene – was one of the genes presenting the highest number of response eQTL as well as one of the highest up-regulations of gene expression. Given their association with changes in the immune response to *M. leprae* sonicate, these response eQTL represent promising genetic candidates for susceptibility or resistance to leprosy and show the functional consequence of genetic risk factors for disease. Importantly, our study identified novel, hitherto unknown pathways of host responsiveness to *M. leprae* and substantially increases our understanding of leprosy pathogenicity.

1425W

A variable selection method for identifying complex genetic models associated with human traits. *E. R. Holzinger[1], J. Malley[2], Q. Li[1], J. E. Bailey-Wilson[1].* 1) National Human Genome Research Institute, NIH, Baltimore, MD; 2) Center for Information Technology, NIH, Bethesda, MD.

Purpose: Standard analysis methods for genome wide association studies (GWAS) are not robust to complex disease models (e. g. multivariable models with non-linear interaction effects), which likely contribute to the heritability of complex human traits. Machine learning methods, such as Random Forests (RF), are an alternative analysis approach that may be more optimal for identifying these effects. One caveat to RF is that there is no standardized method of selecting a set of variables with a low false positive rate (FPR) while retaining adequate power. **Methods:** We have developed a variable selection method called r2VIM. This method incorporates recurrency and variance estimation into RF to guide optimal threshold selection. We assess how this method performs in simulated SNP genotype data with a variety of complex effects (multiple loci with interactions and main effects). **Results:** Our findings indicate that the optimal selection threshold can identify interactions with adequate detection power while maintaining a low FPR in the selected variable set. For example, the optimal VIM threshold had an average detection power of 0. 80 and an average FPR of 0. 11 for a model with a two-locus interaction and no main effects (i. e. a purely epistatic model). However, the optimal threshold is highly dependent on the simulated genetic model, which is unknown in biological data. To address this, we permute the phenotype and re-run r2VIM to generate a null distribution of VIMs. The results from the permuted data are used to choose a selection threshold in the non-permuted analysis by comparing FPR estimates at different VIM thresholds, which does not require knowledge of the underlying genetic model. We tested the permutation method on an array of simulated data. Our initial results show that the best balance between FPR and detection power is produced by selecting the VIM threshold with an FPR of close to 0. 05 in the permuted data. Since our method is selection based (i. e. no modeling), we also implement a novel technique in r2VIM called “entanglement maps” to guide distinction of main effects versus interaction effects. We present visualizations of these results to further aid interpretation.

1426T

Bayesian Kernel Based Modeling and Selection of Genetic Pathways and Genes for Cancer. *Z. Wang[1], S. Chakraborty[2].* 1) BioStat Solutions, Frederick, MD; 2) University of Missouri, Columbia, MO.

Background: Much attention has been given recently to the development of analytical methods that utilize the large quantity of genetic information, which becomes available, thanks to advanced biological technologies. Most of the proposed methods look at the entire set of genes and their impact on a disease. Recently a new philosophy emerged which considers a combined genetic pathways effect on a disease. Under the new philosophy the objective is to identify the significant genetic pathways and the corresponding influential genes within a disease process. Method: A Bayesian kernel machine model which incorporates existing information on pathways and gene networks into the analysis of microarray data is developed to identify significant pathways and genes related to survival time. Each pathway is modeled nonparametrically using a reproducing kernel Hilbert space. Mixture priors on the pathway indicator variable and the gene indicator variable are assigned. An efficient Markov Chain Monte Carlo (MCMC) algorithm is developed to fit our model. Simulation studies and a real data analysis, using, van’t Veer et al. (2002) breast cancer microarray data, are used to illustrate the proposed method. Result: 9 out of 243 pathways are selected and 94 out of 4363 genes are selected by the model. Conclusion: This approach can be used to model both linear and non-linear pathway effects and can pinpoint the important pathways along with the active genes within each pathway. The model can use RNA sequencing data as input and it can be easily adapted to continuous or binary response variables.

1427F

Modified Random Forest Algorithm to Identify gene-gene Interaction in Case-Parent Trios Studies of Oral Cleft. *Q. Li[1], E. Holzinger[1], J. Hetmanski[2], M. Marazita[3], T. Beaty[2], J. Bailey-Wilson[1].* 1) Computational and Statistical Genomics Branch National Human Genome Research Institute National Institutes of Health 333 Cassell Drive, Suite 1200 Baltimore, MD 21224; 2) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; 3) Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh PA, USA.

Non-syndromic cleft lip with or without cleft palate (CL/P) is a common birth defect with strong genetic components. Genome-wide association studies (GWAS) for CL/P have identified multiple genes as influencing risk of CL/P. The case-parent trio design has the advantage of controlling for population stratification and facilitates error checking. However, methods designed to test for gene-gene interaction using trio data are very limited. In previous work, we have modified the random forest (RF) algorithm to identify gene-gene interaction in trio data. Our modified random forest uses a sampling scheme with a regression-based splitting criterion to take into account the family relationship among the cases and ‘pseudo controls’ (genotypic combinations possible given the parental mating type). In this project, to further explore possible gene-gene interaction effects, we applied our method to case-parent trio data from a previous GWAS study of CL/P with multiple ethnic/racial groups from an international consortium. To control for the randomness in the search algorithm, we employed the ‘recurrency variable importance metric’ to finalize the list of top SNPs possibly involved in gene-gene interactions. We have applied our method to a subset of ~400 markers, including those belonging to WNT pathway and those with large marginal effects. Our method identified markers with large marginal effects, (in genes *IRF6* and *C1orf107* in the Asian trios; and in *hcG_1814486* and *UNC5C* in the European trios), as the top SNPs possibly involved in gene-gene interactions. We did not find any evidence of gene-gene interaction within the WNT pathway. We are currently applying this method to the entire genome-wide marker panel using these GWAS data, since this method has the potential to identify possible gene-gene interactions among markers without large marginal effects. The results will be compared with the results from another ensemble method, trioFS, which is part of the R package, trio.

1428W

Causal Inference for Integrative Analysis of Genetic and micro RNA variation. *P. Wang[1,4], X. Chen[2], V. Vladimirov[2], Y. Zhu[3], M. Xiong[1].* 1) University of Texas, School of Public Health, Houston, TX; 2) Virginia Commonwealth University, Richmond, VA; 3) Tulane Univrsity, New Orleans, LA; 4) Fudan University, School of Life Sciences, Shanghai, China.

The rapid development of genomics technologies including microarrays and next generation sequencing (NGS) has provided powerful tools for comprehensive analysis of genomes, including the expression of microRNAs, and sequence variations. These genome-wide profiles of the genetic aberrations could reveal the biomarkers for diagnosis and prognosis of diseases as well as mechanistic insights into disease development and progression. Recent efforts to establish the huge psychiatric genome compendium and integrative omics analyses, so-called "integromics", have extended our understanding on the genome, showing its daunting complexity and heterogeneity. However, the challenges of the structured integration, sharing, and interpretation of the big omics data still remain to be resolved. We often use correlations or mutual information between two genes and Gaussian undirected graphical models to construct co-expression networks as well as regression to identify eQTLs. Although association analysis and regression quite useful in genetic and epigenetic studies of mental disorders, they are difficult to unravel the mechanic mechanism of the complex diseases. To overcome these limitations, gene perturbations that exploit naturally occurring genetic variations can be used to learn causal genotype-micro RNA networks. We propose exact learning causal graphical models for genotype-micro RNA networks from both miRNA-seq and NGS genetic data. This will open a new avenue to use exact learning DAG for causal inference of genotype-micro RNA networks, which will lead to accurate estimation of causal effects of the genetic variants on micro RNA expressions and provide a powerful tool for drug target discovery and precision medicine. The proposed methods were applied to micro RNA and SNPs datasets from the Stanley Medical Research Institute (SMRI) of 58 samples. We first apply our method to each gene and 422 miRNAs and 666 genes passed the p-value threshold 3. 25e-6. Then we fit our model to the 422 miRNAs and the 666 selected genes and get inferred sparse miRNA causal network consisted of 809 nodes and 1391edges, among which 339 genes were significantly connected to micro RNA.

1429T

Comparison and optimization of different centrality measures algorithms used in human gene network analysis. *T. Zhang, J. P. Rice.* Dept of Psychiatry, Washington University in Saint Louis, Saint Louis, MO.

Four centrality measures are one of the important parameters utilized in gene network analysis: 1) Degree centrality which simply counts the number of interactions to a node; 2) Eigenvector centrality which ranks the nodes in a network based on its integrating neighbors; 3) Betweenness centrality where nodes which fall in the shortest path of other nodes have high betweenness, and 4) Closeness centrality which is related to the topology of the nodes in a network. Despite some early studies that utilize the simplest degree centrality measure, most of the recent research projects have implemented the eigenvector centrality measure, such as algorithms modified from Google PageRank algorithm, to evaluate the importance of a gene in a gene network. However, research questions such as to what extent these centrality measures can represent the functional significance of genes or whether any of these centrality measures outperformed others have never been seriously investigated. In this project, we aimed to conduct a study to compare the accuracy and efficiency of these four centrality measures. Two essential gene sets were obtained and served as "true answers" for algorithms comparison and machine learning data training. One gene set containing 1,528 protein coding genes was directly extracted from Online Gene Essentiality Database (OEGG). Another gene set including 1,174 genes was obtained through implementing a variant filtering strategy based on theExome Aggregation Consortium data (ExAc). A comprehensive human gene network which covered 90% of human protein coding genes were constructed based on protein-protein interaction data extracted from STRING, REACTOME and iRefindex. The network visualization and centrality measure calculations were performed using R package igraph. ROC curve was made to make comparison of the four algorithms. In addition, we constructed a logistic regression model to combine the four centrality measures in order to give a more accurate prediction of the functional significance of protein coding genes in the comprehensive gene network. The 10-fold cross-validation approach was used to assess the performance of the prediction model. Our results indicated that the combined centrality measure is more efficient in predicting the functional significance of a gene in a specific human gene network.

1430F

Differential expression of transcript isoforms in schizophrenia. *E. Drigalenko[1], W. Moy[2], J. Duan[2,3], H. Göring[4], P. Gejman[2,3], A. Sanders[2,3], Molecular Genetics of Schizophrenia (MGS) Collaboration.* 1) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 2) Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem, Evanston, IL; 3) Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL; 4) South Texas Diabetes and Obesity Institute, University of Texas Health Science Center, San Antonio, TX.

Transcriptome sequencing (RNA-seq) enables investigation of differential expression of transcript isoforms and genes from subjects with or without illness. Transcript isoform expression level is arguably more biologically meaningful than overall gene expression level since the former often encode different proteins, which can have different biological function. Sensitivity may be improved by avoiding the summing over transcript isoforms needed for aggregate (gene level) analysis. We profiled the transcriptome (>9M total single-end RNA-seq reads per subject) in lymphoblastoid cell lines (LCLs) of 1,278 European ancestry subjects (550 cases and 728 controls) from the MGS collection. We used programs Tophat v2. 0. 12 and Cufflinks2 v2. 2. 1 for read alignment, transcript assembly, and quantification (FPKM) of each transcript isoform and gene. We restricted our differential expression analyses to the 108,346 transcripts and 21,215 genes from Gencode v. 20 annotation that showed detectable expression (FPKM>0) in at least 80% (1,022) of subjects. We used square root as the variance stabilizing transformation for the expression values, and then adjusted for covariates: sex, age, age2, ancestry (five genotypic principal components), LCL characteristics (viral load, growth rate, energy status), and RNA-seq batch. After inverse normalization of residuals, we evaluated the association of expression levels with schizophrenia status by multiple linear regression analysis (R v3. 02 program package). We found 1,265 transcripts and 1,252 genes to be differentially expressed by affection status (Bonferroni corrected p<0. 05). Of these transcripts, 1,015 were in genes also showing significant differential expression in the gene-level analysis, while 250 were in genes not showing significant differential expression. On the other hand, 447 differentially expressed genes (gene-level analysis) did not show differential expression at the transcript isoform level. Differential transcriptomic analyses at the gene and transcript isoform levels may complement each other, providing additional information for identifying disease-associated aberrant expression. We found a number of significant transcripts and genes in HGNC (HUGO Gene Nomenclature Committee) gene families, such as histocompatibility complex, histones, interferons, and interleukins. Many of the detected gene families are immune-related ones, consistent with previous large GWAS and expression studies.

1431W

A HMM-based method to harmonize discordant genotypes obtained from DNA sequencing and genotyping experiments in the same samples. *C. Low-Kam[1,2], D. Rhainds[1], I. Mongrain[1,3], M. -P. Dubé[1,2,3], J. -C. Tardif[1,2], G. Lettre[1,2].* 1) Montreal Heart Institute, Montréal, QC, Canada; 2) Faculty of Medecine, Université de Montréal, Montréal, QC, Canada; 3) Beaulieu-Saucier Université de Montréal Pharmacogenomics Center, Montréal, QC, Canada.

We introduce a new method to merge genotypes from two data sources at divergent locations. Our approach is motivated by a low-pass whole-genome sequencing (WGS) study of the French-Canadian population, where 2002 individuals were sequenced at 5. 7x coverage. In this study, we identified over 32 million variants, of which 51% were not previously reported in public databases. We evaluated our data by comparing the resulting sequences against ExomeChip genotyping data available in the same individuals. The overall genotyping concordance rates between both datasets are 99% and 87% when considering all genotypes or only heterozygous genotypes, respectively. These rates are encouraging because low-pass WGS is theoretically less accurate for heterozygous genotypes, and for rare variants, which are enriched on the ExomeChip. To the best of our knowledge, no method currently exists for merging our datasets at discordant positions. As the confidence is greater for the ExomeChip genotyping data, a reasonable approach would be to choose genotypes from the ExomeChip as the “true” genotypes. A more adaptive alternative consists in using linkage disequilibrium to derive the most likely genotype at each divergent location. We developed a bivariate Hidden Markov Model (HMM)-based framework that extends the imputation method of the software MaCH to output posterior genotype probabilities based on both datasets. Conditional independence properties allow accounting for neighbouring sites variation while remaining computationally tractable. The associated algorithm is implemented in C++. Preliminary results are promising: we tested our method on two datasets with distinct errors derived from the same simulated genotypes. Our method provides a compromise that outperforms the dataset with the lowest error rate. Although the algorithm works best to detect errors in homozygotes for the major allele, it is still able to correct errors in heterozygotes and homozygotes for the minor allele. Finally, the algorithm is robust to parameter initialization. Our methodology can be applied to any datasets where divergent genotypes are collected from two sources, on the same samples.

1432T

Imputing LoF variants in Finnish founder population using population-specific reference panels. *A. Sarin[1,2], I. Surakka[1], K. Karczewski[3], R. Durbin[4], D. MacArthur[3], V. Salomaa[2], A. Palotie[1,3], S. Ripatti[1,5] for SISu project.* 1) Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland; 2) National Institute for Health and Welfare, Helsinki, Finland; 3) Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 4) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK; 5) Department of Public Health, University of Helsinki, Helsinki, Finland.

It has been shown that population specific reference panel improves imputation accuracy especially for the rare and low frequency variants. In our study, we utilized low-coverage whole genome sequencing (WGS) and whole exome sequencing (WES) based reference panels created from the Sequencing initiative Suomi (SISu; http://www. sisuproject. fi) data, and compared imputation accuracy of single nucleotide polymorphisms (SNPs) with MAF < 1% in a set of imputation reference panels that also included the publicly available 1000 Genomes panel. Our results show that the highest imputation accuracy in Finnish samples can be obtained by extending the WGS panel (3882 haplotypes) with the WES panel (3080 haplotypes). By using this combination in the imputation of Finnish Illumina HumanCoreExome data (N=10,489), we observed median squared correlation (r2) of > 0. 8 between true and imputed genotypes within 0. 1-0. 5% allele frequency range. Because of a genetic bottleneck in Finnish population approximately 100 generations ago, there is an enrichment of lost-of-function variants (LoFs) in 0. 5-5% allele frequency range compared to non-Finnish Europeans. We imputed 422 LoFs enriched in Finns and observed that 85% of them can be imputed with good quality (info > 0. 7) using a population based reference panel. A concordance analysis performed for a subset of the data (median MAF=1. 9%) showed that 93% of heterozygotes and 95% of the homozygotes for the minor allele could be captured by imputation while introducing 1. 4% of heterozygote and 0. 6% of homozygote false positives. Our results suggest that population specific, sequencing based reference panel allows better imputation of rare and low frequency variants and makes it possible to utilize the existing genome-wide association study (GWAS) chip data in fine-mapping and association screens of complex diseases and traits more efficiently. Because GWAS chip data is available for 60,000 Finns, there is a potential for novel discoveries. Currently, this is also a cost effective option while the field is moving towards large-scale sequencing studies.

1433F

Chinese population allele frequency estimations based on large-scale non-invasive prenatal testing samples. *H. Xu, F. Chen, X. Jin, Y. Zhang, H. Jiang, X. Xu.* BGI-Shenzhen, Shenzhen, China.

Allele frequency estimations in Chinese people were an important factor for the genetic map in Chinese population and other epidemiology studies including molecular prevalence of genetic diseases. Chinese Han populations in the 1000 Genomes Project released in 2012 have been the most widely-used database for variants especially SNPs in Chinese, providing an intact and precise map of genome variation in Chinese people and accelerating tons of Chinese population studies. With 90 Han people in China sequenced, the sampling proportion was quite small compared with billions of Chinese people and the estimated allele frequencies may be deviated from those of large-scale Chinese population due to sampling randomness. It was estimated that over 1,000,000 pregnant women took sequencing-based non-invasive prenatal testing for fetal aneuploides screening in China from January 2011 to June 2015. Low-coverage (~0. 1X) WGS strategy was mostly used in clinical labs in China, which presented a large-scale and randomly sampled population and constitutes significant genetic databases for Chinese populations with informative phenotypes including territory distributions, maternal age, nationality and regions. Till now, most population-based algorithms for allele frequencies computation were developed and validated in 30X WGS data but were not appropriate for low-coverage sequencing data. It was hard to discover population knowledge from such big but ultra-low coverage data which request specific models to deal with population SNP calling, demanding computational tools and mass storage. Here in this study we developed a maximum-likelihood method to estimate allele frequency in Chinese population and applied it to NIPT data in over 150,000 samples. A Chinese genetic map of over 150,000 Chinese people was built and the allele frequencies in the whole genome in the large-scale Chinese people were studied. We also analyzed prevalence of common single-gene disorders such as thalassemia, DMD, SMA and hearing loss in different regions in China from 2011 to 2015. Our findings were compatible with current epidemiology reports in Chinese populations and showed the pictures of molecular prevalence of genetic diseases in China. It was the greatest population studied with millions of orders of magnitude to our knowledge. Our studies improved the understanding of variants in Chinese populations, promoting more potential uses of NIPT samples in population genetics.

1434W

Joint whole-genome analysis of associations between host and hepatitis C virus diversity in a patient cohort. *V. Pedergnana[1], MA. Ansari[2,3], STOP-HCV Consortium[4], E. Barnes[3], P. Klenerman[3], C. Spencer[1].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford. OX3 7BN; 2) Oxford Martin School, University of Oxford, 34 Broad Street, Oxford OX1 3BD; 3) Nuffield Department of Medicine, University of Oxford, Oxford OX1 3SY; 4).

Joint analysis of both human and viral genomes in the context of infection can inform on underlying molecular interactions and can be used to help patients and doctors make better decisions about treatment. The hepatitis C virus has a highly diverse RNA genome that is difficult to sequence. As part of the STOP-HCV consortium (www. stop-hcv. ox. ac. uk) we used a new, efficient, high-throughput method for sequence-specific enrichment and characterization of whole virus genomes capable of unbiased detection of virus variants to sequence the virus of 560 chronically infected patients. In parallel, we also genotyped over 800,000 SNPs in the patients' genome, using the Affymetrix UK Biobank array, and used statistical imputation to obtain data at over 10 million SNPs as well as classical HLA alleles. Perhaps for the first time, paired human genotyping data and novel whole-genome viral sequencing data allow a systematic study of the relationship between hepatitis C virus diversity and host genomics. We develop and apply a new statistical method to look for evidence that host genetics is associated with viral sequence motifs (so-called footprints) in patients infected with hepatitis C. Using an analysis adjusted for viral and human population structure, we inferred footprints of host immune pressures on the pathogen, including a strong signal of association at a previous described HLA-A footprint in the NS3 viral protein ($P < 10\text{-}13$). Using a Bayesian approach we combine evidence across multiple footprints (in the HLA, but also at other clinically relevant host genes and genome-wide) on the viral genome and relate the observations to gene function. Together these new approaches allow a joint assessment of clinically relevant genetic markers in both patient and pathogen, information that is going to be potentially readily available for clinicians in the new era of stratified medicine.

1435T

The KOPS Registry: A unique cohort for the study of the genetics of obesity and weight loss. *A. G. Comuzzie[1,2], R. Bastarrachea[1,2], J. Kent[1,2], M. Olivier[1,2].* 1) Genetics, Texas Biomedical Research Institute, San Antonio, TX; 2) TOPS Nutrition and Obesity Research Center, Texas Biomedical Research Institute, San Antonio, TX.

We have recently entered into a strategic alliance with the TOPS (**T**ake **O**ff **P**ounds **S**ensibly) organization to establish the TOPS Nutrition and Obesity Research Center in the Department of Genetics at Texas Biomedical Research Institute. TOPS was founded by Esther Manz in 1948 in Milwaukee, WI as a community-based support group for individuals seeking to loose weight. From an original membership of 3 (Mrs. Manz and two friends), TOPS has grown to over 180,000 active members with chapters all over the world, and has helped millions of people meet their weight goals over its 67-year history. A subset of the TOPS membership includes KOPS (**K**eep **O**ff **P**ounds **S**ensibly) members. These individuals have achieved their weight loss goals and have maintained this weight for more than a year. Currently there are 19,034 KOPS members in North America (320 in Texas, and 76 in the San Antonio metropolitan area) who have maintained their weight loss from 1 to 44 years. As part of our new alliance with the TOPS organization, not only will we have access to the general membership, but will be the repository for a KOPS database. As part of this initial effort, we are working with the San Antonio metropolitan area TOPS chapters to initiate data collection for the genomic characterization of the KOPS cohort. This unique resource of individuals who have not only achieved, but maintained significant weight loss, combined with similar data from those individuals who continue to struggle with weight loss and maintenance, will allow for the identification of the genetic contribution to weight loss success. For this study, we will begin by recruiting individuals that are members of TOPS chapters in the San Antonio region to identify both obese and formerly obese members. Given the ready abundance of highly palatable, and calorically dense foods, those individuals who can successfully loose excess weight, as well as maintain this loss in the presence of our modern obesogenic environment represent a unique population. For this initial phase of our study we will administer a survey form to capture eating behaviors and dietary preferences, along with other basic demographic parameters, and initial biological samples for DNA characterization. We will present demographics of our regional clubs and KOPS members, and preliminary results from our surveys to demonstrate the unique nature of this population-based resource.

1436F

A population-specific reference panel empowers genetic studies of Anabaptist participants through improved imputation and variant filtering. *L. Hou[1], R. L. Kember[2], J. C. Roach[3], J. R. O'Connell[4], D. W. Craig[5], M. Bucan[2], A. R. Shuldiner[4], F. J. McMahon[1].* 1) Human Genetics Branch, National Institute of Mental Health Intramural Research Program, Bethesda, MD; 2) University of Pennsylvania, Philadelphia, PA; 3) Institute for Systems Biology, Seattle, WA; 4) University of Maryland School of Medicine, Baltimore, MD; 5) Translational Genomics Research Institute, Phoenix AZ.

Genetic isolates are advantageous for genetic studies of complex phenotypes, owing to decreased genetic heterogeneity and enrichment of otherwise rare alleles. However, population-specific variants and linkage disequilibrium patterns may complicate association studies that rely on imputation at ungenotyped sites. Genetic drift may also confound attempts to filter variants against allele frequencies reported in general reference samples. Here we report the first population-specific reference panel drawn from people of Anabaptist ancestry and perform a preliminary assessment of the panel's value for sequencing studies. High-depth (>30x) whole-genome sequencing (WGS) data were generated on 265 individuals of Amish or Mennonite ancestry. We identified >12M single nucleotide variants (SNVs) and short indels, of which ~19% were novel. Genomes were phased with SHAPEIT2 and 120 unrelated individuals were extracted. To evaluate imputation accuracy, 80 individuals were selected at random as a reference panel while the remaining 40 were set aside as a test panel. In the test panel, genotypes on chromosome 1 were masked at all variable sites not represented on the Illumina Human OmniExpress SNP array. Masked sites were then imputed by IMPUTE2 against: 1) the Anabaptist reference panel, and 2) 100 CEU individuals from the latest 1000G reference panel. Imputation accuracy was measured as the squared correlation coefficient (r2) between imputed allele dosages and masked genotypes. Higher imputation accuracy was achieved with the Anabaptist reference panel, especially for SNVs with minor allele frequencies (MAF) <10%. The greatest advantage was observed for less common alleles. For example, alleles with MAF <1% were imputed at a mean r2of 0. 82 with the Anabaptist panel, vs. 0. 73 with the 1000G panel. At r2 > 0. 5, only 10% of SNVs with MAF <1% could not be accurately imputed with the Anabaptist reference panel, compared to 16% with the 1000G panel. To assess the value of the Anabaptist reference panel for filtering out common alleles, we performed exome-sequencing on 57 additional Anabaptist individuals and identified 1,321 functional coding variants that were uncommon (MAF<1%) in the 1000G and EVS panels. Of these ~28% were not uncommon in the Anabaptist reference panel, and ~2% were enriched >4-fold. In studies of population isolates like the Anabaptists, a population-specific reference panel can yield considerable benefits in rare variant imputation and filtering.

1437W

Measuring the value of GWAS results in a clinical trial setting. *P. Patil, J. Alquicira, J. Leek.* Department of Biostatistics, Johns Hopkins University, Baltimore, MD.

The catalog of published genome-wide association studies (https://www. genome. gov/26525384) represents over a decade of analysis and millions of dollars in research expenditures. One concern raised by both the research community and the popular press has been the relatively modest effect sizes detected from these studies. In this work, we undertake a repurposing of the GWAS catalog and apply a robust method from the analysis of clinical trials to assess the clinical value of the reported SNPs and associated odds ratios. We describe some of the nuances in the records for each SNP, including inconsistencies in reporting of odds ratios, sample sizes, and risk allele frequencies in the GWAS catalog. We present both manual and automatic cleaning and curation solutions that we employed to allow for a large-scale analysis of the catalog. We also present baseline covariate adjustment for treatment effect estimation, and how treating a SNP as a predictor of a phenotype can lead to estimator precision gain in a clinical trial setting. We report which SNPs identified through GWAS may offer a benefit if measured prior to a clinical trial for the phenotype of interest. Although the precision gains we see in most cases are small, these results represent a realistic usage of the information in the GWAS catalog toward true clinical value.

1438T

Equivalence tests for the analysis of genotyping data: Assessing equality of SNPs in study cohorts. *Z. Talebizadeh, H. Dai, A. Shah.* Children's Mercy Hospital and University of Missouri-Kansas City School of Medicine, Kansas City, MO.

Statistical testing is strictly based on null (Ho) and alternative (Ha) hypotheses. The construction of statistical hypotheses will determine the interpretation of results. In most analyses of genotyping data or genome-wide association studies, the null hypotheses assume there is no association between SNP and phenotype. Rejection of null hypotheses will provide strong evidence to indicate the potential associations between SNP and phenotype. The aim of our study is to demonstrate the application of equivalence tests on genotyping data in situations that require testing the presence of equality instead of differences. A series of SNP data will be generated and tested for a hypothetical scenario; i. e. , rule out the impact of tested SNPs on a given comorbidity in a disease group, by demonstrating equality between corresponding means. The constructed Ho is: there is association between SNP and comorbidity in the patient population versus the Ha: there is no association between SNP and comorbidity in the patient population. An equivalence test is warranted to test the constructed hypotheses. Equivalence test has been widely applied in clinical trials to confirm the equivalency in drug efficacy (i. e. , bioequivalency), but rarely on genotype data. In this work we will discuss: 1) differences between equivalence test and differential test, 2) misuse of differential test for equivalence testing in the context of analyzing genotyping data, 3) minimal sample size to establish an equivalence limit. We will perform the empirical Monte-Carlo simulation to assess the Type I error and power for three equivalence tests, and compare them with commonly used differential tests, including Fisher's exact test, Chi-square test and Cochran-Mantel-Haenszel trend test. The impact of variables such as: allele frequency, sample size, and the number of tested SNPs on equivalence of cases and controls will also be assessed. We will perform extensive simulation study to illustrate that misuse of differential tests may cause bias in stating equality of SNPs between cohorts. For genotyping data and genome-wide association analysis, a limited number of equivalence tests are available as compared to very rich pool of differential tests. Our study will contribute to addressing this gap by providing a useful protocol, including examples, for application of such tests on genotyping data. More equivalence tests need to be developed to fulfill the needs of genotype testing analysis.

1439F

Integrating genome-wide meta-analyses of binary and continuous phenotypes. *R. K. Walters[1,2,3], PGC ADHD Workgroup, iPSYCH-SSI-Broad/MGH ADHD Workgroup, EAGLE Consortium.* 1) Analytic & Translational Genetics Unit, Massachusetts General Hospital, Boston, MA; 2) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA; 3) Department of Medicine, Harvard Medical School, Boston, MA.

Complex disorders frequently have corresponding measures of continuous population variation in the phenotype that exist alongside the clinical diagnosis. By capturing sub-clinical variation, studying these continuous phenotypes can offer increased power to detect novel genetic variants linked to the disorder (e. g. Dupuis et al. , 2010 Nat Genet.). On the other hand, dichotomous case/control status may be easer to ascertain from e. g. diagnostic codes in electronic medical records, facilitating rapid sample collection and increasing power through larger sample sizes (e. g. Ripke et al. , 2013 Nat Genet.). Despite the close relationship between the phenotypes, however, it is not possible to directly combine results from these two study designs in meta-analysis due to the differing measurement scales. Therefore, we introduce a framework for jointly meta-analyzing GWAS of both dichotomous disease status and corresponding continuous measures of population variation. First, the GWAS results for the dichotomous phenotype are transformed to the scale of a continuous latent trait under the liability threshold model with appropriate adjustment for case/control ascertainment. Next, the relationship between this latent liability and the continuous phenotype is modeled based on the genetic covariance between the two traits and their respective heritabilities, which can be calibrated based on LD score regression (Bulik-Sullivan et al. , 2015 bioRxiv). Modeling the genetic covariance ensures that the results from different phenotypes are weighted appropriately even when the continuous phenotype is not directly a measure of liability for the dichotomous phenotype. The current work presents thorough simulation studies quantifying the expected increase in power from this joint meta-analysis and the influence of varying genetic architectures contributing to the genetic covariance. We then demonstrate the potential for this methodological approach to improve power for discovery of associated SNPs by integrating GWAS of (continuous) measures of symptoms of Attention deficit/hyperactivity disorder (ADHD; MIM 143465) from the Early Genetics & Lifecourse Epidemiology Consortium (EAGLE) with the latest GWAS of ADHD diagnosis from the Psychiatric Genomics Consortium (PGC).

1440W

Assessing models for genetic prediction of complex traits: visualization and quantitative methods. *S. A. Gagliano[1,2,3], A. P. Paterson[2,3,4,5,6], M. E. Weale[7], J. Knight[1,2,3,5].* 1) Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Ontario, Canada; 2) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada; 3) Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada; 4) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 5) Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 6) Epidemiology Division, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada; 7) Department of Medical & Molecular Genetics, King's College London, Guy's Hospital, London, UK.

Within the past couple of years several *in silico* models have been developed with the intention of predicting which genetic variants are more likely to contribute to the risk of a complex trait given their functional genomic characteristics. Several predictive accuracy measures have been used to assess the performance of these models, but there had been no comprehensive review as to which type of predictive accuracy measures and data visualization techniques are most useful in the context of genetic risk variant prioritization. We undertook such a review, comparing various measures, some of which include: receiver operating characteristic (ROC) curves, Mann-Whitney U test, histograms of classification probability, and the novel use of the quantile-quantile plot (PMID: 25997848). These measures have variable interpretability depending on factors such as whether the dataset is balanced in terms of numbers of genetic variants classified as risk variants versus those that are not. We applied these measures to six models from Gagliano et al. (PMID: 24844982). We found that p-values (from the Mann-Whitney U, the hypergeometric and the generalized Cochran-Mantel-Haenszel tests) were uninformative in determining the most accurate models because statistical tests do not explicitly measure class separation between the risk and non-risk variants. We conclude that the area under the curve (AUC) is a suitable starting place, and for models with similar AUCs, violin plots are particularly useful for examining the distribution of the risk probabilities. This investigation emphasizes the importance of: (1) visualizing the distributions of the probabilities, and (2) exercising caution when making conclusions about model performance based on p-values.

1441T

Measuring the rate and heritability of aging in Sardinians using pattern recognition. *D. Schlessinger[1], E. D. Sun[1], Y. Qian[1], G. R. Abecasis[2], F. Cucca[3], J. Ding[1], I. G. Goldberg[1].* 1) National Institute on Aging, Baltimore, MD. 251 Bayview Boulevard, Suite 100 Baltimore, MD 21224; 2) Department of Biostatistics and Center for Statistical Genetics, University of Michigan Ann Arbor, MI; 3) Institute of Genetic and Biomedical Research, National Research Council, Monserrato, Cagliari, Italy.

It is widely accepted that individuals age at different rates. A method that measures physiological age independently of chronological age could therefore be a first step to reveal contributing mechanisms; but searches for individual biochemical markers of physiological age have had limited success. In this study, we rather assessed the extent to which an individual's chronological age could be determined as a composite score inferred from a broad range of biochemical and physiological data. Data were collected in a longitudinal population study in Sardinia. The study (the "SardiNIA" project at https://sardinia. irp. nia. nih. gov/) includes measures of environmental factors and family structures to facilitate both epidemiological and genetic analyses. We used pattern recognition and machine learning strategies on data for the 6,000 participants in the study, who range in age from 12 to 81. The best predictive models were determined from multiple combinations of dimensionality reduction, classification, and regression algorithms. They reached very strong correlations (R > 0. 9) between predicted and actual ages, and showed relative stability in successive visits of the same individuals (R>0. 5). We then defined an Effective Rate of Aging (ERA) for each participant, a continuous trait measured as the ratio of an individual's predicted value to his/her chronological age. The inference that individuals have a characteristic rate of aging is supported by findings that in the entire cohort, the inferred values of ERA showed genetic heritability of 40%. This has been sufficient to initiate genome-wide association studies that identify the effect of variants in genes affecting telomere length and metabolic activity on the rate of aging.

1442F

The Kaiser Permanente/UCSF Genetic Epidemiology Research Study on Adult Health and Aging: Imputing rare variants in ~100,000 individuals. *T. J. Hoffmann[1], Y. Banda[1], M. N. Kvale[1], L. Shen[2], E. Jorgenson[2], C. Schaefer[2], N. Risch[1].* 1) University of California San Francisco, San Francisco, CA; 2) Division of Research, Kaiser Permanente, Oakland, CA.

Rare genetic variants may be responsible for a significant amount of the uncharacterized genetic risk underlying many diseases. Since sequencing is still much more expensive than microarrays, an efficient approach to testing rare variants may be to impute them into very large existing genome-wide studies that have been genotyped via microarrays. Although many success stories of individual rare variants have emerged, such as the variant with MAF 0. 38% associated with sick sinus syndrom (Holm et al, 2011), or the HOXB13 G84E variant in this cohort (MAF 0. 17%) (Hoffmann et al, 2015), to name just a couple, it is still unclear how well imputation will work in general for rare variants. However, increasing reference panel size has made rare variant imputation more feasible. We explore here the effect of several different reference panels on imputation success using ~100,000 individuals genotyped at over 650,000 SNPs in the Kaiser Permanente Research Program on Genes, Environment and Health (RPGEH) Genetic Epidemiology Research on Aging (GERA) cohort. We find, for example, that when comparing the reference panel of 1,092 individuals from 1000 Genomes to the more recent release with 2,504 individuals, that there is little change in the ability to impute variants with frequency>1%, but we recover approximately twice the variants with frequency<1%. In addition, we also find that an extremely high proportion of variants that were removed for QC reasons were recovered by imputation, which may be a useful strategy to control for batch effects.

1443W

Heritability and genetic correlations among hormone and cytokine levels related to type 2 diabetes. *R. L. Hanson[1], E. J. Weil[1], J. M. Curtis[1], M. Walter[2], R. G. Nelson[1].* 1) Diabetes Epidemiology Clin Res, NIDDK, Phoenix, AZ; 2) Clinical Laboratory Core, NIDDK, Bethesda, MD.

Several hormones and cytokines have been described as related to type 2 diabetes and obesity, but there is little information on the extent to which they are influenced by genetic factors or on their genetic correlations with conventional metabolic variables. We used classical pedigree-based variance components methods to calculate heritability (h2) for hormone levels in 1088 nondiabetic American Indians in a family study. All individuals were examined after an 8-hr fast. Ten "extended" hormones/cytokines were measured: adiponectin, adipsin, amylin, C-reactive protein (CRP), glucose-dependent insulinotropic peptide (GIP), interleukin-6, leptin, plasminogen activator inhibitor-1, resistin and tumor necrosis factor α (TNFα). We also analyzed 8 "core" metabolic variables: body mass index (BMI), fasting glucose (FPG), hemoglobin A1c, fasting insulin, HDL cholesterol, triglycerides, systolic and diastolic blood pressure (SBP, DBP). Significant (P<0. 05) heritability, adjusting for age, sex and ancestry, was observed for all "extended" variables with h2 ranging from 36% for TNFα to 86% for resistin. All "core" variables, except DBP (h2=0) were significantly heritable, and h2 ranged from 27% for fasting glucose to 57% for BMI. We conducted bivariate analyses of the covariance among relatives to partition the phenotypic correlation (rP) between "extended" and "core" variables (except DBP) into genetic and environmental sources (rG,rE). Of the 70 variable pairs 51 had significant phenotypic correlations with |rP| ranging from 0. 06-0. 59. Among these 51 variable pairs, |rG| ranged from 0. 01-0. 71 and |rE| from 0. 02-0. 60; on average 40% of the correlation was due to shared genetic factors. The strongest rG were: 0. 71 for leptin-BMI, 0. 66 for insulin-leptin, 0. 60 for CRP-BMI and -0. 59 for GIP-glucose; the strongest rE were 0. 60 for resistin-BMI and -0. 55 for resistin-HDL. These analyses suggest that many hormones/cytokines related to type 2 diabetes are substantially influenced by genetic factors and that many of these genetic factors partially overlap with those that influence conventional metabolic variables. Further molecular genetic studies of these hormones may help clarify the etiology of type 2 diabetes and obesity.

1444T

A simple yet accurate correction for winner's curse can predict signals discovered only in larger genome scans. *S-A. Bacanu, T. B. Bigdeli, D. Lee, B. P. Riley, V. Vladimirov, F. H. Ayman, K. S. Kendler.* VIPBG, Virginia Commonwealth University, Richmond, VA.

Genome-wide/sequencing scans, henceforth denoted as genome scans, discovered/will discover numerous genetic variants significantly associated with various phenotypes. However, while significant signals explain only a small fraction of the variation for most traits, variants with smaller signals explain a much larger part. To extract the signals of all magnitudes from a genome scan, we propose a very simple method for accurately estimating the mean of all (normally distributed) genome scan statistics (Z-scores). Given that adjusting for winner's curse (multiple testing) is difficult for all normally (χ^2) distributed statistics but simple for their p-values, we propose a two-step procedure which we denote as **F**DR **I**nverse **Q**uantile Transformation (FIQT). First, adjust the p-values for multiple testing using a False Discovery Rate (FDR) approach. Second, estimate the mean of Z-scores as having i) magnitude obtained by back-transforming the adjusted p-values to a (upper tail) standard normal quantile and ii) sign of the original statistics. When compared to competitors, e. g. Empirical Bayes and tail adjustment methods, FIQT has very good performance in terms of (i) squared error loss, (ii) fraction of the explained variability in true Z-score means and iii) computational efficiency. We show that FIQT applied to the first Psychiatric Consortium (PGC) schizophrenia cohort can be used to predict with reasonable accuracy a non-trivial fraction of the 108 regions harboring signals in the second PGC schizophrenia cohort. Finally, we suggest further improvements to FIQT by taking into account the noncentrality/mean of χ^2 statistics for i) the entire genome scan or ii) functionally relevant subsets.

1445F

Study the effects of multiple mediators using DNA methylation. *J. Shen, X. Lin.* Department of Biostatistics, Harvard TH Chan School of Public Health, Boston, MA.

With development of genetic and genomic technologies, it is of increasing interest in biomedical and epidemiological studies to explore the biological mechanism on how an exposure variable affects behavioral or disease outcome. Mediation analysis provides a powerful tool to identify the intermediate genes and biological pathways that mediates the disease causing process. Epigenetic studies for complex diseases have shown that the effect of exposure variables are more likely to be mediated in a collaborative fashion through multiple probes in a region, e. g. , CpG islands or shores, or methylation sites from several genes, e. g. , a genetic pathway or network. Rather than examining a single mediator, it is more desirable to investigate the mediation effect of multiple methylation sites simultaneously, while taking into account correlation among these mediators. We propose a test for the overall indirect natural effect for a group of mediators using the difference in coefficients approach. We employ kernel machine based approach to account for potential interactions and nonlinearity among mediators. A Wald type test is developed with a simple estimator for the variance when comparing coefficients from different models. The performance of the method is evaluated with simulation studies under various correlation structures of the mediators. The proposed test is then applied to the Normal Aging Study to investigate the effect of smoking on coronary heart disease that mediated by the methylation level on genome-wide CpG sites.

1446W

Genetic susceptibility to higher insulin secretion is associated with obesity. *J. N. Todd[1,2,4], C. M. Astley[1,4], J. N. Hirschhorn[1,3,4], J. C. Florez[2,3,4].* 1) Medicine-Endocrinology, Boston Children's Hospital, Boston, MA; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA; 3) Broad Institute, Cambridge, MA; 4) Harvard Medical School, Boston, MA.

High glycemic index diets have been associated with obesity in population studies. It has been theorized that carbohydrate-driven insulin secretion mediates weight gain by driving hunger and caloric intake. In fact, insulin secretion measures (*e. g.* insulin levels 30 minutes after an oral glucose load, first phase insulin) corresponded with weight loss on a low-glycemic diet and future weight gain in observational studies. To better understand the causal relationship between insulin secretion and obesity, we undertook a bi-directional Mendelian randomization analysis using polygenic instruments of both traits. In Mendelian randomization, genetic predictors of an exposure of interest (*e. g.* insulin secretion) are used to identify the association between that exposure and the outcome (*e. g.* obesity) not confounded by shared causes (*i. e.* unmeasured confounders). This same analysis is undertaken using genetic predictors of obesity to assess for reverse causation on insulin secretion. In this way, bidirectional Mendelian randomization can help tease out both the presence and direction of causal effect that contribute to observed correlations between two traits. Using summary statistical data from publicly available genome-wide association studies, we found that a genetic risk score of loci associated with glucose-stimulated insulin secretion at genome wide significance correlated with increased BMI; this was true whether we used SNPs identified by 30-minute insulin levels with or without adjustment for BMI (beta= 0. 089 and *P*=8. 6x10-7, beta=0. 061 and *P*=3. 5x10-6, respectively). Conversely, there was no relationship between a genetic risk score comprised of SNPs associated at genome-wide significance with BMI and 30-minute insulin levels (beta= 0. 005 and *P*=0. 80). This analysis provides evidence that higher levels of glucose-stimulated insulin secretion may play a causal role in obesity.

1447W

Genetic determinants of Multiple Sclerosis susceptibility in US minority populations. *A. Beecham[1,2], L. Amezcua[3], N. Isobe[4], P. Manrique[1], B. T. Lund[3], A. Levy[3], D. Conti[5], G. Beecham[1,2], P. De Jager[6], S. Delgado[7], J. Oksenberg[4], J. McCauley[1,2], International Multiple Sclerosis Genetics Consortium.* 1) John P. Hussman Institute for Human Genomics, Miller School of Medicine, University of Miami, Miami, FL, USA; 2) Dr. John T. Macdonald Department of Human Genetics, Miller School of Medicine, University of Miami, Miami, FL, USA; 3) Department of Neurology, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 4) Department of Neurology, University of California at San Francisco, San Francisco, CA, USA; 5) Department of Preventative Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA; 6) Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, MA, USA; 7) Multiple Sclerosis Division, Department of Neurology, Miller School of Medicine, University of Miami, Miami, FL, USA.

Multiple Sclerosis (MS) exhibits variable prevalence across populations, with European populations having a higher prevalence than either Hispanic or African. In addition to environmental influences, this observation could be partly due to a greater genetic risk in European populations. Genetic association studies in individuals of European descent have identified 110 established MS risk variants in 103 discrete loci outside of the Major Histocompatibility Complex. Our goal was to determine whether these 110 variants also confer risk for MS in genetically admixed samples from Hispanic and African American populations. Targeted genotyping was performed using an Illumina ExomeChip+ BeadChip, inclusive of a dense set of markers within 1Mb of the 110 variants as well as ancestry informative markers (AIMS). Separately in the Hispanic (1056 cases and 1232 controls) and African American (1319 cases and 1173 controls) samples, logistic regression was used to test for an association of the additive genetic effect of each variant on MS risk after controlling for global and local ancestry. Global ancestry proportions of European, African, and Asian admixture were computed by ADMIXTURE with 12,073 AIMS. Local ancestry proportions for each of the 110 variants were computed with RFMix after first phasing the haplotypes within 1 MB of the 110 variants using Beagle. Additionally, a combined genetic risk score for the 110 variants was calculated using odds ratios previously published in studies of European populations, and an association between MS and the risk score was assessed. The genetic risk score was significantly higher in MS cases than controls in both samples (one-sided p=2. 08E-30 in Hispanics and p=2. 93E-14 in African Americans). In Hispanic cases, 94 of 107 of the risk variants had over-representation of the same allele as seen in European cases (40 demonstrating one-sided p < 5. 0E-02), while in African American cases only 78 of 107 of the risk variants had over-representation of the same allele (20 demonstrating one-sided p < 5. 0E-02). A greater proportion of the 110 variants show over-representation of the European risk allele in Hispanic MS cases than in African American MS cases. This could be due in part to the increased European admixture at these loci in Hispanics as compared to African Americans. A more in depth look at the local ancestry at these loci will provide further insight into the ancestral origin of these risk alleles.

1448T

Comparisons of Genetically Inferred and Self-reported Race/Ethnicity in US Adults: Results from the National Epidemiologic Survey on Alcohol and Related Conditions – III. *H. Zhang, J. Jung, R. Goldstein, B. Grant.* NIH/NIAAA, Rockville, MD.

Race and ethnicity are widely used in epidemiologic studies. Accuracy of race/ethnicity classification therefore has important implications for the interpretation of findings. Information on race/ethnicity is typically acquired through self-report. Therefore, it is necessary to assess the validity of self-reported race/ethnicity using genetic data. This study evaluated the consistency between genetically inferred and self-reported race/ethnicity in 15,578 samples collected from the National Epidemiologic Survey on Alcohol and Related Conditions-III (NESARC-III). Self-reported race/ethnicity consisted of: non-Hispanic white, black, Asian, Native Hawaiian/Other Pacific Islander (NHOPI), American Indian/Alaska Native (AIAN); Hispanic; and multiple-race. To infer genetic race/ethnicity, we included 924 samples from 11 Hapmap3 populations, 940 individuals from 51 populations of the Human Genome Diversity Panel and 347 individuals from four 1000-genome project as reference populations. First, we applied PLINK software to generate genome-wide IBS estimates between 15,578 NESARC-III and 2,211 reference samples at shared 10,612 common SNPs. Then, we produced multidimensional scaling plots for visual examination of genetically inferred race/ethnicity. We segmented our samples into 11 genetically inferred groups: European American, Middle East, African American, West African, East African, East Asian, C. S. Asian, Native American, Hispanic, Native Hawaiian/Pacific Islander, and multiple-race. Finally, we used kappa statistics to gauge agreement between self-reported and genetically inferred race/ethnicity. Overall consistency between self-reported and genetically inferred race/ethnicity was excellent (χ=0. 83, 95% CI 0. 82-0. 84). Non-Hispanic whites, blacks and Asians had the highest agreement (consistency rates: 96. 9%, 94. 7%, 90. 3%, respectively). Consistency among Hispanics was fair, with 72. 9% of self-reported Hispanics classified as Hispanic by genetic analysis. Consistency rates were 12. 5% for AIAN and 45. 2% for NHOPI. Self-reported and genetically inferred race/ethnicity agreed well, especially for US whites and blacks. Although the data sources agreed well among Asians, this group is highly heterogeneous. Agreement was fair among Hispanics, mainly because both genes and culture strongly affect self-identification as Hispanic. Improvements are needed to the classification of race in AIAN and NHOPI, among whom consistency rates were lower. .

1449W

A New Approximate Bayesian Computation (ABC) framework based on Local Ancestry to Infer Admixture Events. *M. H. Gouveia[1], T. P. Leal[1], F. S. G Kehdy[1], G. S. Araújo[1], W. C. S Magalhães[1], M. R. Rodrigues[1], G. B. Soares-Souza[1], A. R. Horimoto[2], M. L. Barreto[3], A. C Pereira[2], M. F. Lima-Costa[4], B. L. Horta[5], M. O. Scliar[1], E. Tarazona-Santos[1].* 1) UFMG-Brazil, Belo Horizonte, Minas Gerais, Brazil; 2) Instituto do Coração, Universidade de São Paulo, 05403-900, São Paulo, São Paulo, Brazil; 3) Instituto de Saúde Coletiva, Universidade Federal da Bahia, 40110-040, Salvador, Bahia, Brazil; 4) Instituto de Pesquisa Rene Rachou, Fundação Oswaldo Cruz, 30190-002, Belo Horizonte, Minas Gerais, Brazil; 5) Programa de Pós-Graduação em Epidemiologia, Universidade Federal de Pelotas, CP 464, 96001-970 Pelotas, RS, Brazil.

Numerous methods have been developed to infer, by genomic data, admixture parameters (APs) such as admixture times and/or the strength of admixture events. However, none of these methods use as direct information, the chromosomal segments of continuous specific ancestry (CSSA), also called admixture tracts. The CSSAs lengths are very informative about the time when the admixture event happened, whereas the recombination process fragments the CSSAs in lengths smaller and smaller over time. We implemented a new Approximate Bayesian Computation (ABC) approach based on local ancestry to infer historical APs, conditioning on a model of pulses of immigration. The main steps of this approach are: (1) generation of an informative prior distribution of admixture parameters for each pulse, conditioning on the estimated total continental ancestry; (2) simulation of CSSA lengths, based on the prior distribution; (3) computation of the distance between the simulated and observed CSSA lengths distributions; (4) estimation of the posterior distribution of the admixture parameters for each pulse by retaining the simulated CSSA lengths distributions that are more similar to the observed distribution. We simulated CSSA lengths in accordance to the Brazilian demographic history of three pulses (18-16, 12-10 and 6-4 generations ago) of tri-hybrid admixture. This demographic model conciliates statistical complexity and the real history of admixture. Interestingly, ABC results show that the observed low Native American ancestry was mostly introduced in different regions of Brazil soon after the European Conquest of the Americas, which is consistent with the posterior depletion of the Native American population in Brazil. Also, we inferred a predominantly earlier European colonization in the Northeast (Salvador) versus a more recent immigration in Southeastern and Southern Brazil (Bambuí and Pelotas), consistently with historical records. Conversely, African admixture showed a decreasing temporal trend shared by the three studied populations. Currently, we are improving our ABC framework by testing other demographic models with more than three admixture pulses an implementing validation tests. We are also inferring the admixture dynamic of the 1000 genomes admixed populations (Afro Americans -ASW, Colombians - CLM, Mexicans - MXL, and Puerto Ricans - PUR)and other admixture Brazilian (Southeast) and Peruvian (Quechuas) populations from Tarazona-Santos group LDGH dataset.

1450T

Demographic inferences from 447 complete human genome sequences from 148 populations worldwide. *M. Metspalu[1], L. Pagani[2], D. Lawson[3], A. Kushniarevich[1], R. Mägi[5], L. Saag[1], A. Eriksson[4], A. Manica[4], T. Kivisild[2], International Collaboration effort of the Estonian Centre for Genomics.* 1) Estonian Biocentre, Tartu, Estonia; 2) Department of Biological Anthropology, University of Cambridge, Cambridge, United Kingdom; 3) Heilbronn Institute, School of Mathematics, University of Bristol, Bristol BS8 1TH, UK; 4) Department of Zoology, University of Cambridge, Cambridge, UK; 5) Estonian Genome Center, University of Tartu, Tartu, Estonia.

Complete high coverage individual genome sequences carry the maximum amount of information for reconstructing the evolutionary past of a species in the interplay between random genetic drift and natural selection. Here we use a novel dataset of 447 human genomes sequenced at 40X on the same platform (Complete Genomics) and uniform bioinformatic pipelines. Based on SNP-chip data we generally chose three samples to represent each population of interest. We cover a wide range of mostly Eurasian populations with additional populations from Oceania, South America and Africa. Here we describe the dataset in terms data quality and new recovered genetic variation that originates predominantly from previously subsampled continental regions. Using MSMC, D-statistics and Finestructure we have shown that peopling of the World from Africa is best explained by at least two migration waves (See Lawson et al abstract nr …). Here we expand on these conclusions by investigating short IBD segment sharing patterns using diCal, Hapfabia etc. We also disentangle split times involving the two migrations out of Africa (OoA), by running MSMC separately on genome chunks derived from OoA1 and OoA2. We also present detailed regional population histories in reconstructions of past dynamics of effective population size and population split times.

1451W

Reconstructing the Genetic History of Indigenous Caribbean Populations. *T. Schurr[1], J. Benn Torres[2], M. Vilar[3], C. Melendez[4], G. Torres[2], J. Gaieski[1], M. Stevenson[5], R. Bharath Hernandez[6], Z. Browne[5], W. Waters[5].* 1) Anthropology, University of Pennsylvania, Philadelphia, PA; 2) Anthropology, University of Notre Dame, Notre Dame, IN; 3) Science and Exploration, National Geographic Society, Washington, DC; 4) Liga Guakia Taina-Ke, Humacao, Puerto Rico; 5) The Garifuna Heritage Foundation Inc. , Kingston, St. Vincent; 6) Santa Rosa First Peoples Community, Arima, Trinidad.

In collaboration with the Garifuna/Kalinago of St. Vincent, the First People's Community of Arima, Trinidad, and Taíno descendant communities in Puerto Rico, we are conducting an anthropological genetic study of the prehistoric and historic settlement of the Caribbean. Using genetic data generated with the GenoChip, we are evaluating hypotheses concerning the original settlement of the Greater and Lesser Antilles, as well as the expansion of Carib and Awakan-speaking populations into this region over the past few thousand years. Our initial results suggest that the Greater Antilles were colonized by indigenous populations from South America and possibly Mesoamerica, whereas the Lesser Antilles were settled by only South American groups. In addition, while sharing some indigenous mtDNA (maternal) and Y-chromosome (paternal) lineages in common, populations from the Greater and Lesser Antilles otherwise appear to be largely genetically distinct from each other. Autosomal SNP data from these indigenous Caribbean communities further expand our understanding of the genetic contributions from African, European and South Asian populations since European contact. Overall, this study demonstrates the region's first peoples' ongoing legacy in shaping the genetic diversity of contemporary Caribbean populations.

1452T

Strong selection at MHC in Mexicans since admixture. *Q. Zhou[1,2,3], L. ZHAO[1,2], Y. GUAN[1,2,3,4].* 1) USDA/ARS Children's Nutrition Research Center, Houston, TX; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 3) SCBMB, Baylor College of Medicine, Houston, TX; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Mexicans are recent admixture of Amerindians, Europeans and Africans. We performed local ancestry analysis of Mexican samples from two genome-wide association studies obtained from dbGaP and discovered that at the MHC region Mexicans have excessive African ancestral alleles compared to the rest of the genome, which is the hallmark of recent selection for admixed samples. The estimated selection coefficients are 0. 07 and 0. 09 for two datasets, which put our founding among the strongest known selections observed in humans, namely, lactase gene in northern Europeans and sickle-cell allele in Africans. Inaccurate Amerindian training samples was a major concern for the credibility of previously reported selection signals in Latinos. Taking advantage of the flexibility of our statistical model, we devised model fitting method that can learn Amerindian ancestral haplotypes from the admixed samples, which allow us to infer local ancestries for Mexicans using only European and African training samples. The strong selection signal at MHC remains without Amerindian training samples. One wonders why such a strong selection signal was not discovered by 1000 Genomes project in their analysis of Mexican samples using other competing local ancestry inference models. Our simulation studies suggested that the approach adopted by 1000 Genomes admixture analysis group, which used consensus estimates from four methods, is perhaps to blame. Finally, we pointed out that medical history studies suggested such a strong selection signal is plausible in Mexicans.

1453W

Genetic origins and admixed ancestry characterization of Japanese people. *W. Ko[1,2], K. Higasa[3], M. Narahara[2], F. Matsuda[3], R. Yamada[2].* 1) Faculty of Life Sciences and Institute of Genome Sciences, National Yang-Ming University, Taipei, 112, Taiwan; 2) Statistic Genetics, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, 606-8507, Japan; 3) Human Disease Genomics, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, 606-8507, Japan.

A modern human population found at a certain geographic location is often descended from multiple ethnic groups owning to the complex migration history of human expansion. In Japan, although it has been studied extensively over the past decades, the genetic origins of Japanese people remain controversial. Current genetic evidence supports a dual model which suggested that the Japanese people are constituted mainly by an early settlement of human populations during the Upper Paleolithic period (*i. e. ,* Jomon people) followed by an admixture event with the people migrated from the Korean peninsula around 2300 year ago (*i. e. ,* Yayoi people). However, the genetic origin(s) of the native Jomons remains unclear. Tracing the genomic signatures of admixture history can not only reveal the unknown human migration events but also provide critical information that can facilitate the genetic profiling of disease susceptibility, which is critical for the success of personalized medicine. Here, we analyzed a combined dataset of the whole genome SNP genotyping data from 2,277 individuals sampled globally across >100 populations for a total of 19,290 SNPs (after intersecting the two datasets). We performed principle component analysis to project individuals onto a series of orthogonal axes to reveal the genetic structure among diverse ethnic groups. After separating the genetic components contributed from the populations representing the Yayoi, we identified several candidate populations that share common non-Yayoi ancestry with the modern Japanese people. Our results suggest that the genetic origins of Jomons may consist of multiple migration events from both Southeast and Northeast Asia. Surprisingly, we also identified an additional migration wave from the Hmong population. We assigned local ancestry (LA) on the phased chromosomes of the mainland and Okinawa Japanese by performing RFmix (which used the identified candidate ancestral populations to infer the LA tracts in admixed chromosomes by finding the most likely sequence of ancestries through maximum a posterior estimation). Because an ancient population admixture would allow more recombination events to break LA tracks into shorter segments than a recent admixture event, our results of the LA track-length distributions differ significantly between the Yayoi, Hmong, and Jomon ancestries (in descending order), suggesting that the Hmong migration may have occurred before the Yayoi migration.

1454T

The genetic structure of the Saudi Arabian population. *H. Al-Saud[1], SM. Wakil[1], BF. Meyer[1], M. Falchi[2], N. Dzimiri[1].* 1) Genetics Department, King Faisal Hospital and Research Centre, Riyadh, Saudi Arabia; 2) Department of Twin Research and Genetic Epidemiology, King's College London, London, United Kingdom.

Saudi Arabia is the largest Gulf Cooperation Council (GCC) country. Its population consists of different tribes that originated in the northern, western, eastern, middle and south regions of Saudi Arabia, respectively. Due to political and cultural reasons, there has historically been very limited admixture between different tribes. People from the different Saudi tribes then migrated from Saudi Arabia, contributing to foundation of the populations now inhabiting other Gulf countries. Few population genetics research projects have been conducted on this highly consanguineous population that has been shown to have one of the highest prevalence in the world of recessive disorders and common metabolic diseases, especially diabetes. It is therefore important to identify the genetic substructures of the Saudi population, both to help in tracing the migratory genetic flows that contributed to other Gulf populations, and to permit designing of efficient genetic studies aimed at the identification of risk factors underlying common and rare diseases in the GCC countries. We carried out the largest population genetic study in Saudi Arabia to date, by genotyping 2,150 Saudi nationals sampled from different regions of Saudi Arabia using Axiom GWH-96 Array (Affymetrix) arrays. Model-based and model-free clustering were applied to these data, including in our analyses data on eight populations (encompassing Europe, America, Oceana, East Asia, Central South Asia, Middle East, Africa and Qatari populations) from the Human Genetic Diversity Project (HGDP) data set. We identified clear clustering of the Saudi samples into different subgroups, with some tribes showing similarity with both Central East Asian (Kalash Pakistan, Balochi Pakistan, Sindhi Pakistan, Makrani Pakistan and Brahui Pakistan subpopulations) European (Orkney Islands Europe, Russian Europe and Russian Caucasus subpopulations) and Qatari populations, while other tribes appear to show specificity of background. These data strongly support the presence of genetic stratification within the Saudi population, and suggest the presence of subgroups that are characterized by a unique genetic background different from other Arabian populations. Our findings constitute a valuable resource for the investigation of both general and population-specific genetic risk variants associated with different disorders in this population.

1455W

Recent genetic history of Denmark. *G. Athanasiadis[1,2], F. G. Jørgensen[3], J. Cheng[1,2], P. C. Kjærgaard[2,4], M. H. Schierup[1,2,5], T. Mailund[1,2].* 1) Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark; 2) Centre for Biocultural History, Aarhus University, Aarhus, Denmark; 3) Tørring Gymnasium, Tørring, Denmark; 4) Department of Culture and Society, Aarhus University, Aarhus, Denmark; 5) Department of Bioscience, Aarhus University, Aarhus, Denmark.

Purpose. - Denmark has strong historical bonds not only with Norway and Sweden, but also with Western and Eastern Europe through a series of invasions, conquests and alliances. In addition, within Denmark, industrialization in the second half of the 19th century led to considerable migration from the countryside to the cities. In this work we explore the extent to which such distant and more recent historical events left their mark on the genetic structure of the current Danish population. *Methods.* - We ran an extensive genetic analysis on the *Where Are You From?* data set of ~600 students from 36 high schools across Denmark. Each student provided a saliva sample for DNA analysis and completed an online questionnaire about family origin, education level and basic biometrical data. All participants gave their informed consent and the Ethical Committee of the University of Aarhus approved the study. Genotyping was outsourced to 23andMe and more than 500,000 SNPs were available for analysis. After merging our data with data from POPRES, we ran PCA and ADMIXTURE to detect genetic structure. For more fine-grain effects, we identified each individual's closest genetic relatives through IBD tract sharing and calculated the geographic distance between the individual's place of birth and the weighted average geographic coordinates of their closest relatives. Finally, we explored population structure within Denmark as the result of recent admixture with adjacent populations by use of an IBD-based local ancestry method (i. e. "chromosome painting"). *Results.* - Although Denmark forms a distinguishable cluster from neighboring countries in the PCA plots (compatible with isolation-by-distance), no stong structure was observed within the country. Similarly, ADMIXTURE revealed high levels of homogeneity in the Danish samples compared to other North European countries. However, we did observe significant correlation between PC1 (south-north orientation) and average grandparental geographic coordinates rotated clockwise at ~30°. Also, the IBD-based geographic correlation analysis revealed that Danes tend to live near their closest genetic relatives at a median distance of 100 Km - significantly closer than the random expectation. Finally, chromosome painting revealed strong genetic influence from neighboring Nordic (Sweden and Norway) and Germanic (Germany and Holland) countries and negligible influence from Finland, France and Portugal.

1456T

Assessing the benefits of priors that encourage sparsity for estimating ancestral admixture from genome-wide data. *P. Carbonetto, Y. Wang, K. Noto, M. Barber, J. Byrnes, R. Curtis, K. Chahine, J. Granka, E. Han, A. Kermany, N. Myres, C. Ball.* AncestryDNA, San Francisco, CA.

Several recent papers have demonstrated the benefits of using sparse matrix factorization techniques---sparse factor analysis and non-negative matrix factorization---to infer population structure from genetic polymorphism data. The primary strength of sparse matrix factorization is its flexibility; it can capture a wide range of population structure scenarios, and can do so in a way that often has a natural interpretation. For example, sparse matrix factorization is able to recapitulate a mixture of continuous and discrete population structure, whereas other methods, such as PCA and STRUCTURE, cannot do this. However, we have found that this flexibility can come at a cost: in realistic demographic settings, it incorrectly predicts individual admixture proportions. We hypothesize that this is because sparse matrix factorization does not completely specify an admixture model. Motivated by this, we propose a model-based approach, building on ADMIXTURE, that encourages sparsity in the admixture proportions (or "loadings"). We encourage sparse estimates by introducing an exact L0-norm penalty term in the cost function that penalizes non-zero admixture proportions, then we iteratively solve for the model parameters using a hybrid EM algorithm. This penalty can also be interpreted as a prior on the number of ancestral populations contributing to an individual's genome. We explore the behaviour of penalized and unpenalized admixture estimates in data from the Human Genome Diversity Project. Although the idea of encouraging sparse admixture estimates has been suggested previously, to our knowledge the features of this approach have not been empirically assessed in real genetic data from human populations.

1457W

Whole Genome Sequence Identity by Descent Segment Sharing Across Diverse Populations. *N. Chambwe[1], D. E. Mauldin[1], J. G. Vockley[2], J. E. Niederhuber[2], I. Shmulevich[2], J. C. Roach[1].* 1) Institute for Systems Biology, 401 Terry Ave, North, Seattle, WA 98103; 2) Inova Translational Medicine Institute, Inova Health Systems and Inova Fairfax Medical Center, 3300 Gallows Road, Falls Church, VA, 22042.

The identification of shared genomic segments that are identical by descent (IBD) within pedigrees is a useful tool for human genetics. Firstly IBD analysis is important for determining loci implicated in the inheritance of mendelian traits. Secondly, IBD segments form the basis of relationship detection for algorithms such as Estimation of Recent Shared Ancestry (ERSA). Population haplotype diversity at a particular locus strongly influences the likelihood that a pair of individuals is identified as IBD at that locus, as do technical factors such as mappability of short reads to a reference genome. The statistical properties of IBD sharing at particular locations can vary highly, both due to population genetics and technical influences. Technical influences gain in relative importance for shorter IBD tracts, which can be identified by WGS but are not generally identified by older methodologies. To best evaluate the significance of candidate loci and relationship estimations, these locus-specific statistical properties of IBD blocks must be understood and evaluated in a control dataset. Extensive IBD analyses have been carried out using genotyping data generated from microarray platforms. However, whole genome sequencing (WGS) surveys IBD with greater precision and with markers that have distinct statistical properties from those in genotyping panels. Statistical properties of IBD sharing have yet to be fully explored with data generated from WGS. In order to determine IBD segment sharing from WGS data, we applied Inheritance State Consistency Analysis (ISCA) to more than 5000 individuals from diverse populations (INOVA Genomes). These data were generated using DNA from peripheral blood samples sequenced on the CGI and Illumina platforms. We describe global properties of IBD segments such as the segment length distribution and compare differences across populations and phenotypic categories. IBD segment sharing across populations largely agrees with expectations from population structure. We observed that the proportion of the genome with shared IBD1/IBD2 segments was higher between individuals from non-African populations. Increased variability in haplotype diversity across the genome correlates with both long and short IBD segments. This largely agrees with previous findings using genotyping platforms for long IBD segments, but extends the observation to shorter IBD segments we are able to detect with WGS data.

1458T

Characterizing Brazilian sickle cell anemia patients. *P. RS. Cruz[1], G. Ananina[1], M. A. C. Bezerra[3], A. S. Araujo[3], W. M. Avelar[4], A. C. Amato-Filho[4], F. Cendes[4], F. F. Costa[2], M. B. Melo[2].* 1) Center of Molecular Biology and Genetic Engineering (CBMEG), University of Campinas, Campinas, São Paulo, Brazil; 2) Hematology and Hemotherapy Center/HEMOCENTRO, University of Campinas, Campinas, São Paulo, Brazil; 3) Hematology and Hemotherapy Center of Pernambuco/HEMOPE, Recife, Pernambuco, Brazil; 4)) Neuroimaging Laboratory, Department of Neurology, University of Campinas, Campinas, São Paulo, Brazil.

Sickle cell anemia (SCA) is a devastating condition that affects millions and lacks a safe drug therapy. More than 500,000 newborns worldwide are thought to be affected a year. James B. Herrick reported SCA for the first time in 1910 and by the year of 1949 was the first so called molecular disease to have its mechanism unveiled. Before modern medical care, sickle-cell heterozygotes (*HbAHbS*) had the highest fitness in malaria-endemic regions, since they have a lower susceptibility to host *Plasmodium malariae*relative to the *HbA* homozygote and tend not to develop sickle-cell complications. *HbS* mutation is assumed to have arisen multiple times around 5 kya. Present haplotypes are named after their putative geographical origin (Benin, CAR or Bantu, Cameroon, Senegal and Arab) and contribute to the overwhelming phenotypic heterogeneity observed in SCA. African American is the best-studied population in respect to SCA, whereas Brazilian population is poorly described and differs from the American in origin. Slaves brought into the United States originated mainly from Central West African ports, where the Benin haplotype prevails, while the 2. 4 million African Negroes transported to Brazil, from 1701 to 1843, was predominantly from Angola, Congo, and Mozambique, areas where the CAR haplotype is more frequent. We propose, then, to characterize the Brazilian patients regarding ancestry. Ninety SCA patients were recruited at two Brazilian centers: HEMOCENTRO (Campinas, SP) and HEMOPE (Recife, PE), along with 62 *HbAHbA* allegedly healthy individuals. This study is in accordance to the Basic Principles of the Declaration of Helsinki. The genotyping protocol was carried out using Affymetrix SNP Array 6. 0 (Affymetrix Inc. , CA, USA). We filtered SNP and samples data using PLINK for probes with acceptable contrast quality (QC > 0. 40). Population structure and admixture were analyzed applying EIGENSOFT and ADMIXTURE software, along with HapMap reference populations. We found that SCA patients are more proximate to African origin (by both Fst values and ancestral composition) and that*HbS*haplotype shows linkage disequilibrium consistent with recent selection. As the next step we will further dissect chromosome 11 ancestral haplotypes in fine resolution and scan for positive selection signals. Inferring demographic and ancestral history in Brazilian SCA will provide valuable knowledge on the underlying phenotype.

1459W

Ancient European haplotype enrichment in modern Eurasian populations. *D. Harris[1], T. O'Connor[2].* 1) Graduate Program in Molecular Medicine, University of Maryland School of Medicine, Baltimore, MD; 2) Institute for Genome Sciences, Program in Personalized and Genomic Medicine, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD.

The diversification of modern European populations is a fascinating puzzle that has recently advanced due to the sequencing of ancient European genomes. We analyzed 732 modern West Eurasian individuals using three ancient samples coming from the Lazardis et al. Human Origins Array dataset. Specifically, we determined ancient European haplotype enrichment by calculating pairwise differences (PWD) between each ancient European individual and modern Western Eurasian individuals in 50 SNP blocks. Modern Western Eurasians had the fewest PWD across all population groups with the farming Stuttgart individual and had the most PWD with the Loschbour and Motala12 hunter-gatherer individuals confirming Lazardis et al. observation that modern Europeans are more similarly related to ancient individuals coming from a farming community. We selected SNP blocks, for gene ontology enrichment analysis through the use of GORILLA, based on 1) the 10% of regions with greatest differences of PWD between groups, and 2) the 10% of those regions from the first criterion that most closely correlated with the geography of those groups. Most SNP blocks positively correlated to PC1 (latitude) and PC2 (longitude), therefore we focused on outliers that negatively correlated to biogeography. For SNP blocks that negatively correlated to PC1; "regulation of chondrocyte development", "androsterone dehydrogenase activity", and "antigen processing and presentation of endogenous peptide antigen" had the highest enrichment score in the comparison of the Stuttgart, Loschbour, and Motala12 individuals, respectively. Interestingly, the "alpha-beta T cell receptor complex" and "interleukin-17 receptor activity" (including CD3D,E,G and IL17RC,E) were enriched in the Loschbour and Motala12 comparisons of SNP blocks that were positively correlated to PC2. In addition, the Stuttgart individual had the lowest PWD disparity between all modern populations for the SNP blocks that contain the IL17R and CD3 genes, which potentially indicates selection acting on these immune system haplotypes from the Stuttgart individual consistent with the Stuttgart farmer and modern Europeans' continual close interaction with animals and zoonotic disease exposure. In conclusion, our approach of calculating PWD in small SNP blocks supported prior conclusions made by Lazardis et al. and illuminated small genomic haplotypes that are of importance to the evolution of modern West Eurasian populations.

1460T

Refining the South Asian origin of the Roma people. *B. Melegh[1,2], Zs. Banfai[1,2], M. Kayser[3], B. Melegh[1,2].* 1) Medical Genetics, University of Pecs, Pecs, Hungary; 2) Szentagothai Research Centre, University of Pecs, Pecs, Hungary; 3) Department of Forensic Molecular Biology, Erasmus University, Netherlands.

Purpose: Historical and linguistic studies have suggested that Roma people, living mainly in Europe, migrated into the continent from South Asia about 1000-1500 years ago. Genetic studies, based on the examination of Y chromosome and mitochondrial DNA data, confirmed these findings. Recent genetic studies based on genome-wide Single Nucleotide Polymorphism (SNP) data further investigated the history of Roma and, among many other findings, suggested that the source of South Asian ancestry in Roma originates mainly form the Northwest region of India. **Methods:** In this study, using also genome-wide SNP data, we attempted to refine these findings using significantly larger amount of European Roma samples. We also had the opportunity to use more data of distinct Indian ethnic groups, which provided us a higher resolution of the Indian population. The study uses several ancestry estimation methods based on the algorithmic method principal component analysis and model-based methods that apply Bayesian approach and uses Markov chain Monte Carlo or maximum likelihood estimation. **Results:** According to our analyses, Roma showed significant common ancestry with Indian ethnic groups of Jammu and Kashmir, Punjab, Rajasthan, Gujarat, Uttarakhand states, e. g. with Kashmiri Pandit, Punjabi, Meghawal, Gujarati and Tharu. However, we found strong common ancestry with Pashtun and Sindhi, ethnic groups living in Pakistan. Populations of Northeast India have also strong common ancestry with Roma. These ethnic groups are Brahmin, Kshatriya, Vaish. **Conclusion:** We can conclude, that Northwest India plays an important role in the South Asian ancestry of Roma, but they have similarly strong ancestry with some Pakistani ethnic groups and we can find populations in the east region of North India, which also could function as a source of Indian ancestry of Roma. However, ethnic groups of the southern region of India do not show strong relationship with Roma people, living in Europe.

1461W

Admixture and ancestry patterns of three Brazilian quilombo remnants communities. *C. T. Mendes-Junior[1], D. M. Salvanha[2], R. Z. N. Vêncio[2], L. M. Garrido[3], H. Krieger[3], E. C. Castelli[4], A. L. Simões[5], E. S. Andrade[5].* 1) Departamento de Química, Universidade de São Paulo (FFCLRP-USP), Ribeirão Preto, SP, Brazil; 2) Departamento de Computação e Matemática, Universidade de São Paulo (FFCLRP-USP), Ribeirão Preto, SP, Brazil; 3) Departamento de Parasitologia, Universidade de São Paulo (ICB-USP), São Paulo, SP, Brazil; 4) Departamento de Patologia, Universidade Estadual Paulista (FMB-UNESP), Botucatu, SP, Brazil; 5) Departamento de Genética, Universidade de São Paulo (FMRP-USP), Ribeirão Preto, SP, Brazil.

The Brazilian population results of the admixture between European, African and Amerindian populations. Complex historical factors have generated variables proportions of ancestry contribution between individuals and between different admixed populations. This study aimed to describe the admixture and ancestry patterns of three quilombo remnants populations from Brazil, Gaucinha (GAU, n = 14), Mimbó (MIB, n = 15) and Sítio Velho (STV, n = 15) and of an urban population, Teresina (TES, n = 15). Data from seven population samples from the HapMap Project (CEU, CHB, JPT, ASW, LWK, MKK, YRI, all with n = 15) were used as reference. Brazilian samples were genotyped for around 250 thousand SNPs using the GeneChip® Human Mapping 250K Nsp Array I - Affymetrix®. A set of 409 Ancestry Informative Markers (AIMs) was selected from panels described in the literature. Of these AIMs, 35 were present in the SNP array. Genetic distance (*FST*) estimates were performed using the Arlequin software. Global ancestry was estimated using ADMIX95, ADMIX 2. 0 and Structure 2. 3. 4 softwares. This latter was also used to evaluate population structure and individual ancestry. *FST* did not reveal significant genetic differences between the three quilombos. Nevertheless, they differ considerably in their proportions of ancestry composition, which were consistent among the three methods performed. Therefore, only the values generated from Structure are presented here. GAU and MIB presented low European contribution (0. 167 and 0. 185, respectively). GAU presented an African contribution of 0. 479, while in MIB was of 0. 600. Besides, GAU showed a higher Asian/Amerindian contribution (0. 354) than MIB (0. 215). The African contribution in STV was similar to that of GAU, while the Asian/Amerindian contribution was close to that of MIB. The European contribution in STV is much higher (0. 344) than in GAU and MIB. A possible explanation is that the founders of STV came from wider areas with diverse populations while in GAU and MIB they were direct descendants of black people already present in the region. TES presented different patterns in comparison with the three quilombos, with low values of African ancestry (0. 221) and high European ancestry (0. 546), as expected for an urban population. The individual genetic ancestry and genetic differentiation estimates corroborate all these results. FINANCIAL SUPPORT: FAEPA (Grant 1394/2011), CAPES, FAPESP (2013/15447-0) and CNPq/Brazil (309572/20142).

1462T

Sex-Biased Admixture in the Americas. *S. Musharoff[1], C. R. Gignoux[1], S. Shringarpure[1], M. A. Taub[2], T. O'Connor[3], R. A. Mathias[4], C. D. Bustamante[1], K. C. Barnes[4], CAAPA Consortium.* 1) Genetics, Stanford University School of Medicine, Stanford, CA; 2) Department of Biostatistics, Johns Hopkins University, Baltimore, MD; 3) University of Maryland School of Medicine, Baltimore, MD; 4) Division of Allergy & Clinical Immunology Department of Medicine, Johns Hopkins University, Baltimore, MD.

We studied sex-biased population histories from **C**onsortium on **A**sthma among **A**frican-ancestry **P**opulations in the **A**mericas (CAAPA) high-coverage whole genomes (~30x depth). CAAPA comprises 673 individuals who are African-American, African, Afro-Caribbean (Barbados, Jamaica), or Latin-American (Colombia, Brazil, Puerto Rico, Honduras, Dominican Republic). X chromosomes show a decrease of European ancestry as estimated with ADMIXTURE, consistent with a history of European male-driven colonization. CAAPA Latin Americans have female-biased Native American ancestry (5. 36% mean excess X-chromosomal), male-biased European ancestry (1. 36% mean excess autosomal), and female-biased African ancestry (6. 72% mean excess X-chromosomal). Some CAAPA African-descent populationshave never been studied genetically. The Garifuna from Honduras have very little autosomal European ancestry (2. 2%) but high Native American ancestry (16. 6%). The Afro-Brazilians from Condé have a high proportion of African ancestry (50. 5%). The Cartagena Colombians (from one of two slave ports in South & Central America) have more African ancestry than the TGP Colombians (CLM): on average CAAPA individuals have 31. 1% autosomal and 29. 7% X-chromosomal African ancestry and TGP CLM have 7. 7% and 6. 8%, respectively. Y and MT haplotype analysis support the above sex-biased admixture findings: Afro-Caribbeans have African mitochondria, Latin Americans have a mix of African and Native American mitochondria, yet both groups have mostly European Y chromosomes. We identify three Native American Y haplotypes in the Honduran Garifuna only, highlighting their unique history. Unexpectedly we identified a new subgroup of MT-E1a1a that suggests a connection with the Malagasy slave trade. We apply a novel method to infer sex-biased demography during specific time epochs to autosomal and X-chromosomal site frequency spectra. CAAPA Latin Americans show evidence for a female bias over a longer time scale, male-biased bottlenecks Out-of-Africa and into the Americas, and male-biased admixture events. We analyze ancestry tracts with the program TRACTS to estimate timings and magnitudes of sex-biased admixture events. Overall, our findings recapitulate the complex history of the Americas and highlight key differences between populations based on their local admixture histories. As this is the first time some of these unique populations have been studied, this represents a valuable population and medical genetic resource.

1463W

A novel approach to developing an ancestry informative marker panel for individual identification and population admixture characterization based on common population-specific SNPs. *D. Sengupta[1], A. Choudhury[1], M. Ramsay[1,2].* 1) SBIMB, University of the Witwatersrand, Johannesburg , South Africa; 2) Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

Background: Most ancestry informative marker (AIM) panels to date have been based on allele-frequency differences between populations. An alternative approach is to use population specific SNPs that are observed primarily in a single population. SNP array platforms used for GWAS and fine mapping have had a euro-centric sampling bias, and therefore have not been ideal to clearly differentiate diverse global populations or admixture at the level of the individual. The availability of large scale whole genome sequence data in recent years has made it possible to explore the potential for the design of an AIM panel based on population specific genetic variation. **Methods:** We have designed an AIM panel based on common (MAF greater than0. 05) population specific (CPS) SNPs identified from the 1000 Genomes Project (Phase 1) data. The CPS-SNP set for each population was LD pruned and a sub-set of SNPs representing the population a given number of times was identified using a novel iterative-masking algorithm. **Results:** The performance of the CPS-SNP AIM panel was evaluated by testing the accuracy of ancestry prediction for individuals from population studies with whole genome datasets including Complete Genomics, 1000 Genomes Project (Phase 3, including new populations) and individual genomes in the public domain. Despite being based on low coverage sequence data, the panel of 3,158 CPS-SNPs was able to efficiently identify the probable ethno-geographical origin of individuals, including some that were not represented in the design set. The predictions from frequency based AIM panels are probabilistic and require integration of population genetic data and computational analysis like PCA for ancestry inference. The major advantage of our approach is that the predictions are deterministic and no additional data integration and computational analysis are required. Moreover, the CPS-SNP AIM panel is modular and the size of the panel can be scaled down (or up) by including only SNPs for the ancestry components that are expected to be seen in study participants, or by including AIMs for additional parental populations. **Conclusions:** This approach shows promise for implementation in precision medicine, GWAS, forensics and individual identification. The inclusion of key populations representing Southeast Asia, Oceania and southern Africa, more high coverage data and more efficient algorithms in the panel design would significantly enhance the panel.

1464T

The Demographic Patterns Revealed by New World African Diaspora Genome. *W. Song[1], R. A. Mathias[2], K. C. Barnes[2], T. D. O'Connor[1], CAAPA Consortium.* 1) Medicine, University of Maryland, Baltimore, MD; 2) Medicine, Johns Hopkins University, Baltimore, MD.

One of the great interests in human genetics research is to understand human population structure, demographic patterns, and evolutionary history. New World populations, such as African Americans and Latino Americans with African ancestry, provide good examples for studying large migration and admixture events in recent human history. Three questions in particular are: 1) where did different sources of admixture come from, 2) when did admixture happen, and 3) what is the difference among subpopulations. To answer these questions we will make use of the *Consortium onAsthma among African-ancestry Populations in the Americas (CAAPA),* which contains high coverage (~30× depth) whole genome sequence data of 952 individuals of African ancestry. These individuals were selected from populations in North and South America, the Caribbean and continental Africa to form a large spectrum of New World African Disapora. We merged all CAAPA data with 1963 individuals from the publically available Human Origins genotype data. After filtering rare variation (MAF < 5%), there are 389,397 SNPs in autosomal chromosomes left for the analysis of population admixture in this work. The filtered SNPs are first phased using Shapeit with the reference panel from 1000 Genomes Project. PCA-based local ancestry estimation on the CAAPA dataset is performed with PCAdmix, using the continental reference samples from Human Origin dataset. Ancestry-specific PCA (ASPCA) Analysis of PCAmask, in which ancestry specific regions from Europeans, Africans, and Native Americans are masked in PCA with sub-continental reference panels, reveals that the European ancestry in these New World African Diaspora populations comes from two main parts of Europe: Northwest (English/French) and Southwest (Spanish). We use Malder and Tracts to identify the timing of admixture in these populations. The African and Native-American ancestries admix with each other about 13-16 generations ago and later European ancestry entered into these populations 6-8 generations ago. We show that the origin and time of European introgression are different between New World African ancestry populations. Our results clearly reflect the ancestry patterns of African admixed populations in America and provide a general pipeline to study the evolutionary history of other New World populations.

1465W

The genetics of Bene Israel from India reveals both substantial Jewish and Indian ancestry. *Y. Waldman[1,2], A. Biddanda[1], N. R. Davidson[1], P. Billing-Ross[1], M. Dubrovsky[3,4], C. L. Campbell[5], C. Oddoux[5], E. Friedman[3,4], G. Atzmon[6,7], E. Halperin[2,8,9], H. Ostrer[5,10], A. Keinan[1].* 1) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY; 2) Department of Molecular Microbiology and Biotechnology, Tel Aviv University, Tel Aviv, Israel; 3) Danek Gertner Institute of Human Genetics, Chaim Sheba Medical Center, Tel-Hashomer, Israel; 4) Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel; 5) Department of Pathology, Albert Einstein College of Medicine, Bronx, NY; 6) Departments of Medicine and Genetics, Albert Einstein College of Medicine, Bronx, NY; 7) Department of Human Biology, Faculty of Natural Sciences, University of Haifa, Haifa, Israel; 8) The Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel; 9) International Computer Science Institute, Berkeley, CA; 10) Department of Pediatrics, Albert Einstein College of Medicine, Bronx, NY.

The Bene Israel Jewish community from West India is a unique population whose history before the 18th century remains largely unknown. Bene Israel members consider themselves as descendants of Jews, yet the identity of these ancestors and the time of their arrival to India are unknown, with speculations on arrival time varying between the 8th century BCE and the 6th century CE. Their genetic history remains similarly vague: Previous genetic studies showed that most Jewish Diasporas share ancestry that can be traced back to the Middle East, in accordance with historical records. An exception is the analysis of Bene Israel, for which previous studies did not find clear evidence for Jewish ancestry and stressed instead their similarity to Indian populations (with the exception of a suggestive Middle-Eastern link of the Y haplogroups in a set of four individuals). A challenge to the analysis of Bene Israel is the complex genetic structure of Indian populations, which are the result of ancient admixture of two ancestral populations, one of which more closely related to Eurasians and, hence, Jewish populations. Here, we characterize the genetic history of Bene Israel by collecting and genotyping 18 individuals from the population. Combining with 438 individuals from 32 other Jewish and Indian populations, and additional individuals from worldwide populations, we conducted comprehensive genome-wide analyses based on FST, PCA, ADMIXTURE, identity-by-descent sharing, admixture LD decay (ALDER), haplotype sharing (GLOBETROTTER), and allele sharing autocorrelation decay, as well as contrasted patterns between the X chromosome and the autosomes. Putting the results together point to Bene Israel individuals resembling local Indian populations, while at the same time constituting a clearly separated and unique population in India. They share genetic ancestry with other Jewish populations to an extent not observed for any other Indian population. We show that Bene Israel is an admixed population with both Jewish and Indian ancestry, with the genetic contribution of each of these ancestral populations being substantial. The admixture took place in the last millennium, about 19-33 generations ago. It involved Middle-Eastern Jews and was sex-biased, with more male Jewish and local female contribution. It was followed by a population bottleneck and high endogamy, which has led to the increased prevalence of recessive diseases in this population.

1466T

The Population Structure of Nunavik Inuit People. *S. Zhou[1,2], P. Xie[1], L. Xiong[3], A. Ambalavanan[1], CV. Bourassa[1], A. Dionne-Laporte[1], D. Spiegelman[1], PA. Dion[1,4], GA. Rouleau[1,4].* 1) Montreal Neurological Institute, Montreal, QC, Canada; 2) Department of Medicine, Faculty of Medicine, Université de Montréal, Montréal, QC, Canada; 3) Centre de recherche, Institut universitaire en santé mentale de Montréal, QC, Canada; 4) Department of Neurology and Neurosurgery, McGill University, Montréal, QC, Canada.

Nunavik Inuit, descendants of the Thule people who migrated from the west of Northern Canada at approximately 1,000 AD, are now the inhabitants of 14 villages across the north of Quebec. Previous studies and evidences suggested that contemporaryNunavik Inuit came from a small founder population, and their genetic profiles have shown to undergone selection to adapt to the cold climate and has therefore been uniquely preserved. However, genome-wide investigation have yet to be conducted on this group of people to explore their specific genetic background. Here we used high density bead array to genotype Inuit individualsfrom 13 Nunavikvillages, and compared their genetic patterns with those of aboriginal peoples from Americas, arctic populations from Siberia and Alaska, multiple Asian populations and French-Canadians also in Quebec. We further used exome sequencing on a selected individuals to provide additional population specific information in order to complement our findings. In result, principal component analysis (PCA) and Admixture suggested that Nunavik Inuit have extremely distinctive ancestral profiles compared to any world-wide populations, including other Canadian aboriginals as well as Siberians which commonly believed to share the same ancestors of Inuit. Increased runs of homozygosity (ROH) and elevated coefficients of inbreeding (*F*st) suggested that Nunavik Inuit have a significant amount of isolations and low gene flow in their populations, which further correlated with our findings that certain genes, such as *CPT1A* and its family members in the lipid metabolism pathway, are strongly selected for their unique adaptation to the environment. Our results provide important inputs to the studies of Nunavik Inuit susceptible diseases, it is also noteworthy that the population's unique genetic structure may help the disease gene discovery for certain complex disorders.

1467W

A genetic and socio-economic study of mate choice in Latinos reveals novel assortment patterns. *J. Y. Zou[1,2], D. S. Park[3], E. G. Burchard[3,4], D. G. Torgerson[3,4], M. Pino-Yanes[3,4], Y. S. Song[5,6,7], S. Sankararaman[8], E. Halperin[9], N. Zaitlen[3,4].* 1) Microsoft Research, Cambridge, MA; 2) Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA; 3) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA; 4) Department of Medicine, University of California, San Francisco, CA, USA; 5) Computer Science Division, University of California, Berkeley, CA, USA; 6) Department of Statistics, University of California, Berkeley, CA, USA; 7) Department of Integrative Biology, University of California, Berkeley, CA, USA; 8) Department of Genetics, Harvard Medical School, Boston, MA, USA; 9) Blavatnik School of Computer Science, Tel Aviv University, Tel-Aviv, Israel.

Understanding the patterns and drivers of nonrandom mating in human populations has important implications for genetics and medicine as well as for economics and sociology. Previous studies have provided evidence for assortative mating in diverse populations and have suggested that genetic assortment is significantly smaller than educational assortment in non-Hispanic whites. Nevertheless, there remain important open questions regarding assortment patterns (especially in non-European populations), the relationship between assortment based on genomic factors and socio-economic factors (such as income), and biological mechanisms driving such assortment. In this work we addressed these questions by jointly analyzing detailed socio-economic attributes and high-quality genotype data associated with large cohorts of Mexicans and Puerto Rican individuals sampled from multiple geographic locations. To quantify genomic assortment in the absence of parental genotypes, we developed a novel computational method, ANCESTOR, to accurately infer the genomic ancestry (fraction of the genome derived from European, Native American and African ancestries) of each parent from the offspring's genotype. We applied ANCESTOR to our richly phenotyped Latino cohorts. In ethnically homogeneous Latino communities, we quantified genomic ancestry as a key axis of assortment. Partners are much more likely to share similar genomic ancestries than random individuals. Consistent with this, partners are more closely related—equivalent to between third and fourth cousins in Mexicans and Puerto Ricans—than expected based on random mating. Our analysis showed that socio-economic factors only explain a small portion of the genomic ancestry similarity between partners. Strikingly, after normalizing by population background, partners are more similar in genomic ancestries than in education levels. We found that Puerto Rican couples are especially correlated in genes involved in facial development (above and beyond genome-wide similarities), suggesting that similarity at these genes represents an axis of assortment. We replicated our findings across multiple locations. Our analysis integrated population genomics with quantitative social sciences to address fundamental questions about mate selection.

1468T

Population Structure Analysis of Few Northwest Punjabi Populations based on Microsatellites and Alu Insertion Elements. *M. Kaur, B. Doza.* Department of Human Genetics, Guru Nanak Dev University, Amritsar, India.

Introduction: It is noteworthy that Punjabi population exhibit higher genetic affinity with Central Asia, Middle East and European populations due to previous episodes of foreign migrations and invasions. Furthermore, regional conversions from different castes occurred during the period of Islamic dominance. The present study was aimed to to analyze the genetic structure, relationship and gene flow in five endogamous population groups, based on microsatellite loci and Alu insertion elements. **Methodology:** A total of 751 subjects, (154 Jat Sikh, 148 Mazhbi Sikh, 151 Brahmin, 148 Ramdasia and 150 Muslim) were recruited from the Northwest Punjab. Two kinds of markers; microsatellite and Alu insertion elements were used to analyze the genetic structure and relationships among five studied population groups. The genotypes thus obtained were subjected to extensive statistical analyses **Results:** Significant deviations from HWE were observed in almost all the populations and all the loci. In the present studied populations the genetic diversity (DST) was almost similar, except for Muslim population group. Significantly low coefficient of determination (FST) was observed among the five studied population groups. The genetic distance, genetic identity, AMOVA, structure, phylogenetic, principal component, multidimensional scaling and centroid analyses revealed similar results. In the entire analyses the Brahmin and the Mazhbi Sikh formed same cluster, while Jat Sikh and Ramdasia were placed close to their cluster, whereas, Muslim population group maintained the genetic distance from all the other population groups. **Conclusions:** Overall, a low level of genetic differentiation, observed in the studied population groups, especially, Jat Sikh, Mazhbi Sikh, Brahmin and Ramdasia, indicates that genetic drift might have been small or negligible in shaping the genetic structure of the Northwest Punjabi populations. The work provides the evidence that for populations living in same geographic contiguity, ancestry is the governing factor in genetic differentiation. Some amount of affinity with the European, Central Asian and Middle East populations could be attributed to many Islamic invasions from Middle East and Central Asia, specially Iran and Iraq, and the various religion conversions during the expansion of Islamic faith into the Indian sub-continent.

1469W

A new population structure analysis approach specifically designed for whole genome sequence data. *M. Robinson[1], W. S. W. Wong[2], B. D. Solomon[2,3], J. G. Vockley[2,3], I. Shmulevich[1], G. Glusman[1], J. E. Niederhuber[2].* 1) Institute for Systems Biology, Seattle, WA 98109, WA; 2) Inova Translational Medicine Institute, Inova Health System, 3300 Gallows Road, Falls Church, VA 22042; 3) Department of Pediatrics, Virginia Commonwealth University School of Medicine, 1201 E Marshall St, Richmond, VA 23298.

We developed a new approach to population structure analysis that recognizes differences between closely related populations better than standard methods involving principal component analysis (PCA) followed by k-means clustering. We find that PCA's normalization overweights common variants. We developed a novel method, Scaled Singular Value Decomposition (SVD), that weights variants equally and takes full advantage of the rarer variants now observable by whole-genome sequencing (WGS). As a result, population structure within continents is clearly resolved. Scaled SVD also facilitates analysis of admixture by locating variants as well as samples on the same principal components. Furthermore, k-means clustering ignores the ordered nature of principal components and the hierarchical structure of populations. We thus developed abcTree, a top-down hierarchical clustering algorithm specifically intended for population structure identification. abcTree employs adaptive Bonferroni correction to resist over-interpreting population structure. We evaluated our new methods (separately and in combination) on two large WGS cohorts: the 2504 samples (26 ascertained populations) of the 1000 Genomes Project, and an ethnically diverse cohort of 3483 founders from family trios in Northern Virginia, acquired by the Inova Translational Medicine Institute (ITMI, www. inova. org/itmi). The ITMI cohort is annotated with a rich set of multi-omic and phenotypic data, and metadata including self-reported country of birth. We analyzed multiple sets of variants, including common variants sampled from a genotyping array and variants of all frequencies randomly sampled from WGS data. In all combinations, Scaled SVD and abcTree separated samples better than PCA and k-means clustering. For the 1000 Genomes cohort, our method better recapitulated the annotated populations. For the ITMI cohort, the clusters corresponded better with self-reported country of birth. Precise identification of a patient's population of origin enables improved care via better diagnosis of inherited diseases and assessment of disease risks. It also increases power in genome-wide association studies (GWAS). We anticipate that these improvements in population structure analysis will be increasingly valuable as whole genome sequencing enters clinical practice. .

1470T

PCA-seq: Population Structure Inference with Rare Variants from Sequencing Data. *T. Thornton, J. Kirk.* Department of Biostatistics, University of Washington, Seattle, WA.

Advancements in high-throughput sequencing technologies have allowed for unprecedented insight into the vast amount of rare variants across the human genome. Recent studies have found that rare variants are likely to show greater geographic clustering than common variants, and also may provide a powerful resource to delineate fine-scale patterns of population structure. In recent years, principal components analysis (PCA) has been the prevailing approach for population structure inference (and correction) with common variants from high-density single nucleotide polymorphism (SNP) genotyping data. Widely used PCA-based methods, such as EIGENSTRAT (Price et al. , 2006),however, may not give reliable population structure inference when applied to sequencing data with rare variants due to the genetic relatedness matrix (GRM) that is used in standard PCA approaches becoming unstable with low frequency variants. Here, we propose PCA-seq, a new PCA-based approach for population structure inference with rare variants from sequencing data. PCA-seq uses a modified GRM in a PCA, where this GRM can appropriately incorporate low frequency variants, as well as common variants, allowing for robust inference on population structure. In simulation studies, we demonstrate that PCA-seq provides a substantial improvement over PCA with EIGENSTRAT for inference on population structure with rare variants. We further demonstrate the utility ofPCA-seq in an application to sequencing data from the 1000 Genomes Project wherePCA-seqidentifies different patterns of population structure for rare variants and common variants.

1471W

Population analysis of INDELS markers in a sample of individuals from Santander, Colombia. *N. A. Trujillo, C. I. Vargas, A. Castillo, A. M. Gil, A. L. Pico.* Santander, Universidad Industrial de Santander, Bucaramanga, Santander, Colombia.

Nowadays, the insertion-deletion markers (INDELs) have been the focus of several researches, because this type of polymorphism has interesting features like low mutation rates and multiplexing capability, which makes these markers more suitable to investigate population admixture events. In the same way, the study of genetic diversity in admixed populations as Colombia, it is quite important because it can reveal different aspects associated with the history of each individual, being important for different application fields such as clinical and forensic genetics. According to this, the aim of this work was to determine the genetic structure of a population group from the six regional provinces of Santander, Colombia through the typing of 38 autosomal markers (INDELs), expanding the knowledge of the genetic history in our country. We analyzed a total of 500 unrelated individuals, who were genotyped for 38 Indels by multiplex PCR. In order to increase the information related to these markers, we reported allele frequencies, and genetics parameters such as heterozygosity, Hardy Weinberg equilibrium, genetic distance and linkage disequilibrium were evaluated in the population. Additionally, we performed an analysis of individual ancestry proportions using reference samples from European, African, Asian and Hispanic populations by Fondevila M. *et al*, 2012. In the Hardy Weinberg test, we did not obtain statistically significant differences in 32 markers of 38 INDELs analyzed in the population of Santander (P > 0. 01). The populations represented in the six regional provinces presented a high genetic diversity intragroups, which ratified the high variability between individuals. Also, the results revealed a highly diverse genetic composition from all continental populations as it was expected. With this study, it is expected to generate new knowledge about the genetic composition of the population in Santander from each marker typified. In addition, the database reported by this project will be the first obtained for our state, and it will generate new knowledge to the genetic history of Colombia. Also, these markers will constitute a valuable tool for the analysis of important aspects such as ancestry and the possible association of some Indels with a greater predisposition to the development of complex genetic diseases in our population. **Key words:** Insertion-deletion polymorphisms, human identification, Santander, population structure, admixture.

1472T

Personalized genomic medicine challenges due to ancestry: disparities in the African Diaspora. *M. D. Kessler[1], M. Taub[2], R. Mathias[2], K. Barnes[2], L. Yerges-Armstrong[1], T. D. O'Connor[1], CAAPA Consortium.* 1) University of Maryland, Baltimore, MD; 2) Johns Hopkins University, Baltimore, MD.

Health disparities in personalized genomic medicine can create both economic and interpretive difficulties in treating underrepresented minorities. This is especially true in clinical genomics, where most previous studies have been focused on European ancestry individuals. Using 642 whole genome sequences from the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) project, which represents an essentially healthy cohort coming from many groups of the African Diaspora, we evaluate typical filters and databases used in personalized genomic medicine. The ANNOVAR program was used to annotate variants for ClinVar and Human Gene Mutation Database (HGMD) information, allele frequencies, protein function, and 11 in silico predictions of deleteriousness. We excluded any variant with an MAF > 0. 05 in ANY of the 1000 Genome super populations or in the Exome Sequencing Project, as well as variants that were called deleterious by at least 2 in silico predictors. We find that likely-pathogenic variants and variants of unknown significance have a strong positive correlation with African ancestry (r = 0. 990, p < 10-10 and r = 0. 995, p < 10-10 respectively). We also find that pathogenic pre-filtered variants in HGMD are strongly positively correlated with African ancestry (r = 0. 992, p < 10-10), but this correlation disappears with filtering. ClinVar does not have a significant pre-filtered correlation, but when filters are applied, a significant negative correlation appears (r = -0. 639, p = 0. 008). In addition, ClinVar's correlation has changed over time amidst monthly updates, with the most extreme switch happening between March and April of 2014 (r = 0. 733 to r = -0. 683). By starting with April's variant list, and selectively including/excluding SNPs differing between April and March depending on their correlation values and significance, we identified 68 SNPs as the major drivers of the correlation pattern. Using the most recent version of ClinVar, we calculated per gene correlations to identify a subset with the strongest signal, and note them as those that require particular care in clinical analysis due to their large ancestry-related bias. We do not find that these genes are enriched for Mendelian, GWAS catalogue, or X-linked genes. These results also have important financial implications in clinical genetics, with ancestry impacting both variant validation and clinician variant review.

1473W

Percent African admixture is associated with telomere length in a healthy adult population. *L. R. Yanek[1,2], K. R. Iyer[1], M. A. Taub[3], D. Vaidya[1,2,3], B. G. Kral[1,2], L. C. Becker[1,2], M. Armanios[1], D. M. Becker[1,2,3], R. A. Mathias[1,2,3].* 1) Medicine, Johns Hopkins University, Baltimore, MD; 2) GeneSTAR Research Program, Johns Hopkins University, Baltimore, MD; 3) Public Health, Johns Hopkins University, Baltimore, MD.

Africans and African Americans (AA) have been shown to have longer leukocyte telomere length (LTL) than persons with European ancestry (EA), but the extent to which this is a function of a finer scale of admixture (i. e. percent African ancestry) remains unknown. We examined whether the percentage of African admixture is associated with telomere length in 283 healthy subjects (average age = 42. 3, age range = 21-80 years, n female = 161; n AA = 127). Telomere length was calculated from whole genome sequence (WGS) data (>30x coverage on the Illumina HiSeq platform) for 7 contiguous repeats of the telomere motif (TTAGGG or CCCTAA) using the approach of Ding et al (2014). Admixture was estimated using 50,000 LD-pruned SNPs in STRUCTURE using three ancestral groups to calculate the % African and % European ancestry. Standard linear models were used to evaluate the association between telomere length and admixture estimates. In the AAs, average African ancestry was 80% (range: 45-98%) and average European ancestry was 20% (range: 2-55%). In the EAs average European ancestry was 99% (range: 72-99%) and average African ancestry was 1% (range: 0. 01-19%). As previously observed, an overall comparison between the two groups reveals longer telomere length in AAs than EAs (84364 vs 78560 kb, p=0. 0008). On a continuous scale, % of African ancestry was significantly correlated with telomere length (r=0. 22, p=0. 0002). Furthermore, in AAs, there is a strong association between % African ancestry and telomere length (each percent increase in African admixture was associated with an increase in telomere length of 6284 kb, p=0. 0036). Given minimal variation in the EAs, there is no observed association with % African ancestry and telomere length in this group (p=0. 1552). We confirm the prior observations that telomere length is different between AAs and EAs, and show here that within African Americans the % of African admixture is a significant predictor of telomere length. Future studies of racial differences in telomere length may need to account for differences in the proportion of African ancestry among subjects, particularly within African Americans.

1474T

Comparative analysis of transcriptome of different neurons in the human brain. *L. Yang, B. Su.* State Key Lab of Genetic Resource and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China.

Von Economo neurons(VENs) are specialized projection neurons with a characteristic spindle-shaped soma and thick basal and apical dendrite. VENs are distributed in only a few brain areas associated with intelligence and emotion. Humans have larger cell body and greater number of VENs than non-human primates. It was speculated that the emergence of VENs may be correlated with primate brain evolution and origin of human intelligence. However, the functional role of VENs remains elusive. Using laser capture microdissection and RNA sequencing, we analyzed the transcriptomes of VENs and three non-VEN types of neurons. We identified a set of VEN specific or preferential expression genes, which serves as important candidates for dissecting VENs' function and their potential contribution to the origin of human intelligence.

1475W

Sequence and Amplicon Diversity Among Chimpanzee Y Chromosomes. *M. T. Oetjens[1], F. Shen[1], Z. Zou[2], J. M. Kidd[1,2].* 1) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI, 48109, USA; 2) Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI, 48109, USA.

The male-specific region of the Y chromosome is a unique resource for tracking the history of populations due to the lack of recombination. Y chromosomes are also enriched for large, nearly identical repetitive regions known as amplicons which harbor many of the genes essential for spermatogenesis. In humans, copy-number variation of specific amplicons has been demonstrated. Sequence diversity on the unique segment of the MSY is greatly reduced compared to the autosomes, an observation consistent with the action of strong purifying selection. Similar studies of Y-chromosome diversity in other primate species have been limited. Here, we analyze 9 chimpanzee (representing three subspecies: Eastern, Western, and Nigerian-Cameroon) and two Bonobo male whole genome sequences (WGS) produced by the Great Apes Genome Project to assess nucleotide and ampliconic copy number diversity across the *pan* lineage. To assess nucleotide diversity, we applied a series of filters designed to remove regions of high-self identity such as the ampliconic tracts and heterochromatin from our callset. In total, we identified 23,946 SNVs (without human) across 4. 2 million callable sites. Comparisons with autosomal, X chromosome, and mitochondrial sequences from the same samples indicate that nucleotide diversity on the chimpanzee MSY is reduced relative to neutral expectations. Consistent with purifying selection, we observed a significant bias in synonymous over non-synonymous variants (p = 1. 60 x 10-5; Fisher's Exact Test) in single copy X-degenerate genes. We applied Quick-mer, a novel pipeline for paralog-specific copy-number analysis based on exact kmer matches to assess the ampliconic diversity among chimpanzees. We observe that the copy number of Y chromosome amplicons is variable amongst chimpanzees and bonobos and identify several lineage-specific patterns. The inferred copy number of testis-expressed genes varies by population, which may reflect underlying selection specific to chimpanzee subspecies and bonobos. For example, we find *TSPY* to range in copy number between 9. 2 (Western Chimpanzees) and 32. 1 (Bonobos) and *DAZ,* 1. 76 (Easterns) - 4. 11 (Westerns). Recurrent switchpoints of copy-number change along the ampliconic tracts across chimpanzee populations were also observed, which may be the result of localized genome instability or selective forces. These results support the hypothesis that strong purifying selection is inherent to the MSY in the *pan* and *homo* lineages.

1476T

Trends in DNA methylation with age replicate across diverse human populations. *S. Gopalan[1], O. Carja[2], M. Fagny[3,4], E. Patin[3,4], M. S. Kobor[5], H. Fraser[6], A. Froment[7,8,9], M. Feldman[6], L. Quintana-Murci[3,4], B. M. Henn[1].* 1) Department of Ecology & Evolution, Stony Brook University, Stony Brook, NY, United States; 2) Department of Biology, University of Pennsylvania, Philadelphia, PA, United States; 3) Department of Genomes and Genetics, Human Evolutionary Genetics, Institut Pasteur, 75015 Paris, France; 4) Centre National de la Recherche Scientifique, URA 3012, 75015 Paris, France; 5) Centre for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, BC, Canada; 6) Department of Biology, Stanford University, Stanford, CA, United States; 7) Institut de Recherche pour le Développement, 75006 Paris, France; 8) Muséum National d'Histoire Naturelle, 75005 Paris, France; 9) Centre National de la Recherche Scientifique, UMR 208, 75005 Paris, France.

The process of aging is associated with marked changes in patterns of genomic DNA methylation, including a global loss of methylation with increasing age. Several specific CpG sites also exhibit linear changes in their DNA methylation levels with chronological age. Futhermore, it has been demonstrated that methylation data from a relatively small number of CpG sites can accurately predict an individual's chronological age. So far, however, models of DNA methylation and aging have primarily focused on European and North American populations. The substantial genetic diversity and environmental variation that exists among human populations has not been explicitly accounted for, and its effects on the process of epigenetic aging process remain unknown. In this study, we investigate DNA methylation trends in two African hunter-gatherer populations from two tissue types, blood and saliva (n = 181): the Baka pygmies from the Central African forests of Cameroon and the Πιτι K-homani San from the Kalahari Desert in South Africa. We conducted an epigenome-wide association study on each population to identify CpG sites where methylation levels exhibit a significant and linear association with age. We confirm that a CpG site in the promoter of the gene ELOVL2 exhibits a significant and remarkably congruent relationship with aging when compared to previously studied European populations. We further identify significant hypomethylation with age at a CpG site in the promoter the D-aspartate oxidase gene (DDO). Expression of DDO has previously been found to increase with age in several species. We also test the performance of a previously published age prediction model to estimate chronological age in diverse populations and from different tissue types. We find that several age-related CpG sites replicate consistently across populations, demonstrating their utility as accurate biomarkers for age with potential applications in the fields of forensics and genomics. These results demonstrate how the differences in epigenome-wide trends across ethnic groups may provide important insights into the phenomenon of epigenetic aging in humans.

1477W

On estimating the shared genetic basis of complex phenotypes between populations. *B. C. Brown[1], N. Zaitlen[2].* 1) Department of Computer Science, UC Berkeley, Berkeley, CA; 2) Department of Medicine Lung Biology Center, UC San Francisco, San Francisco, CA.

Many phenotypes vary dramatically in their distributions around the world, and the extent to which these differences are driven by genetic rather than environmental factors is unknown. In this work we develop a framework for estimating the correlation of causal variant effect sizes between two populations (χ_g) from summary statistics. When χ_g=1 all causal variants between populations have the same effect size and when χ_g=-1, all have opposite effect sizes. This estimate therefore encapsulates the extent to which gene-gene, gene-environment, or other modifiers alter effect sizes between populations. Previous approaches to this problem (Yang et al 2013 NG) rely on linear mixed models and have three disadvantages. First, they require access to primary genotypes and phenotypes. Second, these methods are computationally expensive. Finally, they estimatethe correlation of allele-frequency normalized effect sizes(χ_z), or more intuitively, the correlation of the variance explained per causal variant. While our framework can also estimate χ_z, interpretation of this statistic is difficult; even when all causal SNPs have identical effect sizes variations in allele frequencies will give χ_z<1. Our method is based on maximizing the likelihood of $\chi_g$given the observed summary statistics and genotype covariances computed from a matched reference panel. We show that the joint distribution of the summary statistics is multivariate normal and that a dramatic speed-up can be had by using only the diagonal elements of the covariance matrix, resulting in an estimator that can be fit in time of order NM, where N is the reference panel sample size and M is the number of SNPs. Our framework can also estimate the total heritabilityof a phenotype in each population and the co-heritabilitybetween two phenotypes in one population, thus providing novel estimators of those quantities from summary statistics, and extends naturally to partitioning by functional category. We have shown via extensive simulation over real genotypes in two populations that our estimator of $\chi_g$is unbiased with reasonable standard error: with 50k individuals and 50k SNPs the standard error is 0. 090. Furthermore, our heritabilityestimator has half the standard error of LD score regression (Bulik-Sullivan 2015 NG): 0. 039 vs 0. 084 at the same number of SNPs and individuals, facilitating application to smaller regions. Finally, our method is robust so assuming only a small fraction p<1 of the SNPs are causal.

1478T

mQRF: A random forest model on meQTL detection by using chromatin states, haplotype structure and chromatin interaction information. *Y. Liu[1,2], M. Kellis[1,2].* 1) MIT, Cambridge, MA; 2) Broad Institute, Cambridge,MA.

mQRF: A random forest model on meQTL detection by using chromatin states, haplotype structure and chromatin interaction informationYaping Liu, Manolis KellisCSAIL in MIT, Broad Institute of MIT and Harvard, USADNA methylation, as an essential epigenetic mark, plays a critical role in gene regulation possibly via interactions with genetic variants, transcription factors and other epigenetic marks. Recent advances in Epigenome Wide Association Study (EWAS) provide the opportunities to study these interactions in order to fully understand the function of DNA methylation. One of the important steps to decipher the mechanism of genetic influences on DNA methylation is the detection of methylated quantitative loci (meQTL) using accurate and highly sensitive computational methods. The methods currently used to discover meQTL, however, are directly adapted from expression quantitative loci (eQTL) detection methods. Here, we explored the variety of relationships between SNPs, methylation and gene expression in different chromatin environments, which indicated that it is important and necessary to design a specific method for meQTL detection. meQTLFinder, a random forest model present here, incorporated biological information associated with DNA methylation and greatly increased the power on meQTL detection even in a limited sample size. mQRF quantitatively identified the relationship between genetic variation and DNA methylation, DNA methylation and enhancer, DNA methylation and gene expression in different chromatin states among different tissues by using Roadmap Epigenome project datasets. DNA methylation values from microarray or next generation sequencing studies were further normalized based on different chromatin states, sequence context and reference tissue type. A generalized linear model widely used in eQTL studies was applied to calculated association significance level between SNP variation and CpG methylation level. Later, normalized Hi-C contact frequency and haplotype structure between meQTL was also incorporated into the training set of mQRF's random forest model to increase the power of weak associated meQTL detection. Grant support from NIH-1U01HG007610-01.

1479W

The genomic and epigenomic properties of sexual dimorphism in human meiotic recombination. *C. Bhérer, C. L. Campbell, A. Auton.* Department of Genetics, Albert Einstein College of Medicine, Bronx, NY.

Sexual dimorphism in meiotic recombination is widespread across species. In humans, males tend to have significantly lower recombination rates than females over the majority of the genome, but the opposite is true close to the telomeres. These broad-scale differences have been known for decades, and yet little is known about the fine-scale differences between the sexes and the factors influencing them. Based on recombination events inferred from pedigree data consisting of more than 100,000 meioses, we have constructed sex-specific genetic maps at a previously unachievable resolution. Compared to previous maps, our maps display an improved correlation to LD-based maps and improved localization of events to hotspots of recombination. Using wavelet analysis, we characterize the sexual dimorphism in recombination across different scales and its relationship to genomic and epigenomic features. We show that the variation in female and male rates are highly correlated over a range of scales, and investigate the genomic features associated with the differences. Interestingly, male recombinogenic regions are more strongly associated with certain repeat elements and epigenetic marks, implying that male and female hotspots may be differently regulated by the hotspot localizing protein PRDM9. Finally, fine-scale analysis of genomic regions with a significantly higher rate in female than male and vice-versa further revealed a number of genomic features enriched in dimorphic regions compared to matched non-dimorphic regions. These findings provide potential clues to the origins of sexually dimorphic recombination rates.

1480T

Insights into meiotic recombination from characterization of non-crossover and crossover events in a mouse humanized at PRDM9. *R. Li, N. Altemose, E. Bitoun, AG. Hinch, E. Hatton, P. Donnelly, G. Zhang, B. Davies, SR. Myers.* Wellcome Trust Center for Human Genetics, Oxford University, Oxford, United Kingdom.

Meiotic recombination is initiated by programmed DNA double strand breaks (DSBs), repairing as crossovers (COs) or non-crossovers (NCOs). Although most DSBs resolve as NCOs, some basic properties of these events remain uncertain, because NCO tracts are often short and so difficult to detect. In mice and humans, many DSBs occur in hotspots positioned by the DNA-binding protein PRDM9, although >30% of human crossovers occur outside known hotspots. To study PRDM9 alleles from both species, we crossed and sequenced transgenic *M. m. domesticus* (B6) mice humanized at PRDM9, with wild-type *M. m. castaneus* (CAST) mice, which have a high SNP density relative to B6. We identified 81 NCO and 106 CO events in four F2 progeny, whose chromosomes are mosaics of the CAST and B6 genomes due to maternal and paternal recombination. All NCOs comprised simple unbroken tracts of SNPs, implying complex NCO is very unusual, despite recent human results. We estimate a mean NCO tract length of only 34 bases, with an approximately 10:1 NCO to CO ratio in meiosis. Remarkably, almost all NCOs and CO's overlap DMC1 or H3K4me3 ChIP-seq peaks from F1 males, which mark DSBs and PRDM9 binding sites respectively. Therefore few or no truly female-specific recombination hotspots exist and previous apparent "non-hotspot" recombination (e. g. in humans) may reflect variation at hotspots or PRDM9, or insufficient power. Most NCO events occurred very near to PRDM9 binding motifs: up to 69% actually contain the motif, and NCO conversions show a GC-bias. The CAST PRDM9 allele shows dominance over the human allele for both CO and NCO events but nevertheless humanization reprograms recombination sites. CAST-controlled NCOs preferentially occurred on the B6 chromosome, but human PRDM9-controlled NCOs did not explained by CAST PRDM9 binding motif erosion on the CAST background, and implying no strong meiotic control of the number of DSB's occurring on individual chromosomes exists. Surprisingly, two >8kb deletions (one genic) found in individual F2 mice are both near (<2kb) human hotspots. Although recombination proceeds normally, in general, at humanized hotspots, it may therefore be that introducing the human allele, which has evolved outside the mouse genome, increases the frequency of such rearrangements.

1481W

Estimating Rates, Tract Lengths and Contributing Factors to Non-Allelic Gene Conversion. *A. Harpak[1], X. Lan[2], J. Pritchard[1,2].* 1) Biology, Stanford University, Stanford, CA; 2) Genetics, Stanford University, Stanford, CA.

Gene conversion is the unidirectional transfer of genetic sequence from a "donor" to an "acceptor". In one of its modes, Non-Allelic Gene Conversion (NAGC), the donor and the acceptor are homologous sequences in different genomic loci. The parameters dominating NAGC are poorly understood. For example, estimations of the rate of NAGC empirical studies range across 7 orders of magnitude. This is despite the fact that interlocus gene conversion has been implicated as the cause of various genetic diseases and may be an important factor in the concerted evolution of many human gene families. Recent tandem gene duplications are of focal interest in studying NAGC as this process is contingent on high sequence similarity between donor and acceptor. In this work, we estimate the rates of NAGC in humans and other primates using ~700 pairs of recent tandem duplicates. For this purpose, we align families of duplicates across humans and 4 other primates. We then use the patterns of sequence divergence to estimate the rate and tract length distribution of NAGC with two independent methods. In the first method, we consider correlation in divergence patterns between two nucleotides in the same gene. The rates of mutation and NAGC determine the levels of divergence expected. The physical distance between the pair of nucleotides determines the probability that a NAGC event would cover both nucleotides, resetting divergence in the two concurrently. Using this model, we estimate the set of parameters that maximize the composite likelihood of the observed divergence patterns across all pairs of nucleotides in all gene families. In the second method, we make explicit use of the fact that a NAGC event would change the genealogy of the affected region across the gene family in primates. Genealogies may vary along the gene sequence, and the rates of transitions between genealogies are determined by the rate and length scale of NAGC. We construct a Hidden Markov Model where observed states represent observed divergence levels and hidden states represent underlying genealogies. This method also outputs a map of putative NAGC events across recent duplicates. We regress the density of NAGC events to various genomic features to reveal features putatively driving accelerated NAGC rates. This work contributes precise, cross validated estimates of the rate and length distribution of NAGC, suggests factors enhancing NAGC and measures its effect on human genetic variation.

1482T

Geometry of ancestral links: Pathogenic clues. *M. Jeanpierre.* Genetics, Hopital Cochin, Paris, France.

Inferences about negative selection can provide important information about the probability of a given variant being the cause of a genetic disease. This is particularly important for non-coding variants, for which it is difficult to carry out functional tests. Recent variants are generally considered much more likely to cause disease than very old. As rare variants may or may not be deleterious, frequency is only a very approximate indicator of pathogenicity. The shape of the genealogy for a particular variant may, theoretically, represent the effects of positive and negative selection more accurately than frequency alone. The algebraic description of haplotype patterns according to any topology of graphs representing genealogical trees is possible, but the resulting expressions rapidly become unbound, even for simple graphs with only a few branches, as the number of terms is approximately proportional to the factorial of the number of branches. We therefore need to break this problem down into simpler, more tractable questions. The distribution of nodes along the branches from a common ancestor is shaped by selection and could be represented graphically. The resulting representation of the true complexity would be easier to understand than that conveyed by bulky algebraic expressions. This approach is more efficient than a two-step strategy beginning with the construction of a large set of possible trees and followed by attempts to make sense of the resulting forest of probable trees. Isolated variants considered separately are rarely correlated with large effects, intervals containing clusters of variants may be more meaningful. Due to the many biases involved, length-biased sampling and related questions must be taken into account in the set of functions defining the distribution of haplotype sizes for any type of genealogy.

1483W

Whole genome sequences are required to fully resolve the linkage disequilibrium structure of human populations. *R. J. Pengelly, W. Tapper, A. Collins, S. Ennis.* University of Southampton, Southampton, United Kingdom.

Introduction: An understanding of linkage disequilibrium (LD) structure in the human genome underpins much of medical genetics, and provides a basis for disease gene mapping and investigation of biological mechanisms such as recombination and selection. Whole genome sequencing (WGS) provides the opportunity to determine LD structure at maximal resolution. **Methods**: We generated LD maps according to the Malécot-Morton model as implemented in LDMAP. We then compared maps produced utilising high depth whole-genome sequencing (WGS) data with LD maps produced utilising the array-based HapMap dataset, for representative European and African populations. **Results**: Maps derived from the WGS data achieve much greater resolution of LD structure (~ 5 fold), which translates to improved utility. The improved resolution allows the reduction in size of candidate regions flanking GWAS signals by 17%. The absence of ascertainment bias in WGS genotyping improves the population representativeness of the maps, and highlights the extent of uncaptured variation present in African populations using array genotyping methodologies. This uncaptured variation is a key reason for the poor reproducibility of GWAS results in African populations. Given the extended time to an effective population bottleneck for African populations, larger sample sizes and higher marker densities are required to fully resolve the LD structure. As such, we are currently generating maps from larger datasets, for release as a resource to the community. **Conclusions**: Array-based LD mapping is unrepresentative of actual LD structure, particularly in African populations. The Malécot-Morton model provides a highly efficient means of quantifying LD, and WGS LD maps will provide a rich resource for medical genetics, as well as for recombination biology. This work is funded under the UK Biotechnology and Biological Sciences Research Council (BBSRC) 'Sparking Impact' scheme.

1484T

Regulation of crossover placement in human meiotic recombination. *C. L. Campbell[1], N. A. Furlotte[2], N. Eriksson[2,3], D. Hinds[2], A. Auton[1].* 1) Department of Genetics, Albert Einstein College of Medicine, Bronx, New York 10461; 2) 23andMe Inc. , Mountain View, California 94043; 3) Present address: Coursera, Mountain View, California 94041.

Recombination is a key step in meiosis, serving multiple functions. Crossing over between homologous chromosomes shuffles existing genetic variation, while failure of the chromosomes to synapse and recombine can lead to nondisjunction and aneuploid gametes. Recent studies have shown that the placement of crossovers is under careful regulation, but despite this, displays tremendous variability between males and females in the location and frequency of recombination throughout the genome. To further address sex differences in recombination we conducted a pedigree analysis using 4,209 human families obtained through a research agreement with the personal genomics company 23andMe. From this, we are able to observe 18,302 meioses; 9,152 from females and 9150 from males. We find that males have a higher hotspot usage than females, a 4. 6% increase. We find that the number of crossovers rises sharply in older mothers, in support of previous reports of a maternal effect of crossover count. Finally, we estimate the level of crossover interference, a phenomenon that acts to ensure even spacing of recombination events. Males show a stronger overall strength of interference, while in females, a greater proportion of events escape this effect and this proportion increases with maternal age. This suggests that crossover placement appears to become increasingly deregulated in older mothers.

1485W

Inferring the parental haplotype source of germline-transmitted de novo duplications from population genotype data. *Y. Liu[1,2], S. Vattathil[1,2,3], L. Huang[2], X. Xiao[4], R. B. Scharpf[5], C. Huff[1,2], I. Ruczinski[6], S. Arur[7], T. H. Beaty[8], P. Scheet[1,2].* 1) Graduate School of Biomedical Sciences, University of Texas at Houston Health Science Center, Houston, TX, USA; 2) Dept. of Epidemiology, University of Texas MD Anderson Cancer Center, Houston , TX, USA; 3) Dept. of Genome Sciences, University of Washington School of Medicine, Seattle, WA, USA; 4) Center for Genomic Medicine and Department of Community and Family Medicine, Geisel School of Medicine at Dartmouth, Hanover, NH, USA; 5) Dept. of Oncology, Johns Hopkins Unviersity, Baltimore, MD, USA; 6) Dept. of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA; 7) Dept. of Genetics , University of Texas MD Anderson Cancer Center, Houston , TX, USA; 8) Dept. of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA.

De novo germline-transmitted duplications arise during gamete formation (possibly driven by meiotic recombination) and represent the genesis of a genetic polymorphism that has been associated with both complex and Mendelian forms of human disease. The formation of this type of duplications could be divided into two categories, i. e. , there are two possible sources for the extra copy of DNA in the offspring: the parent in which the duplication arose either (1) transmitted one copy from each homologous chromosome ("tri-allelic", arising from inter-chromosomal recombination), or (2) transmitted two identical copies from a single chromosome ("bi-allelic", arising from intra-chromosomal recombination). Here we apply a novel statistical model to two large population family-based genetic datasets to estimate the relative contributions of these two processes to the origins of copy number duplications. One dataset contains 2078 trios from a study of craniofacial abnormalities and the other contains 1346 trios from a study of autism. In our procedure, we first use trios to identify *de novo* duplications in offspring. We then check for consistency among plausible combinations of transmitted haplotypes and observed SNP array data (genotypes implied by the "B allele" frequencies, BAFs) for the offspring (duplication carrier), assuming empirically informed distributions for the BAFs in regions of duplication. Our preliminary results from both datasets indicate a tendency for duplications to derive from the same parental chromosome (bi-allelic), however the proportions differ between data sets (55%, 89%) warranting further investigation. Interestingly, we found examples in both datasets with evidence of a crossover event within the de novo duplication, thus part of the duplicated region manifests as tri-allelic and part as bi-allelic. Our analysis present the first population-level investigation into a phenomenon critical to the generation of this specific form of genetic polymorphism.

1486T

Pleiotropic constraint in the Human Transcription Factor Network. *K. Chesmore, J. Bartlett, C. Cheng, S. Williams.* Genetics Dept, Dartmouth College, Hanover, NH.

It has long been argued that pleiotropy can constrain gene evolution. We can now investigate the pleiotropic mechanisms that drive such constraints because of the recent expansion of large-scale molecular data. No class of genes has been more thoroughly characterized than transcription factors due to the efforts of the ENCODE consortium. Here we examine constraints on the human transcription factor network as a function of the number of partners with which each protein and protein domain interacts. We show not only increased evolutionary constraint with an increasing number of molecular partners, but also as a product of each domain's role. Additionally, we show that as the number of protein-protein interacting domains increases for a given transcription factor, the constraint on each domain decreases. In summary, we demonstrate that not only does pleiotropy, as defined by the number of molecular partners, constrain protein evolution in general, but that the constraints differ by functional role.

1487W

The birth of a human-specific neural gene by incomplete duplication and gene fusion. *M. L. Dougherty[1], X. Nuttle[1], B. Nelson[1], J. Huddleston[1,2], O. Penn[1], M. Y. Dennis[1], L. Harshman[1], K. Penewit[1], F. Antonacci[3], M. Malig[1], K. Munson[1], M. Duyzend[1], E. E. Eichler[1,2].* 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Howard Hughes Medical Institute, Seattle, WA; 3) Université degli Studi di Bari, Bari, Italy.

The human-specific gene *HYDIN2* was generated by an interchromosomal duplication event that copied 360 kbp of the large ciliary protein-encoding gene *HYDIN* (hydrocephalus-inducing) from chromosome 16q22. 2 to 1q21. 1, a hotspot for genomic rearrangements. The segmental duplication, however, included neither the promoter and first five exons nor the final two exons of the original progenitor gene, leading to speculation that the duplicate may not produce a functional protein product. We coupled RT-PCR with long-read (SMRT) sequencing of cDNA to characterize transcription at this locus. 5' RACE indicated a transcription start site outside of the duplication, and by targeting the duplication boundaries, we observed fusion products on both the 5' and 3' ends of the segmental duplication as well as extensive splicing diversity. A novel promoter and first exon were acquired by an adjacent duplication block and fusion transcripts with an *NBPF* core duplicon are observed downstream. This juxtaposition of duplications leads to a new potential translation start site and long open-reading frame consistent with the emergence of a novel human-specific gene. Expression profiling by RT-PCR and RNA-seq confirmed a neural-specific expression pattern distinct from the ancestral locus. Interestingly, we also find evidence of accelerated amino-acid replacements for both *HYDIN* and *HYDIN2* indicative of potential subfunctionalization of both copies after duplication. By comparing the 360 kbp duplication block to homologous primate sequence, we were able to date the duplication event to 3. 1 million years ago (+/- 0. 5 mya). Despite its recent origin, the duplication is nearly fixed for copy number in the human species (99. 9%). We surveyed a diversity panel of 2134 human genomes and found only one individual with a diploid copy number <4 for both paralogs. This is especially striking in light of the susceptibility of the 1q21. 1 region to extensive deletion and rearrangement. These data suggest a model of rapid gene innovation by gene fusion of adjacent segmental duplications, altered tissue expression, and subfunctionalization of *HYDIN* and *HYDIN2* during the transition from australopithecines to the genus *Homo*.

1488T

An unsteady molecular clock in primates. *P. Moorjani[1,2,5], CE. Amorim[1,5], P. Arndt[3], M. Przeworski[1,4].* 1) Department of Biological Sciences, Columbia University, New York, NY, USA; 2) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Max Planck Institute for Molecular Genetics, Berlin, Germany; 4) Department of Systems Biology, Columbia University, New York, NY, USA; 5) Contributed equally.

Evolutionary events, such as the human-chimpanzee speciation or human migration Out-of-Africa, are usually dated under the assumption of a "molecular clock"—a constant rate of substitutions per unit time. Although this assumption is central to our understanding of the chronology of human evolution, its validity remains unclear. Among mammals, for example, it is well known that there exists substantial variation in substitution rates, potentially due to differences in their generation times or other life-history traits. To characterize the effect in primates, we analyzed genome-wide autosomal data from 12 primates. Focusing on putatively neutral sites and accounting for possible effects of methylation and biased gene conversion, we found that substitution rates in Old World Monkeys are ~30-40% higher than in apes. Rates vary even among apes: the gorilla branch ~6-10% longer than the human one. Importantly, the extent of rate variation differs among mutation types, with both transitions and transversions at CpG sites exhibiting a more clock-like behavior. As a result, not only the total rate, but the mutational spectrum is evolving among primates. This indicates that the molecular clock in primates is unstable and suggests that events in primate evolution are most reliably dated using CpG sites. We present updated divergence times in light of these considerations.

1489W

Long-term survival of duplicate genes despite absence of subfunctionalized expression. *X. Lan[1], J. K. Pritchard[1,2,3].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Biology, Stanford University, Stanford, CA; 3) Howard Hughes Medical Institute, Stanford University, Stanford, CA.

Gene duplication is a fundamental process in genome evolution. However, young duplicatesare frequently degraded into pseudogenes by loss-of-function mutations. One standardmodel proposes that the main path for duplicate genes to avoid mutational destructionis by rapidly evolving subfunctionalized expression profiles. We examined this hypothesisusing RNA-seq data from 46 human tissues. Surprisingly, we find that sub- or neofunctionalizationof expression evolves very slowly, and is rare among duplications that arosewithin the placental mammals. Most mammalian duplicates are located in tandem andhave highly correlated expression profiles, likely due to shared regulation, thus impedingsubfunctionalization. Moreover, we also find that a large fraction of duplicate gene pairsexhibit a striking asymmetric pattern in which one gene has consistently higher expression. These asymmetrically expressed duplicates (AEDs) may persist for tens of millionsof years, even though the lower-expressed copies tend to evolve under reduced selectiveconstraint and are associated with fewer human diseases than their duplicate partners. Wesuggest that dosage-sharing of expression, rather than subfunctionalization, is more likelyto be the initial factor enabling survival of duplicate gene pairs.

1490T

The effect of life history on the rate of the molecular clock and implications for great apes. *G. Amster, G. Sella.* Columbia University, NY.

Estimates of species split times in the great ape phylogeny rely on the molecular clock. The estimates have recently been nearly doubled, because pedigree based estimates of the mutation rate approximately halved previous estimates based on fossil calibration. These revisions, however, rely on assuming a constant mutation rate per generation, which appears inconsistent with our current understanding of mutation in hominins, and might also be inconsistent with the fossil record. Here we show that less naïve models of the molecular clock that account forlifehistoryfactors and our current knowledge about their effects on mutation rates suggest a potentially large increase in estimates of split times. Notably, moderate differences in generation time between the sexes, changes in male puberty age, and to a lesser extent, changes in the rate of germ cell division during spermatogenesis, would have had a substantial effect on the rate of substitutions in great apes. We also show thatlifehistorycould help explain the observed ratios of X to autosomes divergence in great apes, and suggest modifying inferences based on this ratios in other taxa.

1491W

Linked selection dominates the genomic landscape of Great Apes. *A. Woerner[1], K. Veeramah[2], M. Hammer[3].* 1) Genetics GIDP, University of Arizona, Tucson, AZ; 2) Dept of Ecology and Evolution, Stony Brook, NY; 3) Arizona Research Laboratories, Division of Biotechnology, University of Arizona, Tucson, AZ.

Early studies that have attempted to utilize putatively neutral loci in the genome for demographic inference have focused on pseudogenes, introns, and repetitive elements such as Alus. More recent studies have attempted to reduce the impact of directional and purifying selection as well as background selection and genetic draft by finding loci that are either in regions of low gene density and high recombination rate, or are at a maximal distance from genes in genetic units. However nucleotide diversity not only reduced near genes across great apes, but also near noncoding conserved elements. Because ~5% of primate genomes are composed of evolutionarily conserved elements, with ~2/3 of these elements being noncoding, skews in diversity in regions in linkage with such elements has large effects on genomic patterns of nucleotide variation. Here, we show that the combination of multiple conserved noncoding elements has more of an influence on nucleotide diversity than the distance to the nearest gene. Using an approach that models the combined effects of linkage to coding and noncoding elements, we predict levels of polymorphism on both the autosomes and the X chromosome for humans, chimpanzees, bonobos and gorillas. Based on our predictions we identify loci that are “effectively neutral” across various ape taxa. Using the relationship between diversity and the local recombination rate as a proxy for the effects of directional selection on neutral sites, we show that unlike other loci in the genome, our effectively neutral loci show no significant correlation between the local recombination rate and diversity. We conclude that “neutral” sequence is extremely rare in the genome. Finally, we use these neutral regions to estimate the relative effective population size of the X chromosome vs. the autosomes in African and non-African human populations, and in gorillas, chimpanzees and bonobos.

1492T

The genetic handicap principle: a severely deleterious mutation can be tolerated if the genome-wide mutation load is sufficiently low. *K. Popadin[1,2], S. Peischl[3], R. Sailani[4], A. Letourneau[1], F. Santoni[1], M. Garieri[1], S. Nikolaev[1], D. Meyer[5], L. Excoffier[3], S. E. Antonarakis[1].* 1) University of Geneva Medical School, Department of Genetic Medicine and Development, Geneva, Switzerland; 2) University of Lausanne, Center for Integrative Genomics, Genopode Building, Lausanne, Switzerland; 3) Institute of Ecology and Evolution, Computational and Molecular Population Genetics (CMPG), Bern, Switzerland; 4) Stanford University, Stanford School of Medicine, Stanford, CA; 5) University of Sao Paulo, Department of Genetics and Evolutionary Biology, Sao Paulo, Brazil.

Every human genome harbors hundreds of slightly-deleterious mutations. Even though most of these mutations individually have a small effect, their cumulative effect (mutation load) could substantially reduce the fitness of an individual. Here we investigate an interaction between a severely deleterious mutation and the mutation load, formulating the genetic handicap principle: an individual bearing a severely deleterious variant (i. e. a genetic handicap) is viable only if the genome-wide mutation load is sufficiently low. We develop a population-genetic model, which predicts that live-born individuals carrying a handicap mutation (causing, for example, 50-90% probability of miscarriage) show a reduction (5-20%) in their number of slightly-deleterious mutations as compared to controls.

To test this prediction, we used data on trisomy of chromosome 21 (T21), a frequent chromosomal abnormality in humans which is associated with high miscarriage rates. Three lines of evidence support the view that live-born T21 individuals have passed through the strong embryonic selection and thus have decreased load of slightly-deleterious mutations.

First, on chromosome 21 there is a deficit of homozygotes for rare derived alleles in the T21 cohort (N=338), suggesting a selection against homozygotes for deleterious variants.

Second, the T21 cohort on chromosome 21 shows an excess of loss-of-expression (LOE) regulatory variants, which are considered to be compensatory under the condition of trisomy, and deficit of gain-of-expression (GOE) regulatory variants, which are considered to be deleterious under the condition of trisomy, consistent with embryonic selection in favor of LOE and against GOE variants.

Third, assuming that each gene has an optimal expression level and all deviations from the optimum are deleterious, we compared inter-individual variation in fibroblast expression level between live-born T21 individuals (N=8) and controls (N=8) and revealed whole-transcriptome decreased variation in expression level in T21 cohort, driven mainly by dosage-sensitive genes, considered to be the most important for the health of an embryo. This can reflect a potential selection against T21 fetuses with non-optimal pattern of expression.

We conclude that the negative fitness consequences of severe mutations such as trisomy can be partially compensated by a reduced genome-wide load of slightly-deleterious mutations.

1493W

Prevalence of Ebola viral entry resistance in a diverse population. *P. F. Cherukuri, T. Vilboux, P. Kothiyal, A. Black, G. Eley, K. C. Huddleston, R. K. Iyer, B. D. Solomon, J. G. Vockley, J. E. Niederhuber.* Inova Translational Medicine Institute, Inova Health System, Falls Church, VA. USA.

Whole-genome sequencing (WGS) is fast evolving into a population genetics tool to estimate effect of sequence variants on human health and fitness. However, predicting the role of variants that confer protection remains challenging. Given this, we hypothesized that use of large population datasets and integrative-omics data would delineate role of protective variants. To test this hypothesis, we used variants ascertained by WGS in specific genes that confer resistance to Ebola viral entry and pathogenesis. Methods: We performed trio-based WGS on individuals enrolled in various clinical protocols, primarily as part of a broad study on use of population-based genomic sequencing. As a pilot study we performed WGS on 3,536 ethnically diverse individuals who are prospectively followed as part of a clinical/genomic study. WGS was performed on Illumina platform. Each individual's ancestry was defined by pairwise Pearson Correlation Coefficient based hierarchical clustering using 1000 Genomes Project defined population genotype markers and matching self-identified ethnicities. Variants in genes (e. g. *NPC1*) known to play key roles in viral attachment and entry into human cells were evaluated for variant burden. Additional genes within the viral entry pathway with no clear known role in uptake of viruses were used as controls. Results were compared to frequencies in public exome databases. Additionally, RNASeq was performed on 237 individuals to test for gene-expression patterns for loss–of-function variants. Results: We screened the WGS data for heterozygous variants with predicted large functional effect in the gene set. Specifically, we prioritized variants that resulted in evolutionarily unanticipated missense, frame-shift, splice-site altering or protein truncating mutations. We identified 69 novel variants (not in dbSNP), of which 7 were missense and 4 protein truncating in *NPC1* in heterozygous state, compared to 3 missense variants in control gene dataset. RNASeq allele-specific measurement of truncated products showed skew in distribution of reads at heterozygous sites compared to normal controls, suggesting non-sense mediated decay of truncated message. From this, we predicted that 0. 11 - 0. 37 % of our diverse population would have resistance to Ebola virus based on the identified variants. Though there are obstacles in interpreting WGS data, our pilot study on viral entry resistance serves as an example of role of integrative-omics in population genetics.

1494T

Chromosome-wide scan to identify rapidly mutating Y-STRs. *T. Willems[1,2], M. Gymrek[1,2,3,4,5], Y. Erlich[2].* 1) Computational and Systems Biology Program, MIT, Cambridge, MA USA; 2) New York Genome Center, New York, NY USA; 3) Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA USA; 4) Broad Institute of MIT and Harvard, Cambridge, MA USA; 5) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA USA.

Precise estimates of the time to the most recent common ancestor play a pivotal role in various Y chromosome applications. These include tracing patrilineal surnames and anthropological statuses and discriminating between individuals for forensic analyses. Current tests rely on a few dozen Y-chromosome short tandem repeats (Y-STRs) due to their relatively high mutation rates. However, these tests would be substantially improved by incorporating additional rapidly mutating Y-STRs. Unfortunately, identification of these loci is encumbered by the high cost and low throughput of the traditional method for estimating Y-STR mutation rates, which involves collecting and genotyping tens of thousands of father-son pairs using capillary electrophoresis. To address these limitations, we developed a novel mutation rate estimation method that leverages the advent of population-scale whole-genome sequencing datasets and bioinformatics tools capable of genotyping STRs. Our approach constructs a phylogeny using Y-SNP genotypes, overlays WGS-based STR genotypes and utilizes message passing and numerical optimization to converge to the maximum likelihood mutation model. In addition, our approach has the added benefit of considering a range of mutation models that allow for multistep mutations and length constraints. Using in-depth simulations, we confirmed that our method correctly accounts for PCR stutter artifacts and obtains unbiased mutation rate estimates across all considered mutation models. Application of the method to data from the 1000 Genomes project resulted in estimates for nearly 350 Y-STRs, representing the largest such set to date. Comparison of these estimates with those from father-son studies indicated a high degree of concordance for loci that have been previously characterized. In addition, we identified over 100 previously uncharacterized Y-STRs with mutation rates greater than 10-3. 5. Altogether, our study provides a broadly applicable method for estimating Y-STR mutation rates from whole-genome sequencing cohorts and vastly expands the number of identified Y-STRs with high discriminative power.

1495W

The Qatari Genome: A Population-Specific Tool to Facilitate Precision Medicine in the Middle East. *KA Fakhro[1,2], MR Staudt[3], F Agosto-Perez[3], TL Vincent[3], A Robay[1], JA Malek[1], R Badii[4], A Al-Nabet Al-Marri[4], C Abi Khalil[1], D Stadler[5], M Zirie[6], A Jayyousi[6], J Salit[3], JG Mezey[3, 7], RG Crystal[3], JL Rodriguez-Flores[3].* 1) Weill Cornell Medical College in Qatar, Doha, Qatar; 2) Sidra Medical and Research Center, Doha, Qatar; 3) Department of Genetic Medicine, Weill Cornell Medical College, New York, New York; 4) Laboratory Medicine and Pathology, Hamad Medical Corporation, Doha, Qatar; 5) Weill Cornell Medical College in Qatar, Doha, Qatar; 6) Department of Medicine, Hamad Medical Corporation Doha, Qatar; 7) Department of Biological Statistics and Computational Biology, Cornell University, Ithaca, NY.

Precision medicine will be driven by the sequencing of humans from a wide geographical representation, for which accurate variant calling and personalized interpretation will be critical to deciphering relevance to individual health. In this study, the value of having a geographically appropriate reference genome is shown for both calling accuracy and bioinformatic analysis, by building a population-specific reference genome (QTR1) for the indigenous Arab population of Qatar. This reference was constructed by leveraging allele frequency data from exome and genome sequencing of 1,158 Qataris (0. 4% of population), and tested by aligning the exome and genome of an "n+1" individual sequenced on four platforms to both the standard GRCh37 reference genome and QTR1. Using a population-centric reference reduced the number of variants called in exomes by 14% and genomes by 23%, dramatically improving calling efficiency and cost of storage and analysis as sequencing cohort sizes grow. Furthermore, when analyzing variants linked to disease or pharmacogenetic phenotypes in the entire Qatari cohort, we observed that 295 of 2,330 (12. 6%) were in fact the major allele in Qataris, and may have been incorrectly classified as being deleterious alleles using a standard genome interpretation workflow based on GRCh37. We also refined automated annotation with manual curation to identify 198 Mendelian conditions with known disease-causing mutations observed in Qataris. By comparison to three large publically available exome databases, 26 conditions appear with causative allele frequencies highest in Qataris, including diseases with known elevated incidence in this population such as homocystinuria (CBS [MIM 613381]) and cystic fibrosis (CFTR [MIM 602421]), or favorable conditions such as hyperalphalipoproteinemia (APOC3 [MIM 107720]). Consistent with high consanguinity rates in Qatar, individuals homozygous for rare recessive mutations are observed despite the small sample size, including diseases such as ectodermal dysplasia (KRT85 [MIM 602767]), retinitis pigmentosa (RP1 [MIM 603937]), and arterial tortuosity (SLC2A10 [MIM 606145]). Notably, only a minority of all discovered mutations presented here are currently on the Qatari national newborn screening panels. We will provide this allele frequency data and the QTR1 reference sequence as open access tools for use in precision medicine in Qatar and in ethnically similar populations in the Arabian Peninsula.

1496T

Association of polymorphism *ADIPOQ* rs2241766 (45 T/G) and *ADIPOQ* rs1501299 (276 G/T) with obesity in breast cancer patients. *A. Mendez-Hernandez[1], R. PEREZ-MORALES[2], M. GALLEGOS-ARREOLA[3], J. ESPINOSA-FEMATT[1].* 1) DEPARTMENT OF RESEARCH, UNIVERSIDAD JUAREZ DEL ESTADO DE DURANGO, GOMEZ PALACIO, DGO. , MEXICO; 2) DEPARTMENT OF MOLECULAR BIOLOGY UNIVERSIDAD JUÁREZ DEL ESTADO DE DURANGO, Gómez PALACIO, DGO; 3) LABORATORY OF MOLECULAR GENETICS, DIVISION OF MOLECULAR MEDICINE, CIBO, IMSS, GUADALAJARA, JAL.

Background: Obesity is both a known risk factor for development of breast cancer and a negative prognostic factor among women with breast. Adipokines as adiponectin are produced in adipose tissue including mammary fat. Polymorphisms in the gene encoding adiponectin (*ADIPOQ*) have been associated with a decrease in the levels of this hormone, promotion and evolution of breast cancer. *ADIPOQ* rs1501299 genotype G/G and rs2241766 T/T are associated with decreased adiponectin levels, obesity, insulin resistance and breast cancer. Therefore our aim is to evaluate the association of polymorphisms *ADIPOQ* rs2241766 (45 T/G) and rs1501299 (276 G/T) with obesity in patients with breast cancer. Methods: DNA genomic of 301 samples (UMAE, HOSPITAL OBSTETRICS AND GYNECOLOGY, CMNO, IMSS) was included in the study. The *ADIPOQ* rs2241766 (45 T/G) and rs1501299 (276 G/T) genotyping was determinate by Taqman® probe and allelic discriminated was determinate by StepOne software v. 2. , of Applied Biosystems. The association was determinate by odds ratio (OR). Results: the genotype *ADIPOQ* rs2241766 T/G was associated with overweight and obesity [OR = 2. 53 (95% C. I. = 1. 13 - 5. 64), p = 0. 01] in patients with breast cancer, while the adipoq rs2241766 G/G genotype was found as a protective factor to overweight and obesity [OR = 0. 48 (95% I. C. = 0. 23 to 0. 98), p = 0. 04], no association was found with polymorphism rs1501299 (276G/T). Conclusion: the *ADIPOQ* rs2241766 (45 T/G) may be an indicator of risk for overweight and obesity and explain the possible mechanism of obesity in the progression of breast cancer in our population.

1497W

Exploring variation in mosaicism and the mutation rate over time in the mammalian genome. *S. J. Lindsay, M. E Hurles.* Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, United Kingdom.

It is currently challenging to study heterogeneity in gametogenesis and mutation rate variation at different stages of the reproductive cycle explicitly in humans. Therefore, we have designed and carried out an experiment in order to investigate these attributes in laboratory mice. We surveyed the variance in contemporaneous gametes and the influence of parental age at set time points on the mutation rate by generating a tissue bank from the pups in consecutive litters between two reciprocal crosses (129S χ(B6 χ| B6 χ(129S χ) over their fertile lifespan. By whole genome sequencing, *denovo* mutation discovery and validation from multiple pups from the first and last litters produced by the same breeding pairs, we were able to investigate the age effect in the maternal and paternal lines. We found that laboratory mice display a broadly similar trend in paternal age effect as humans, although there is a degree of variance within each litter. Through further deep sequencing of validated *denovo* mutations in different tissues and multiple pups from the same litter, as well as their parents, we were also able to study germline and somatic mosiacism in the parents and offspring, offering insights into rate mutation rate variation between individuals over time.

1498T

A genome-wide association study revealed an association between HLA locus and positivity of rheumatoid factor in Japanese patients with rheumatoid arthritis. *T. Hosozawa, K. Ikari, A. Taniguchi, H. Yamanaka, S. Momohara.* Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Shinjuku, Japan.

Background. Many studies show that numerous genetic factors are associated with the onset of rheumatoid arthritis (RA). While the diagnostic utility and specificity of anti-cyclic citrullinated peptide antibodies for RA are the focus of many recent studies, we concentrated on another important diagnostic marker associated with RA disease activity and joint destruction, rheumatoid factor (RF). RF is an autoimmune antibodyrecognizing the Fc fragment of IgG, which has been used for more than half a century to diagnose RA. The genetic background of RF-positive RA patients is still not clear. The purpose of this study was to identify genes related to RF using a genome-wide association study. **Methods**. The present study was part of a Japanese RA cohort project established by the Institute of Rheumatology, Rheumatoid Arthritis (IORRA). DNA samples of 2423 Japanese RA patients were obtained from the IORRA DNA collection. SNP genotyping was performed using the Illumina HumanOmni2. 5 Exome-8 BeadChip. Quality control of the genotype data and analyses were performed using PLINK2. We then compared RF-positive and RF-negative for each SNP using the chi-square test. **Results**. All SNPs that satisfied the genome-wide significance level of $p<5. 0\times10\text{-}8$ were located in the Human leukocyte antigen (HLA) region of chromosome 6, which includes the HLA-DRB1 gene. Our result suggests that HLA polymorphisms are associated with RF-positive status in RA patients. Further studies using HLA imputation tools are required to confirm these results.

1499W

Contributions of short tandem repeats to the phenotypic variation and heritability of a simulated gene expression trait. *E. C. Glassberg[1], Y. Erlich[2,3], M. Gymrek[3,4,5,6], M. Feldman[1], J. Pritchard[1,7,8].* 1) Department of Biology, Stanford University, Stanford CA; 2) Department of Computer Science, Fu Foundation School of Engineering, Columbia University, New York NY USA; 3) New York Genome Center, New York, NY USA; 4) Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA USA; 5) Broad Institute of MIT and Harvard, Cambridge, MA USA; 6) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston MA USA; 7) Department of Genetics, Stanford University, Stanford CA USA; 8) Howard Hughes Medical Institute, Chevy Chase MD USA.

An important aim of statistical genetics is to determine the genetic architecture of complex traits; however, for many complex traits studied to date, a large proportion of the heritability appears to be missing. One explanation for this missing heritability is that variants that are poorly tagged by the common SNPs used in current association studies, such as short tandem repeats (STRs), make significant contributions to phenotypic variability. STRs are highly polymorphic in humans, many STRs are known to have functional consequences, and linkage between STRs and common SNPs is estimated to be lower than between pairs of SNPs at comparable physical distances. Further, STR length variation has recently been shown to influence the expression of genes in *cis* (Gymrek et al 2015). As a result, STRs may play an important role in the determination of human complex traits. Further, due to the high mutation rate and multi-allelic nature of STR variation, current association studies are underpowered to detect the effects of STR alleles on complex traits. Therefore, causal STR variants may contribute to missing heritability. However, the different mutational profiles of SNPs and STRs affect both the expected contributions of each class of variant to phenotypic variation and our power to detect associations between variable loci and a trait of interest. Here, we compare a simple, two-phase model (from diRienzo et al 1994) of mutation and phenotypic effect of an STR locus on a simulated gene expression trait to an analogous, infinite sites model for SNPs, combined with a model for selection on gene expression. Simulating Wright-Fisher sampling, we explore (i) the amount of phenotypic variation produced by SNP and STR loci at various levels of mutation, phenotypic effect, and selection, and (ii) the ease with which causal associations are detected under various parameter sets. This work provides insight into the dynamics of selection on functional STR loci and the mutational and selection regimes in which STR variants contribute to missing heritability.

1500T

Human Mitochondrial Indel Discovery via *de novo* Assembly and its Applications. *M. Su[1], N. Stoler[2], B. Arbeithuber[3], B. Rebolledo-Jaramillo[2], W. Guiblet[4], K. D. Makova[1], A. Nekrutenko[2].* 1) Department of Biology, Penn State University, University Park, Pennsylvania, USA; 2) Department of Biochemistry and Molecular Biology, Penn State University, University Park, Pennsylvania, USA; 3) Institute of Biophysics, Johannes Kepler University Linz, Linz, Austria; 4) Integrative Biosciences, BioInformatics and Genomics Program, The Huck Institutes of the Life Sciences, Penn State University, University Park, Pennsylvania, USA.

Mutations in mitochondrial DNA (mtDNA) in human contribute to a number of severe genetic diseases, including LHON (Leber's hereditary optic neuropathy) and MELAS (mitochondrial encephalomyopathy lactic acidosis and stroke-like episodes), which are transmitted and observed in maternal lineages. The mutations in mtDNA are manifested as not only point mutations but also insertions/deletions (indels). Due to insufficient DNA repair and a toxic environment rich in radical oxygen species, the mutation rate of mtDNA is estimated to be ~10 times higher than that of nuclear genome. The presence of more than one mtDNA variant within an individual is called heteroplasmy. Analyzing 39 healthy mother-child pairs from central Pennsylvania, we have recently shown that on average a person carries 1-2 heteroplasmic point mutations in the mtDNA of blood or buccal cells, and one in eight mothers carries a disease-associated mtDNA mutation as a heteroplasmy. These findings drew our attention to the identification of indels in mtDNA, where they are common in low complexity regions (i. e. repetitive sequences). However, detecting low frequency indels is a challenge as they are often obscured by read alignment artifacts. In this study, we developed a new indel discovery method using *de novo* assembly. Experimental validation of indels was performed using droplet digital PCR and duplex sequencing. We discuss the identification of indel sites, the mutation rate of indels, the biological importance of indel mutations, and the applications of indel discovery analysis.

1501W

An overview of the *HLA-C* promoter and 3'untranslated region variability and haplotype structure. *E. C. Castelli[1], I. O. P. Porto[1], T. H. A. Lima[1], J. Ramalho[1], R. V. Buttura[1], M. A. Paz[1], L. P. Felício[2], C. T. Mendes-Junior[3], L. C. Veiga-Castelli[4], E. A. Donadi[4].* 1) School of Medicine, UNESP - Univ. Estadual Paulista, Botucatu, São Paulo, Brazil; 2) Federal University of Goiás (UFG), Goiânia, Goiás , Brazil; 3) Departamento de Química, Universidade de São Paulo (FFCLRP-USP), Ribeirão Preto, São Paulo, Brazil; 4) Division of Clinical Immunology, School of Medicine of Ribeirão Preto, University of São Paulo - USP, Ribeirão Preto, São Paulo, Brazil.

The *HLA-C* gene seems to be the most unusual among the classical class I genes of the Major Histocompatibility Complex (MHC). Besides presenting a reduced variability and decreased expression when compared to *HLA-A* and *HLA-B*, its encoded molecule is able to interact with a series of receptors of the KIR family (Killer cell Immunoglobulin-like Receptor) responsible for inhibitory signals to Natural Killer cells and it is the only classical MHC gene expressed at the cytotrophoblast during pregnancy. *HLA-C* is the least explored among the MHC classical class I genes, in part because histocompatibility laboratories generally evaluate only *HLA-A* and *HLA-B*, and usually only the segment encoding the peptide-binding site is explored. Here we explore the variability and haplotype structure of the *HLA-C* promoter segment within 600 bases before ATG and the entire 3' untranslated region (UTR), from the stop codon to the end of the 3'UTR segment (616 bases), considering 100 individuals from the Southeast Brazil (São Paulo State). For this purpose, a next-generation sequencing strategy was developed coupled with bioinformatics approaches to enhance the mapping quality of HLA-related sequences. Considering the promoter segment, 34 variable sites were detected, with 32 found at least in two samples and 21 were already described by the official HLA database (IMGT/HLA). These variable sites were arranged into 16 haplotypes, five of them counting for more than 70% of all promoter sequences. For the 3'UTR segment, 47 variable sites were detected, all presenting polymorphic frequencies and arranged into 22 haplotypes, in which seven count for more than 75% of all the 3'UTR sequences. The promoter and 3'UTR nucleotide diversity was 0. 016828 and 0. 021434, respectively, what is close to the value observed for the *HLA-C* coding segment considering data from the 1000Genomes Project (0. 025024), but much higher than the one described for the entire human genome (0. 00075). Considering that polymorphisms at the promoter and 3'UTR segments might affect the binding of transcriptional factors and specific microRNAs, respectively, it is possible that these frequent and divergent haplotypes might be associated with different expression profiles and might be affected by different microRNA subsets. Finally, the *HLA-C* promoter and 3'UTR are highly polymorphic and should be considered for functional and clinical studies.

1502T

Characterization of the microsatellite mutation process at every locus in the genome. *M. Gymrek[1,2,3,4], T. Willems[4,5], N. Patterson[2], D. Reich[2,6,7,9], Y. Erlich[4,8,9].* 1) Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA USA; 2) Broad Institute of MIT and Harvard, Cambridge, MA USA; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA USA; 4) New York Genome Center, New York, NY, USA; 5) Computational and Systems Biology Program, MIT, Cambridge, MA USA; 6) Department of Genetics, Harvard Medical School, Boston, MA USA; 7) Howard Hughes Medical Institute, Harvard Medical School, Boston, MA USA; 8) Department of Computer Science, Fu Foundation School of Engineering, Columbia University, New York, NY, USA; 9) Equally supervised this work.

Microsatellites, or short tandem repeats (STRs), are comprised of repeating motifs of 1-6bp that span over 1% of the human genome and contribute to over 40 Mendelian disorders. These loci are prone to polymerase slippage during replication, resulting in mutation rates that are orders of magnitude higher than those of point mutations, conferring high evolvability. Revealing the mutation process of STRs is important to address questions in medical genetics, forensics, and population genetics. However, most studies of microsatellite polymorphism have focused on a highly ascertained set of loci that are extraordinarily polymorphic and easy to genotype, providing biased mutation models. Here, we harnessed novel bioinformatics tools and an analytical framework to estimate the mutation rate at each STR in the human genome. First, we used our lobSTR algorithm to generate the most comprehensive microsatellite polymorphism dataset to date, consisting of nearly 1. 5 million loci across 300 samples sequenced to high coverage by the Simons Genome Diversity Project. These samples originate from diverse genetic backgrounds and maximize our power to observe STR evolution across a wide range of time scales. Genotypes show more than 93% concordance with capillary electrophoresis panels and accurately recover population structure. Next, we developed an analytical model of the microsatellite mutation process that imposes a length constraint on allele size and shows greater consistency than traditional stepwise models with empirical data. Our model allows us to obtain maximum likelihood estimates of mutation rate and length constraint parameters at each microsatellite in the dataset by correlating genotypes with local sequence heterozygosity across the 300 samples. We extensively tested this model on simulated and observed genotypes and accurately recovered mutation rates for markers with published de novo rates. Applying our method at each STR locus in the genome, we find that nearly all loci show evidence of a length constraint, and a subset of microsatellites show extremely high mutation rates of 10-2/locus/generation. We used this call-set to analyze sequence determinants of microsatellite variation, assess patterns of variation in coding regions, and scan for microsatellites under selective pressure.

1503W

Gender-based differences on SNP polymorphisms in candidate genes to obesity: study of the general Mexican Mestizo population. *E. A. Hernández-Tobías[1], R. Camacho-Mejorado[1], C. Santana[2], G. Noris[2], L. Torres-Sánchez[3], M. A. Meraz-Ríos[4], R. Gómez[1].* 1) Departamento de Toxicología, Cinvestav-IPN, Mexico City , Distrito Federal , Mexico; 2) Laboratorio Biología Molecular Diagnóstica, Querétaro, Qro. , Mexico; 3) Instituto Nacional de Salud Pública, INSP, Cuernavaca, Morelos, Mexico; 4) Departamento de Biomedicina Molecular, Cinvestav-IPN, Mexico City, Mexico.

Obesity is a complex multifactorial trait associated with lifestyle and environmental factors, whose rates have risen significantly in the last few decades, reaching pandemic proportions worldwide. Recent studies support that heredity factors are a cornerstone in the development of obesity, which is genetically heterogeneous. Given the complex character of this trait, the identification of genetic markers is an active area of research for further preventive diagnosis and optimal treatment. However, remarkable discrepancies have been found among diverse research groups, related to ethnicity differences. In addition, population-based studies suggest a clear physiological difference between men and women such as disparity in fat distribution, causing differences in incidence and progression of disease, consequently, gender may be a confounder factor. Hence, it is important to know the genetic frequencies in the general population to validate possible associations in further case-control studies. Using a multiethnic population from Central Valley of Mexico, 200 men and 200 women (400 chromosomes), we studied eleven SNPs within or near of six candidate genes for obesity that regulate process such as: adipogenesis (PPARG, PGC1A), inflammation (ADIPOQ and CDH13), lipid storage (LYPLAL1) and overeating disorders (MC4R), through allelic discrimination assays by quantitative PCR. Our results showed that all loci presented similar genetic distributions between sexes, except rs1501299 and rs864265 located on ADIPOQ gene and rs2820446, near LYPLAL1 gene. Regarding ADIPOQ gene, a higher proportion of homozygous state was observed in men than women (~ 0. 78 vs. ~ 0. 54; P = 0. 001). In contrast, the heterozygote state for LYPLAL1 was more prevalent in women than men (~ 0. 60 vs. ~ 0. 1; P = 0. 018). We further analyze Hardy-Weinberg equilibrium for each gender, finding departures in ADIPOQ-rs1501299 and LYPLAL1-rs2820446 only in men, characterized by an excess of homozygotes (*Fis* = 0. 425 and 0. 850 respectively, P < 0. 0001). Results suggest that sex differences in obesity and other comorbidities related to inflammation may have an important genetic regulation background, reflected by differences among sexes in the general population. Thus, we suggest that further studies should consider this information to undertake their predict disease risk with the same accuracy for men and women. *Project supported by funds provided by Kellogg's Nutrition and Health Institute (to R. Gómez).* .

1504T

Influence of deleterious and disease variants on gene expression in human populations. *D. Lu, C. Zhang, K. Yuan, L. Tian, M. Shi, Y. Yuan, S. Xu.* Partner institute for Computational Biology, Shanghai, China.

Deleterious variants in human have been known younger than neutral variants of the same frequency, meanwhile, the age of gene origin is associated with gene expression level, gene haploinsufficiency and gene regulation. We hypothesized that expression levels have been regulated via the effective transcripts, which could make higher expression level on those genes with larger number of deleterious variants. Such 'genetic burden compensation', as we proposed, could affect the distribution of genes with higher expression level in human populations. Here we examined this hypothesis by assessing the number of deleterious variants of genes with different gene origin ages, and analyzing gene expression level based on transcriptomic data and genomic data obtained from the 1000 Genomes Project. Our results showed that the increased number of damaging mutations at younger genes have higher expression level. This observation has great implications as many human specific genes with young age might have played crucial roles during human speciation. We suggest the genetic burden compensation on such younger genes have paved a way and facilitated to human speciation and evolution.

1505W

The Impact of Promoter Polymorphisms on Infant Outcomes and Cytokine Concentration in Preterm Breast Milk. *K. L. Baumgartel[1], M. W. Groer[2,3], S. M. Cohen[1], D. Ren[4], D. L. Spatz[5], Y. P. Conley[1,6].* 1) Health Promotion and Development, University of Pittsburgh School of Nursing, Pittsburgh, PA; 2) College of Nursing, University of South Florida, Tampa, FL; 3) College of Medicine, University of South Florida, Tampa, FL; 4) Health and Community Systems, University of Pittsburgh School of Nursing, Pittsburgh, PA; 5) University of Pennsylvania School of Nursing, Philadelphia, PA; 6) Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA.

The immune protection offered through breast milk is especially important for preterm infants. Interleukins (ILs), found in breast milk but in varying concentrations, may provide preterm infants with protection against complications. The purpose of this study was to: 1) examine the relationship between maternal IL genotypes and weekly milk concentrations of IL4, IL6, and IL10, 2) describe the trajectories of milk IL change over the first three weeks postpartum, 3) examine whether maternal IL genotypes predict milk IL trajectories, 4) examine if weekly IL levels and/or IL trajectories predict infant outcomes, and 5) explore a relationship between maternal IL genotypes and infant outcomes. An ancillary study was conducted among mothers who delivered preterm infants (n=64) and maternal DNA was extracted from breast milk whey for genotyping using TaqMan. Trajectory modeling was used to identify IL subgroups. Subset analyses by ethnicity (Caucasian, African American, Hispanic) were performed. After controlling for gestational age at delivery and ratio of mom's own milk to total milk, maternal minor allele absence (MAA) was significantly associated with fecal calprotectin (FC) in Caucasians (rs1800795 p=0. 0222; rs1800796 p=0. 0158; rs1800872 p=0. 0196). There were significant associations in the total population between FC and rs1800795 GG genotype (p=0. 0213),rs1800871 TT genotype (p=0. 027), rs1800872 AA genotype (p=0. 0158), and rs1800896 AA genotype (p=0. 0045). There was an inverse trend toward significance among Caucasians between rs1800795 MAA and milk IL6 (p=0. 0966). Subsequently, there is a significant relationship between milk IL6 concentration and FC in Caucasians (p=0. 0290). Trajectory analyses resulted in linear group shapes, with two distinct subgroups in IL6, and three subgroups in both IL4 and IL10. SNPs were not associated with trajectory groups. IL6 group 1 membership was significantly associated with intraventricular hemorrhage (OR=6. 275, p=0. 0412), and there was a trend toward significance with FC (p=0. 0822). This study reveals an association between maternal IL SNPs and infant outcomes. This study also suggests that SNPs influence IL milk concentrations, and IL milk levels are associated with outcomes. Trajectory analyses revealed distinct linear-shaped IL milk groups at varying levels and group membership was associated with outcomes. More research is needed to determine clinically relevant milk IL levels and mechanisms for IL variability.

1506T

Profile of pathogenic alleles in healthy Lithuanian population. *V. Kucinskas, T. Rancelis, I. Domarkiene, E. Pranckeviciene, I. Uktveryte, L. Ambrozaityte.* Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, Vilnius, Lithuania.

Introduction. Comparison of genomic variants identified in a particular exome/genome data with known variants of clinical significance is of key importance in diagnostics, prognostics and personal medicine. **Methods.** We analyzed 98 exomes of healthy Lithuanian individuals representing Lithuanian population. DNA samples extracted from venous blood were sequenced using SOLiD 5500 system according to the manufacturer's protocols (Life Technologies, USA). Functional annotation of genomic variants was performed using ANNOVAR including frequencies of genomic variants from 1000 Genomes project (1000G), ESP6500 project and ClinVar database. To compare alleles' frequencies exact Fisher test was used (α=0. 05). **Results.** Each individual exome was found to have 30k to 50k genomic variants. An average of 46 single nucleotide variants (SNV) per individual exome were assessed as pathogenic by ClinVar. 334 unique SNVs in 98 exome group were classified as pathogenic. Taking into account that all study individuals were healthy, they were subdivided into two groups: alternative allele carriers and homozygous for an alternative allele. Variants classified as pathogenic in individuals homozygous for an alternative allele were excluded from analysis (111 variants). Additionally, 46 variants were excluded due to high allele frequency (> 50%) in other populations. A number of pathogenic alleles had higher frequency in Lithuanian population compared to other populations. In a number of genes these pathogenic SNVs were identified: **Stargardt disease 1** (OMIM#248200) allele frequency in the Lithuanian population (LIT) 2. 6%, all populations' allele frequency from 1000G (ALL) 0. 3% (p=7. 993x10-4), European population from 1000G (EUR) 0. 4% (p=7. 165x10-3); **muscle AMP deaminase deficiency** (OMIM#615511) – LIT 7. 1%, ALL 1. 1% (p=7. 368x10-8), EUR 3. 7% (p=2. 607x10-2); **pigmented nodular adrenocortical disease** (OMIM#610475) – LIT 2%, ALL 0. 2% (p=9. 909x10-4), EUR 0. 5% (p=4. 163x10-2); **severe combined immunodeficiency disease** (SNOMEDCT#31323000) – LIT 5. 1%, ALL 0. 6% (p=1. 162x10-6), EUR 2. 2% (p=2. 286x10-2). **Conclusion.** Our identified distribution profile of genomic variants of high confidence in small Lithuanian population enriches current state of scientific knowledge and gives rise to the development of new diagnostic strategies. This study is part of the LITGEN Project (VP1-3. 1-MM-07-K-01-013) funded by the European Social Fund under the Global Grant Measure.

1507W

The effects of single nucleotide polymorphisms on the function of *HSPA1A*, a key regulator of the cellular stress response in humans. *R. M. Oliverio, K. Hess, M. C. Siracusa, N. Nikolaidis.* Biological Science, Center for Applied Biotechnology Studies, and Center for Computational and Applied Mathematics, California State University, Fullerton, Fullerton, CA.

An essential question in molecular evolution is how mutations enable organisms to adapt and survive their environments. At the cellular level, this same question relates to the ability of cells to adapt and survive stressful conditions, such as disease. Cellular adaptation to stress is regulated and maintained by the cellular stress response system. Molecular chaperones, and, in particular the 70-kDa heat shock proteins (Hsp70s) are key orchestrators of the stress response system, and alterations in their function have direct physiological consequences. Therefore, it is of vital importance to determine how natural mutations alter the function of these proteins and how such changes affect cellular and organismal adaptation. To answer these questions we tested whether natural single nucleotide polymorphisms (SNPs) found on *HSPA1A*, the major stress inducible Hsp70 gene in humans, alter protein function. Specifically, the wild-type human *HSPA1A* sequence was subcloned into mammalian expression vectors, and the mutated gene variants were generated using site-directed mutagenesis. We first determined whether any of these mutations affected the intracellular localization of HSPA1A within mammalian cells. These experiments were performed by tagging HSPA1A to GFP, and using fluorescent dyes to stain the nucleus, mitochondria, and lysosomes, to view where the proteins localized via confocal microscopy. These assays revealed that the mutants and the WT protein had similar subcellular localization. Next, we determined whether the mutations affected the ability of HSPA1A to prevent cell death by inhibiting the formation of protein aggregates caused by poly-glutamine carrying huntingtin proteins. This assay determined that two of the mutations caused increased cell death as compared to the WT, suggesting that some mutations alter the chaperone function of HSPA1A. Furthermore, live-dead assays showed that human cells carrying some of these mutations had significantly less resistance to heat stress than cells expressing the WT protein. Given that these natural variants are either population-specific or clinical we suspect that the observed functional differences alter the ability of cells, and the individuals carrying them, to cope with stress and adapt to environmental perturbations.

1508T

Whole-genome reference panel of Tohoku Medical Megabank Organization (ToMMo) and variant annotations. *Y. Yamaguchi-Kabata, N. Nariai, Y. Kawai, Y. Sato, K. Kojima, M. Tateno, F. Katsuoka, J. Yasuda, M. Yamamoto, M. Nagasaki.* Tohoku Medical Megabank Organization, Tohoku University, Aoba-ku, Sendai, Japan.

Although reports of common genomic variants and their frequencies are accumulating for various populations, many of low-frequency variants remain undetected or have unknown frequencies. A catalogue of genomic variants from whole-genome sequencing and estimates of variant frequencies for each population are needed to provide a foundation for genomic medicine. Tohoku University Tohoku Medical Megabank Organization (ToMMo) have sequenced whole genomes of 1,070 cohort participants, and constructed the whole-genome reference panel (1KJPN). We started partial public release of whole-genome Japanese reference panel, and opened a website “integrative Japanese Genome Variation Database (iJGVD; http://ijgvd. megabank. tohoku. ac. jp/)”. The current release of iJGVD provides SNV frequency data obtained from the 1070 individuals. The first release contains data of about 4,300,000 SNVs selected by the criteria: (1) they are on autosomes, (2) their minor allele frequency is greater than 5%, and (3) they have been reported in dbSNP138. This database can be used to search those SNVs and get their allele frequency with allele count, and the genome browser provides graphical views of the genomic location of SNVs with locations of known genes and other SNVs in dbSNP. Allele frequencies of SNVs in iJGVD were compared with those of SNVs in HapMap JPT (Japanese from Tokyo) individuals. The allele frequencies in the two populations were very similar. We are annotating variants of 1KJPN with biological and medical information, and identified overlaps between the SNVs and known pathological SNVs from the Human Gene Mutation Database (HGMD). We believe that our open variant data in iJGVD will be useful in medical genomics, especially for comparison of allele frequencies in iJGVD with those of the patient group for a target disease to identify disease-causing variants. Furthermore, information of pathogenic variants and their frequencies in the population will be utilized for comparison with those in other populations.

1509W

Novel validated Alu elements insertions from Mobile-Element Scanning. *J. Feusier, D. J. Witherspoon, W. S. Watkins, L. B. Jorde.* Human Genetics, University of Utah, Salt Lake City, UT.

Alu and LINE elements are retrotransposons that have been very active during recent human evolution. Specifically, the Alu Yb8 and Yb9 subfamilies account for about 1/3 of recent polymorphic Alu insertions in humans. It is likely that many thousands of Alu elements in human genomes have not been discovered yet. To address this issue, we captured Alu Yb8 and Yb9 elements from more than 400 members of diverse populations using Mobile-Element Scanning (ME-Scan), a multiplexed, targeted high-throughput sequencing method. ME-Scan uses primers specific to the Yb8/9 subfamily sequence to selectively amplify these families. We captured 15,911 insertions absent from the reference genome that were only present in one individual in the study and attempted to validate 60 of these insertions using locus-specific PCR. The evidence supporting each insertion is quantified by the number of supporting reads and by the number of predicted independent library capture events (number of molecules). We determined that 7 distinct molecules were representative of true positive insertion events. Additionally, we tested a set of high-quality insertions in exons and discovered that one individual had an Alu insertion event inside the second exon of *METTL20*. Specifically, the Alu element captured the first seven nucleotides of the exon as part of the target site duplication and inserted before the start of the exon. This exon capture event left the exon intact, but it may affect splice sites and disrupt protein translation. In conclusion, we identified and validated non-reference Alu elements in addition to discovering a rare insertion event inside of an exon.

1510T

Detecting signatures of positive selection associated with musical aptitude in the human genome. *I. Jarvela[1], X. Liu[2,5], C. Kanduri[1], J. Oikkonen[1], K. Karma[3], P. Raijas[4], L. Ukkola-Vuoti[1], Y-Y. Teo[2,5].* 1) Department of Medical Genetics, University of Helsinki, P. O. Box 63, 00014 University of Helsinki, Finland; 2) NUS Graduate School for Integrative Science and Engineering, National University of Singapore, Singapore 117456, Singapore;; 3) DocMus Department, University of the Arts Helsinki, P. O. Box 86, 00251 Helsinki, Finland; 4) Joensuu Conservatoire, Rantakatu 31, 80100 Joensuu, Finland; 5) Saw Swee Hock School of Public Health, National University of Singapore, Singapore 117597, Singapore.

Music perception and practice are complex cognitive functions of the human brain that are well preserved in human evolution. Musical aptitude is a common trait and the prerequisite for music perception and performance. Musical aptitude can be defined as the ability to understand and perceive the rhythm, pitch, timbre, tone duration, and the formal structure in music. Here, we have applied genomic and bioinformatic approaches to assign the regions of positive selection associated with musical aptitude in the human genome. We compared whole genome genotyping data of 647. 417 SNPs from the Illumina HumanOmniExpress 12 1. 0V SNP chip in 74 cases (mean age 44. 41 years, females N=36) and 74 controls (mean age 44. 99 years, females N=47) characterized for musical aptitude. Musical aptitude was defined using three music test scores, auditory structuring ability test (Karma Music test) and Carl Seashore's pitch and time tests, that formed a combined music test score; COMB (range 75-150) in this study. The mean COMB score was 137. 41 (range 125-148) for cases and 100. 30 (range 76-117) for controls. The definition of the phenotype was based on pattern recognition mimicking sound prototypes in songbirds combined with sensory capacities to detect pitch and duration. We assigned signatures of positive selection using three selection methods: haploPS, XP-EHH and FST. We show that regions under positive selection contain several genes affecting auditory perception, steroid hormone secretion, cognitive performance and memory and song perception and production of songbirds. Among them *VLDLR*, a target of human *FOXP2*, that is critical for human speech development. The selection region containing *GPR98* was identified with both haploPS and FST. *GPR98,* involved in inner ear development, is found to be under positive selection in songbird lineage, thus making it a plausible candidate gene for the evolutionary advantage of human musical aptitude. Moreover, *GPR98* is also associated with vision suggesting a common biological origin of sensory perception skills. Gene ontology classification of the identified genes by window statistics revealed the enrichment of genes responsible for inner ear development (GeneTrail2, Hypergeometric distribution test, fdr 0. 04). The findings are consistent with the evolutionary conservation of genes related to auditory processes and cognition and provide first empirical evidence for signatures of positive selection for musical aptitude.

1511W

Understanding high-altitude adaptations in Tibetans using population genetic and genetic association analyses. *C. Jeong[1], B. Basnyat[2], G. Childs[3], S. Craig[4], M. Neupane[5], D. Witonsky[1], C. Beall[6], A. Di Rienzo[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Oxford University Clinical Research Unit, Patan Hospital, Kathmandu, Nepal; 3) Department of Anthropology, Washington University in St. Louis, St. Louis, MO; 4) Department of Anthropology, Dartmouth College, Hanover, NH; 5) Mountain Medicine Society of Nepal, Kathmandu, Nepal; 6) Department of Anthropology, Case Western Reserve University, Cleveland, OH.

Indigenous human populations in the Tibetan plateau are thought to be better adapted to their native environments than those of lowland origin, as reflected in their distinct physiological traits and their lower prevalence of high-altitude related diseases. However, the identity and impact of adaptive phenotypes are poorly understood beyond two candidate genes with signatures of positive selection and with functional involvement in hypoxia: *EGLN1* (egl nine homolog 1) and *EPAS1* (endothelial PAS-domain containing protein 1). To identify genetic variants, phenotypes, and environmental factors associated with high-altitude adaptations in Tibetans, we analyzed a cohort of 996 ethnic Tibetan women born and raised at altitudes (2,982-4,052 m) in Nepal, whose genotypes were determined over 660,000 loci. Phenotypes collected in the field include hemoglobin concentration (Hb), arterial oxygen saturation (SaO2), and fertility traits such as the numbers of pregnancies (PG) and live births (LB). Genome-wide significant associations (linear mixed model (LMM) $p \geq 4.40\times10\text{-}9$) were found between PG and LB and SNPs within *CCDC141* (coiled-coil domain containing 141). We could not detect a genome-wide significant association for either Hb or SaO2. However, association signals were present for candidate genes in Tibetans (*EPAS1* for Hb, LMM $p \geq 4.90\times10\text{-}5$; *HIF1A* for SaO2, LMM $p \geq 6.70\times10\text{-}4$) and for genes associated in other populations (*PRKCE* for Hb, LMM $p \geq 4.07\times10\text{-}4$), thus confirming a role for oxygen homeostasis systems in Tibetan adaptations. Interestingly, the *PRKCE* gene does not harbor signatures of positive selection, despite its physical position close to *EPAS1* and association with Hb. Next, we searched for genomic regions harboring signals of recent positive selection from multiple statistics, using 344 unrelated Tibetans. Both *EPAS1* and *EGLN1* show multiple signals, such as extreme allele frequency change, reduced haplotype diversity and excess haplotype sharing between high-altitude East Asian populations. We also found additional genes with multiple signals and potential functional implication for high-altitude adaptations, including *SLCO1A2*, *EP300* and *XYLT1*. Our study presents a multi-dimensional approach to understanding the genetic basis of human adaptations to local environments, incorporating both population genetic and genetic association analyses, as well as measures of short-term reproductive success.

1512T

A population genetics perspective on quantitative traits. *Y. B. Simons[1], K. Bullaughey[2], R. R. Hudson[2], G. Sella[1].* 1) Biology, Columbia University, New York, NY; 2) Ecology & Evolution, University of Chicago, Chicago, Illinois.

Genome wide association studies (GWAS) have begun uncovering the genetic architecture of a wide array of human quantitative traits, including morphological traits like height and BMI as well as complex diseases like schizophrenia and diabetes. Interpreting GWAS results in evolutionary terms can provide us with an unprecedented insight into the evolution of quantitative traits in humans and may help guide the design of future mapping studies. Evolutionary processes, such as mutation, selection, drift and pleiotropy, shape the genetic architecture of quantitative traits but very few existing models incorporate them in a way that's meaningful for GWAS interpretation. We extend Fisher's Geometric Model to quantitative variation and use it to obtain predictions of the genetic architecture of quantitative traits under different evolutionary scenarios. Under this model, we relate the phenotypic effects of variants to the selection acting upon them and we see that weakly and strongly selected variants are expected to carry more of the variance than nearly neutral sites and therefore be easier to detect in GWAS. However, variants under very strong selection would be too rare to be detected. Pleiotropy is represented by the dimensionality of the trait space and pleiotropic effects weaken the relationship between effect size and selection. The overall effect of pleiotropy on GWAS success is to effectively increase GWAS power. Our analysis suggests that the increase in GWAS success with increasing study size may be highly sigmoidal for some traits and that the increase may be quite dramatic once a large enough study size is reached. This model may provide the basis for using GWAS results to infer the strengths of evolutionary forces shaping quantitative traits in humans.

1513W

Global diversity in the TAS2R38 bitter taste receptor: revisiting a classic evolutionary PROPosal. *D. Risso[1,2], M. Mezzavilla[3], L. Pagani[2,4], A. Robino[3], G. Morini[5], S. Tofanelli[6], M. Carrai[6], D. Campa[6], R. Barale[6], F. Caradonna[7], P. Gasparini[3], D. Luiselli[2], S. Wooding[8], D. Drayna[1].* 1) National Institute on Deafness and Other Communication Disorders, NIH, Bethesda, MD 20892, USA; 2) Department of BiGeA, Laboratory of Molecular Anthropology and Centre for Genome Biology, University of Bologna, via Selmi 3, 40126 Bologna, Italy; 3) Institute for Maternal and Child Health - IRCCS "Burlo Garofolo", Trieste, Italy; University of Trieste, Trieste, Italy; 4) Division of Biological Anthropology, University of Cambridge, CB2 1QH, Cambridge, UK; 5) University of Gastronomic Sciences, Piazza Vittorio Emanuele 9, Bra, Pollenzo 12042, CN, Italy; 6) Department of Biology, University of Pisa, Via Ghini 13, 56126 Pisa, Italy; 7) Biological, Chemical and Pharmaceutical Sciences and Technologies Department, STEBICEF, Université degli Studi di Palermo, V. le delle Scienze, Edificio 16, 90128 Palermo, Italy; 8) Health Sciences Research Institute, University of California at Merced, 5200 North Lake Road, Merced, CA 95343, USA.

The ability to taste phenylthiocarbamide (PTC) and 6-n-propylthiouracil (PROP) is a classic polymorphic trait that is mediated by the *TAS2R38* bitter taste receptor gene. These taste phenotypes have been shown to be correlated with the ability to taste other taste-active compounds, as well as with food habits. Nonetheless, several features of its evolutionary significance and population dynamics are still unresolved. In particular, it is not clear why the worldwide frequency of the *TAS2R38* non-taster AVI haplotype is very high, almost equivalent to that of the taster PAV haplotype. While the long-standing hypothesis suggests that balancing selection has been acting on this locus, other theories have emerged more recently. We performed a detailed analysis of the *TAS2R38* gene and its surrounding regions in a sample of 5511 individuals belonging to 104 different worldwide populations. Our results show no departures from neutral expectations. This suggests that recent demographic events have had a major role in shaping the genetic diversity at this locus, suggesting a reconsideration of the classic hypothesis. We also hypothesize that interactions with the adjacent maltase-glucoamylase (*MGAM*) gene may have contributed to the current distribution of PAV and AVI haplotypes. One hypothesis is that the distribution of the uncommon *TAS2R38* AAI haplotype is interpretable as the product of a recent recombination event that occurred in Africa, after the Out Of Africa (OOA) event. Collectively, our results offer novel insights into the evolutionary history of the *TAS2R38* gene, showing a relaxation of the selective forces previously acting on this gene, and providing a new hypothesis for the observed present-day worldwide distribution of AVI and AAI haplotypes.

1514T

Clarifying the disputed role of *FOXP2* in modern human origins. *E. G. Atkinson, B. M. Henn.* Dept. of Ecology and Evolution, Stony Brook University, Stony Brook, NY.

Identified for its pivotal role in the development of spoken language, the *FOXP2* gene is also known for its controversial role in human evolution. Early genetic work identified a selective sweep for two derived amino acid substitutions in *FOXP2* during recent human evolution (within the past 200,000 years), supported in large part by detection of an extremely low Tajima's *D* value at the gene. When the genomes of other ancient hominids were found to contain the same fixed genetic variants, however, the conflicting timelines between the signals of selection obtained from the molecular sequence of the gene as compared to divergence time estimates between humans and other ancient hominid species were irreconcilable. Selection for these two amino acids thus appears not to be human-specific, yet many papers continue to work from a hypothesis of positive selection of *FOXP2* in humans. Here, we comprehensively re-analyze *FOXP2* with next-generation genomic datasets comprising hundreds of individuals and thousands of SNPs. Specifically, we test for fine-scale molecular patterns in the gene and between various human populations in order to resolve estimates of selection. We are unable to replicate the original negative *D* signal in the expanded human genomic datasets, despite having many more variants, more diverse individuals, and greater statistical power. We can, however, mimic the negative *D* result when running calculations on a subset of the HGDP genomic dataset with a sample of human populations comparable to the original work; i. e. one-third Africans and two-thirds individuals who underwent the Out-of-Africa expansion. The *D* signal thus appears to have been due to the pooling of Africans and non-Africans together for analyses, which increases the number of segregating sites relative to pairwise genetic differences. Such a result seems to have been an unintended consequence of a small sampling strategy. We apply additional selective sweep statistics and haplotype analysis to this locus to evaluate evidence for selection over the past 200,000 years, finding indications of balancing selection in Africans but not non-Africans. *FOXP2* does not appear to have undergone a recent selective sweep, as had been previously proposed.

1515W

Adaptation via gene flow in Africa: a novel approach using ancestry to identify regions of the genome under natural selection. *G. B. J. Busby[1], G. Band[1], Q. S. Le[1], E. Leffler[1], K. Rockett[1,2], D. Kwiatkowski[1,2], C. C. A. Spencer[1,2], MalariaGEN.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN United Kingdom; 2) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge. CB10 1SA United Kingdom.

Patterns of genetic diversity in human populations are the result of a complex demographic history, which is expected to shape variation across the genome, and natural selection, which acts at specific loci. By modelling haplotype similarity we describe shared ancestry within and between 58 worldwide populations with a focus on sub-Saharan Africa. Using a novel approach, we scan the genomes of individuals within these groups to identify loci where there is a significant deviation in local ancestry away from genome-wide expectations. Such signals may identify regions of the genome under natural selection.

Our approach identifies both known and novel beneficial loci that have been shared across populations by gene flow. For a West African pastoralist group called the Fulani, the region of the genome which has maintained the highest proportion of European ancestry contains the lactase persistance gene (*LCT*), and the region with the highest African ancestry contains the Duffy gene (*DARC*), mutations in which provide protection against *Plasmodium vivax* malaria. Across multiple African populations we identify other loci where natural selection has potentially changed the contribution of haplotypes from ancestral populations, including at the Major Histocompatibility Complex and at two loci known to be associated with malaria susceptibility: *ATP2B4* and *HBB*. We additionally see a ten-fold increase in hunter-gatherer ancestry at *SERPINB11* – a gene previously implicated as playing a role in host-parasite interactions – in the Fulani, who have historically hadlower levels of malaria than sympatric populations. Our observations are directly relevant to studies of disease genetics in Africa, and highlight an important – and general – role for gene flow in introducing new adaptations.

1516T

Joint Analysis of Natural Selection and Disease Association. *K. A. Hartman, R. D. Hernandez.* Biological & Medical Informatics, University California, San Francisco, San Francisco, CA.

Natural selection acts to promote genomic variants that drive phenotypic attributes that confer a fitness advantage or to remove those variants that are detrimental to fitness. Depending on the form of selection and the time since the selection took place, different signals will be left in the genomes of modern populations. The field of population genetics has developed a suite of genome-wide scans to detect these signals based on divergence between species or populations, rates of synonymous and nonsynonymous substitution, allele frequency distribution, and length of homozygous haplotypes. As natural selection can only act on genomic loci that cause fitness differences and as the genetic components of disease represent substantial phenotypic differences, it is reasonable to expect a large interplay between natural selection and disease. Indeed previous work showed Mendelian diseases are subject to greater than expected levels of purifying selection and that tag SNPs for a variety of complex diseases have greater than expected levels of background selection. This observation suggest that signals of selection can potentially aid the search for disease associated loci by identifying functionally important sites. We have developed a framework to jointly analyze natural selection and genome-wide association study (GWAS) statistics. The method identifies which signals of selection are predictive of associated sites for a particular phenotype and then uses these signals to refine phenotypic association. An intermediate result of the method is to identify the historical selection pressures that have acted on a particular phenotype. In order to validate the methodology, we have collected publicly available data from GWAS examining a variety of diseases and traits and evaluated signals of selection evaluated in the European populations of the 1000 Genomes Project. We use our framework for a joint analysis of GWAS and natural selection on each phenotype to predict additional sites that were not significant in the original studies and compare these predictions to sites found in replication studies. This framework is widely applicable to many traits, and can strengthen genome-wide associations to any phenotype likely to have a fitness effect.

1517W

Reintroduction of a Homocysteine Level-related Allele into East Asians by Neanderthal Introgression. *Y. Hu[1,2], Q. Ding[1,2], Y. He[2], S. Xu[2], L. Jin[1,2].* 1) State Key Laboratory of Genetic Engineering and Ministry of Education Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China; 2) CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS), Shanghai, China.

Here we present the analysis of Neanderthal introgression at the dipeptidase gene *DPEP1*. Neanderthal origin for the introgressive haplotypes was demonstrated using an established three-step approach. This introgression was under positive selection and reached a frequency of >50% in East Asians. This introgression introduced an allele (rs460879-T) into East Asians, and it was implicated with plasma homocysteine level in Filipino. However, the same allele was also found in non-East Asians, but not of Neanderthal introgression. It is likely that the functional rs460879-T was lost in East Asians, and was reintroduced through Neanderthal introgression. This study suggests that Neanderthal could reintroduce functionally important pre-existing allele into populations where the allele was lost, and shed new light on the understanding of the contribution of Neanderthal introgression to the adaptation of non-Africans.

1518T

OCA2 confers convergent skin lightening of East Asians during recent human evolution. *B. Su[1], Z. Yang[1,2], H. Zhong[3], J. Chen[4], X. Zhang[1], H. Zhang[1], X. Luo[1,2], S. Xu[5], H. Chen[6], D. Lu[5], Y. Han[7], L. Li[8], L. Fu[8], X. Qi[1], Y. Peng[1], K. Xiang[1], Q. Lin[1,2], Y. Guo[1], M. Li[1], X. Cao[1], Y. Zhang[1], L. Zhang[4], X. Guo[9,10], S. Dong[9], F. Liang[9], J. Wang[9,10], A. Willden[1], Q. Li[7], A. Meng[4], H. Shi[1].* 1) State Key Laboratory of Genetic Resources and Evolution, Kunming Instititue Zoology, Chinese Academy Sciences Kunming, China; 2) Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing, China; 3) Department of Pathology and Immunology, Baylor College of Medicine, Houston, USA; 4) State Key Laboratory of Biomembrane and Membrane Engineering, School of Life Sciences, Tsinghua-Peking Center for Life Sciences, Tsinghua University, Beijing, China; 5) Max Planck Independent Research Group on Population Genomics, Chinese Academy of Sciences and Max Planck Society (CAS-MPG) Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China; 6) Center for Computational Genomics, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China; 7) College of Life Science, Liaoning Normal University, Dalian, China; 8) No. 1 School of Clinical Medicine of Kunming Medical University, Kunming, China; 9) BGI-Shenzhen, Shenzhen, China; 10) Department of Biology, University of Copenhagen, Copenhagen, Denmark.

Skin lightening among Eurasians is considered an adaptation to high latitude environments, likely occurred independently in Europe and eastern Asia due to convergent evolution. In Europeans, several responsible genes for lightening have been found, but for East Asians the situation remains elusive. We conducted a genome-wide comparison between dark-skinned Africans and Austro-Asiatic speaking aborigines and light-skinned northern Han Chinese, and identified a pigmentation gene *OCA2* showing unusually deep allelic divergence between them. An amino acid substitution (His615Arg) of *OCA2* prevalent in most eastern Asian populations, but absent in Africans and Europeans, was significantly associated with skin lightening in northern Han Chinese. Further transgenic and targeted gene modification analyses in zebra fishes and mice both recapitulated the phenotypic effect of the *OCA2* variant, resulting from a decreased melanin production. Our results indicate that *OCA2* plays a key role in the convergent skin lightening of East Asians during recent human evolution. .

1519W

Human nose shape differentiation is due in part to adaptation to local temperature. *A. A. Zaidi[1,2], P. Claes[3], B. C. Mattern[2], C. Hughes[4], M. D. Shriver[1,2].* 1) Genetics, Pennsylvania State University, University Park, PA, USA; 2) Department of Anthropology, Pennsylvania State University, University Park, PA, USA; 3) KU Leuven, ESAT/PSI - UZ Leuven, MIRC - iMinds, Medical IT Department, Leuven, Belgium; 4) Department of Anthropology, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

The evolutionary reason for the differentiation of nose shape across human populations is an unsolved mystery. It has long been speculated that differences in nose shape among populations are temperature- and humidity-related adaptations to climate rather than being driven primarily by genetic drift or sexual selection. Functional studies of nasal morphology suggest that the nasal passage is useful in conditioning inspired air so as to prevent damage to the mucosal lining of the lungs. We are primarily interested in investigating whether local adaptation to climate is responsible for the divergence of nose shape across populations. We used a dense-correspondence quasi-landmark based method to quantify the three dimensional nasal shape of 459 individuals of West-African, East-Asian, Northern-European, and South-Asian ancestry. We then used Qst-Fst comparisons to show that the size and shape of the nares, as well as the surface area of the nose are more differentiated across populations than expected by genetic drift alone. Partial Mantel tests and spatial generalized linear mixed models show that both phenotypes are strongly correlated with temperature during the last glacial maximum and the Holocene after correcting for genetic distance, suggesting a role of temperature in the evolution of nose shape. We did not find significant correlations with humidity. In a sample of 486 individuals with mixed West African and European ancestry, we show that these traits are highly correlated with genetic ancestry. We further used a linear mixed model as implemented in the program GCTA to show that a set of ~500K autosomal SNPs explain a large proportion of the variance in nose shape in a sample of 1,721 individuals of European ancestry. These results suggest that both aspects of nose shape are heritable. Finally, we identified putative loci associated with nose shape in a sample of 324 individuals of mixed African and European ancestry and show that these loci exhibit signatures of accelerated evolution across populations. Our results provide strong support for the hypothesis that differentiation in nose shape across populations has been driven by local adaptation to temperature.

1520T

Identifying the target of selective sweep. *A. Akbari[1], G. Tesler[2], R. Ronen[3], Y. Lin[4], N. Rosenberg[5], V. Bafna[4].* 1) Department of Electrical & Computer Engineering, University of California San Diego, La Jolla, CA; 2) Department of Mathematics, University of California San Diego, La Jolla, CA; 3) Bioinformatics Graduate Program, University of California San Diego, La Jolla, CA; 4) Department of Computer Science & Engineering, University of California San Diego, La Jolla, CA; 5) Department of Biology, Stanford University, Stanford, CA.

Methods for detecting the genomic signatures of natural selection have been heavily studied, and have been successful in identifying many selective sweeps. For the vast majority of these sweeps, the favored allele remains unknown, making it difficult to distinguish carriers of the sweep from non-carriers. Because carriers of ongoing selective sweeps are likely to contain a future most recent common ancestor, identifying them may prove useful in predicting the evolutionary trajectory – for example, in contexts involving drug-resistant pathogen strains or cancer subclones. In this research, we start with the development and analysis of a new statistic, the Haplotype Allele Frequency (HAF) score, assigned to individual haplotypes in a sample. The HAF score naturally captures many of the properties shared by haplotypes carrying an adaptive allele. We provide a theoretical model for the behavior of the HAF score under different evolutionary scenarios, and validate the interpretation of the statistic with simulated data. We develop an algorithm (PreCIOSS: Predicting Carriers of Ongoing Selective Sweeps) to identify carriers of the adaptive allele in selective sweeps, and we demonstrate its power on simulations of both hard and soft selective sweeps, as well as on data from well-known sweeps in human populations.

We also investigate the problem of identifying the favored mutation itself, without using functional information. We developed a tool to reconstruct possible genealogical scenarios, using clusters of recurrent mutations as guides to ancestral conserved haplotypes. We use a novel algorithm to select conserved haplotype segments that best separate carriers and non-carriers identified by PreCIOSS. In simulations, and on real data-sets with known sites under positive selection, our tool is able to short-list the favored mutation with high sensitivity and specificity. Applying this method on a 600 kbp (163 sites) region including the lactase (LCT) gene region centered at the known favored mutation in European (CEU) population, we obtained a small list of 3 candidates that includes the C/T-13910 (rs4988235) mutation, for which the T allele was found to be 100% associated with lactase persistence in the Finnishpopulation. Similar results were observed for other selective sweeps including PSCA in African (YRI), and ADH1B and EDAR in Asian (CHB+JPT) populations from the Hapmap2 project.

.

1521W

A new measure of positive selection reveals genome-wide polygenic adaptation for complex traits in humans' recent evolutionary past. *E. A. Boyle[1,2], N. Telis[3], Y. Field[1,5], J. K. Pritchard[1,4,5].* 1) Genetics Dept, Stanford University; 2) School of Medicine, Stanford University; 3) Biomedical Informatics, Stanford University; 4) Biology Dept, Stanford University; 5) Howard Hughes Medical Institute.

Geneticists have long been stymied by the challenge of characterizing highly polygenic traits in human populations. While genome-wide association studies have permitted the profiling of more than 1,000 human traits in less than ten years, the implications of known or suspected SNP-trait associations have not been fully elucidated. Concurrently, other groups focusing on identifying how selection on genomic regions has shaped human evolution have struggled to identify the functional underpinnings of detected signal. While cases of selection acting outside the paradigm of hard sweeps are known in the literature, the relative dearth of such examples may suggest a gap unaddressed by current methods for detecting complex genetic adaptation. To investigate the potential contribution of polygenic, genome-wide adaptation in recent human history, we have developed a new metric – the Singleton Density Score – that aggregates signals of positive selection using singleton density across individuals within the Avon Longitudinal Study of Parents and Children (ALSPAC). We relate this novel metric to other genome-wide annotations and recapitulate prior work showing polygenic genetic architecture for height-favoring alleles in northern Europe. Finally, we apply this measure to over two-dozen other complex traits to quantify the effect of recent polygenic adaption on human quantitative traits in present-day populations and find evidence for selection acting to increase cranial size in Europeans' recent evolutionary history.

1522T

Estimating The Selective Effect For Each Gene Using Large Scale Population Data. *C. A. Cassa[1], D. M. Jordan[1], D. J. Balick[1], D. Nusinow[1], D. Beier[2], S. Sunyaev[1].* 1) Division of Genetics, Harvard Medical School/ Brigham and Women's Hospital, Boston, MA; 2) University of Washington, Seattle Children's Hospital, Seattle, WA.

Large sequencing datasetsenableunbiased queries into the significance and potential consequences of DNA sequence variation. Studies have assessed the functional importance of differential mutational load and the depletionof variants in genes. Here, we use the genome-wide distribution of expected and observed protein truncating variants (PTV) in a set of exomes to estimate the strength of selectionfor each gene. Next, we use this selective effect data to characterize the functional and clinical associations of genes throughout the spectrum of selective effects. We identify a set of understudied genes with high potential for developmental importance or essential function.

Characterization of the strength of selection (*s*) and dominance (*h*) coefficients for human genes is of great interest for evolutionary research and clinical interpretation. Estimates of these parameters in humans have proved elusive, as inference techniques used in other organisms generally require cross-breeding over several generations. Newly available large-scale population sequencing data provides insight into the selection that prevents PTVs from being observed in the general population. Using 60,706 patientexomeswithout severe Mendelian disorders from Exome Aggregation Consortium (ExAC), we use allele frequency information to estimate the strength of heterozygote selection in the form of the combined parameter s_het = *h*s*. Statistical power to infer these parameters in greatly increased by the simplifying assumption that all PTVswithin a given geneincur the same selective disadvantage,which should be particularly high in genes intolerant to such mutations. This allows us to treat PTVs in each gene as a single bi-allelic locus, where observations of PTVs throughout each gene are treated as multiple observations of the same PTV allele. We generally expect that mutations will enter the population in each gene at a localized rate of *U*, as estimated by a mutational model, and arefiltered by selection in proportionto their localized *hs* value. The forces of mutation and selection act against one another to generate and filter PTVs in each gene, and the ratio of these two values should be proportional to the observed number of PTV alleles in the population. For genes under extreme purifying selection (s~1), we can use the combined parameter *hs* to estimate the dominance coefficient *h*using the observed values of the mutation rate and variant frequencies for each gene.

1523W

The lingering load of archaic admixture in modern human populations. *K. Harris[1,2], R. Nielsen[2,3].* 1) Stanford University, Stanford, CA; 2) University of California Berkeley, Berkeley, CA; 3) Center for Bioinformatics, University of Copenhagen, Copenhagen, Denmark.

Founder effects and bottlenecks can damage fitness by letting deleterious alleles drift to high frequencies. This almost certainly imposed a burden on Neanderthals and Denisovans, archaic hominid populations whose genetic diversity was less than a quarter of the level seen in humans today. A more controversial question is whether the out-of-Africa bottleneck created differences in genetic load between modern human populations. Some previous studies concluded that this bottleneck saddled non-Africans with potentially damaging genetic variants that could affect disease incidence across the globe today (e. g. Lohmueller, et al. 2009; Fu, et al. 2014), while other studies have concluded that there is little difference in genetic load between Africans and non-Africans (e. g. Simons, et al. 2014; Do, et al. 2015). Although previous studies have devoted considerable attention to simulating the accumulation of deleterious mutations during the out-of-Africa bottleneck, none to our knowledge have incorporated the fitness effects of introgression from Neanderthals into non-Africans. We present simulations showing that archaic introgression may have had a greater fitness effect than the out-of-Africa bottleneck itself, saddling non-Africans with weakly deleterious alleles that accumulated as nearly neutral variants in Neanderthals. Assuming that the exome experiences deleterious mutations with additive fitness effects drawn from a previously inferred gamma distribution, we predict that the fitness of the average Neanderthal was about 50% lower than the fitness of the average human, implying the existence of strong selection against early Neanderthal-human hybrids. This is a direct consequence of mutation accumulation during a period of low Neanderthal population size that is thought to have lasted ten times longer than the out-of-Africa bottleneck (Pruefer, et al. 2014). Although our model predicts some transmission of deleterious Neanderthal variation to present-day non-Africans, it also predicts that many Neanderthal alleles have been purged away, depleting conserved genomic regions of Neanderthal ancestry as observed empirically by Sankararaman, et al. (2014). Our results imply that the deficit of Neanderthal DNA from functional genomic regions can be explained without the action of epistatic reproductive incompatibilities between human and Neanderthal alleles.

1524T

Pleiotropic effects of immune responses explain variation in the prevalence of fibroproliferative diseases. *M. Huang[1], SB. Russell[2], JC. Smith[3], JS. Trupin[4], SM. Williams[1].* 1) Department of Genetics, Geisel School of Medicine, Dartmouth College, Hanover, NH 03755; 2) Vanderbilt Genetics Institute, Division of Dermatology, Department of Medicine, Vanderbilt University, Nashville, TN 37091; 3) NHC Healthcare, Lewisberg, TN 37091; 4) 143 School St. Marshfield, VT 05658.

Many diseases are differentially distributed among human populations. Differential selection in ancestral environments that coincidentally predispose to disease can be an underlying cause of these unequal prevalence patterns. Some of the selected genes may be pleiotropic, affecting multiple phenotypes and resulting in more than one disease or trait. Patterns of pleiotropy may be helpful in understanding the underlying causes of an array of conditions in a population. For example, several fibroproliferative diseases are shown to be more prevalent and severe in populations of sub-Saharan ancestry. We propose that this disparity is due to selection for an enhanced Th2 response that confers resistance to helminthic infections, and concurrently increases susceptibility to fibrosis due to the profibrotic action of Th2 cytokines. Through an extensive review of the literature, we show higher relative frequencies of many fibroproliferative diseases in black populations than in white. We provide evidence that instead of conferring a selective advantage with respect to the specific disease, this pattern is best explained by selection for resistance to diseases caused by helminthic worms, prevalent in sub-Saharan Africa. Resistance to helminths is achieved by a shift in the immune system toward an enhanced Th2 immune response that coincidentally results in fibrosis. We show the allelic distributions of variants that affect Th2 response are almost universally consistent with this hypothesis. Furthermore, by comparing the differentiation between African and European populations with respect to Th2, Th1 and TGFB genes, we show that there is greater differentiation in variants in Th2 related genes compared to the genetic background between African and non-African populations, an observation consistent with selection favoring Th2 enhancement in Africa. In summary, we provide evidence that adaptation of the immune system has shaped the genetic structure of these human populations in ways that alter the distribution of multiple fibroproliferative diseases.

1525W

Signatures of positive natural selection targeting human microsatellites and estimation of relevant mutational and selective parameters. *R. Johnson[1], B. Payseur[2], R. Haasl[1].* 1) University of Wisconsin-Platteville, Department of Biology, 1 University Plaza, Platteville, WI, 53818; 2) University of Wisconsin-Madison, Laboratory of Genetics, 425-G Henry Mall, Madison, WI 53706.

Microsatellites are common genetic polymorphisms in which the copy number of a short nucleotide motif is variable. Although the vast majority of microsatellites are thought to evolve neutrally, variation at several microsatellites is implicated in human diseases including Huntington disease and Friedrich ataxia. Recent empirical data also suggest that some microsatellites possess adaptive repeat lengths, which have been selected for by positive natural or artificial selection. The goal of our research was to analyze candidate microsatellite loci in the human genome for signatures of positive natural selection. We genotyped 200 microsatellites in 200 human individuals from eight 1000 Genomes populations and applied a method based on approximate Bayesian computation (ABC) to infer neutral versus selective evolution. In cases where a model of positive selection was favored, we estimated the strength of selection and the mutational profile of the microsatellite. Here, we present the first results of our analysis, focusing on five microsatellite loci found in an intron or basal promoter of the genes *PAX3*, *GRIN2B, NOS2*, *SHBG*, and *EGFR*. In each case, we provide compelling evidence that the distribution of variation found at the microsatellite has been shaped by positive natural selection. Furthermore, through our joint inference of mutational and selective parameters, we demonstrate the complicated interplay of natural selection and mutation at these highly mutable variants. In particular, we show that the repeat length of adaptive alleles at a microsatellite locus affects the efficacy of positive selection as longer microsatellites tend to possess higher mutation rates, thereby generating less adaptive alleles with appreciable frequency.

1526T

Evolution of the 'fused' gene family's repeat region in primates. *V. Romero, I. Inoue.* Human Genetics, National Institute of Genetics, Mishima, Shizuoka, Japan.

Epidermal Differentiation Complex (EDC) on chromosome 1 is a dense cluster of genes expressed during the late differentiation of epidermis and is responsible for maintaining hydration and structure of skin. The cluster organization of this complex might suggest adaptation during evolution and is strongly associated with a variety of skin disorders. EDC includes the CE precursor family, S100A family and the 'fused' gene family. The 'fused' genes members are: Cornulin, Filaggrin, Filaggrin-2, Hornerin, Repetin and Trichohyalin. These genes have similar organization showing an initial S100 domain followed by a tandem repeat region and a C-terminal domain. In this study, we focus on the repeat region of each gene and the evolution pathway among Marmoset, Macaque, Baboon, Orangutan, Gorilla, Chimpanzee and Human. We demostrate that each member of the 'fused' genes is undergoing different selective pressure, which infers a complex model of adaptive evolution for the EDC. As an example, Filaggrin2/Filaggrin number of repeats varies between species; Human has 14, Chimpanzee 13, Macaque 14, Baboon 22 and Marmoset 8. The similarity between these repeats ranges from 70% to 100%. Phylogenic analysis separated all Marmoset-repeats apart, first repeats from all species, and then the rest of Macaque and Baboon, and Chimpanzee and Human pair together. Maximum-likelihood analysis detected positive selection in 7 codons of all branches, with each branch under different selective pressure and positive selection in 2 codons of the Marmoset-branch. We also inferred 15 duplications and 10 losses events by reconciling the gene tree with the species tree. Additional studies were performed with the rest of 'fused' genes.

1527W

Signals of Selection in Genes Associated with Autoimmune Diseases. *R. Rothwell[1], S. Zoellner[1,2], BRIDGES Consortium.* 1) Biostatistics, University of Michigan, Ann Arbor, MI; 2) Psychiatry, University of Michigan, Ann Arbor, MI.

Recent epidemiological studies estimate autoimmune diseases collectively affect at least 5% of individuals worldwide. This group of diseases, including Crohn's disease, Celiac disease, multiple sclerosis, rheumatoid arthritis, lupus, spondyloarthropathy, type 1 diabetes, and ulcerative colitis, while presenting a wide range of symptoms, share multiple associated loci. In particular, there is substantial evidence for associations within loci of the major histocompatibility complex (MHC). The best characterization is for the human leukocyte antigen (HLA) loci, in which studies have implicated pathogen-driven balancing selection. Non-HLA shared loci, however, also show evidence of being directly causative in disease, though less is known of their evolutionary origin. Understanding this origin is important to developing the underlying genetic model of disease, therefore providing biological insights and leading to future medical interventions. It is also of interest to the general study of human evolution, exploring the process of adaption and the development of phenotypic diversity. One hypothesis is variants in these loci were previously selected for some protection from infectious diseases or offered some other evolutionary advantage. To assess this idea, commonly referred to as the "thrifty gene hypothesis", we use whole genome sequencing data to look for selection signals at previously implicated non-HLA autoimmune associated loci. The BRIDGES consortium dataset, one of the most extensive data sets to be used in sequencing scans, consists of 3765 European individuals sequenced to an average coverage of 9. 2x. Specifically, our analyses includes site frequency spectrum-based tests, (Tajima's D, Fay and Wu's H), linkage-disequilibrium based neutrality tests (extended haplotype homozygosity, integrated haplotype score), and inter-species neutrality tests (McDonald-Kreitman). We calculate these statistics in windows across the genome, generating an empirical distribution. To test for significance while accounting for population effects such as exponential growth, we compare the windows within our genes to this empirical distribution. We therefore conduct a comprehensive search for signals of previous positive selection, ongoing selection, and evolutionary adaptation, providing a complete description of the evolution of these autoimmune-associated loci.

1528T

Natural Selection on *HFE* in Asian Populations Contributes to Enhanced Non-heme Iron Absorption. *K. Ye[1], C. Cao[1], L. Xu[2], K. O'Brien[1], Z. Gu[1].* 1) Division of Nutritional Sciences, Cornell University, Ithaca, NY, USA; 2) Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences and Graduate University of the Chinese Academy of Sciences, Shanghai, China.

HFE, a major regulator of iron (Fe) homeostasis, has been suggested to be under positive selection in both European and Asian populations. While the genetic variant under selection in Europeans (C282Y) has been relatively well-studied, the adaptive variant in Asians and its functional consequences are still unknown. Identifying the adaptive *HFE* variants in Asians will not only elucidate the evolutionary history and the genetic basis of population difference in Fe status, but also assist the future practice of genome-informed dietary recommendation. Using the HapMap Project data, we observed significant iHS values around *HFE* only in Asians, suggesting the presence of positive selection and population-specific selection pressure. We further identified a candidate adaptive haplotype that is common in Asians (52. 35-54. 71%) but rare in Europeans (5. 98%) and Africans (4. 35%). The T allele at tag SNP rs9366637 (C/T) captured 95. 8% of this Asian-common haplotype. FST statistics and differences of derived allele frequency (χDAF) for rs9366637 are significant between Asians and Europeans or Africans. Previous association studies have shown that the T allele of rs9366637 is associated with increased levels of circulating ferritin in women. We further showed that individuals carrying T/T or C/T at rs9366637 have significantly reduced *HFE* expression compared to C/C in lymphoblastoid cell lines from the Hapmap Project and in another four tissues/cells from GTEx, including lung, tibial nerve, EBV-transformed lymphocytes and prostate. As reduced expression of *HFE* theoretically would lead to enhanced non-heme Fe absorption, we hypothesized that the Asian-common haplotype of *HFE* was selected for enhanced non-heme Fe absorption in Asian populations, which has a long history of consuming plant-based Fe-poor diet. To test this hypothesis, we recruited 57 women of Asian descent and measured Fe absorption using stable isotopes in those homozygous at rs9366637, representing homozygous carriers (T/T) and non-carriers (C/C) of the Asian-common haplotype. We observed a 22% higher absorption in homozygous carriers compared to the non-carriers. Additionally, compared with a group of age-matched Caucasian women, Asian women exhibited significantly elevated Fe absorption after controlling for Fe status. Our results indicate population differences in Fe homeostasis and suggest that natural selection on *HFE* may have contributed to elevated Fe absorption in Asians.

1529W

Prevalence of an archaic high altitude adaptive *EPAS1* haplotype in the Himalayas. *Q. Ayub[1], S. Hackinger[1], T. Kraaijenbrink[2], Y. Xue[1], M. Mezzavilla[1], G. van Driem[3], M. A. Jobling[4], P. de Knijff[2], C. Tyler-Smith[1].* 1) Wellcome Trust Sanger Institute, Hinxton, Cambs, United Kingdom; 2) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 3) Institute of Linguistics, University of Bern, Bern CH-3012, Switzerland; 4) Department of Genetics, University of Leicester, Leicester, United Kingdom.

Genetic, biochemical and morphological changes have enabled humans to adapt to living at high altitudes in Asia, Africa and South America. High altitude adaptation in Tibetans is reportedly influenced by introgression of a 32. 7 kb long haplotype from the Denisovans, an extinct branch of archaic humans. This haplotype lies within the endothelial PAS domain protein 1 (*EPAS1*), a transcription factor acting in the hypoxia inducible factor pathway. A parallel study indicated that the same haplotype had probably entered the Tibetan population from the Sherpa, a high altitude adapted population from Nepal, thus suggesting that most likely the Denisovan introgression occurred in a population ancestral to the Sherpa and Tibetans. We genotyped 22 single nucleotide variants (SNVs) in this region in 1,550 Eurasian individuals, including 1,233 from Bhutan and Nepal residing at altitudes ranging from 86 – 4,550 m above sea level. Derived alleles for 5 SNVs (rs115321619, rs73926263, rs73926264, rs73926265, rs55981512) that characterize the core Denisovan haplotype (AGGAA) were present at high frequency not only in Tibetans and Sherpa, but also among many ethno-linguistic groups from Bhutan and Nepal. The frequency of the Denisovan core haplotype in these populations shows a significant correlation with altitude (Spearman's correlation coefficient = 0. 797, p-value 6. 996 x 10-12). The Denisovan derived alleles were also observed at frequencies of 3-14% in the 1000 Genomes Project African samples and an additional 10 East and South Asian samples shared the Denisovan haplotype that extends beyond the 32 kb region. These additional samples enabled us to refine the haplotype structure and identify candidate functional variants that might be driving the selection signal.

1530T

Genetic Basis of Polygenic Adaptation in Indigenous Siberian Populations Inferred using Exome Sequencing Data. *P. Hsieh[1], B. Hallmark[2], TM. Karafet[3], MF. Hammer[1,3], RN. Gutenkunst[1,4].* 1) Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ; 2) Interdisciplinary Program in Statistics, University of Arizona, Tucson, AZ; 3) Arizona Research Laboratories, Division of Biotechnology, University of Arizona, Tucson, AZ; 4) Department of Molecular and Cellular Biology, University of Arizona, Tucson, AZ.

Siberia is one of the coldest environments on Earth and has great seasonal temperature variation. Recent archeological studies indicate that humans have occupied Siberia for at least ~45,000 years, and persisted through the Last Glacial Maximum in North Eurasia. As early modern humans dispersed from their ancestral tropical African homeland into much cooler environments, long-term settlement in Siberia undoubtedly required biological adaptation to severe cold stress, dramatic variation in photoperiod, as well as limited and highly variable food resources. Humans are the only primate species other than the Japanese macaque that has adapted to boreal conditions—where temperatures remain far below freezing for more than half the year—pointing to intense selection pressures that likely drove the enhancement of physiological processes that generate and conserve heat. Physiological evidence, such as differences in basal metabolic rates and brown adipose tissue, suggests genetic adaptions in Arctic populations to life at high latitude. Because many of these physiological traits, including body mass and metabolic processes, are highly polygenic, we sought signatures of polygenic selection in Siberian populations. We sequenced exomes of individuals from two indigenous Siberian populations: the Nganasans (N = 21), who are the northernmost indigenous group in the world, and the Yakuts (N = 21), who live in the coldest regions on our planet. To detect polygenic selection, we performed gene-set enrichment analysis using pathways from the NCBI Biosystems database as well as a set of candidate genes that have been previously implicated in cold adaptation. The significance of the candidate gene sets for polygenic selection was assessed using whole-exome coalescent simulations to account for potential biases caused by demographic processes and heterogeneity in mutational and recombination rates across the entire genome. Our results thus give insight into the complex polygenic basis of adaptation to life in cold environments in human populations.

1531W

Population structure analysis in a Southwest American Indian population: potential for understanding natural selection in disease susceptibility. *P. Kumar, W-C. Hsueh, A. K. Nair, P. Piaggi, S. Kobes, R. L. Hanson, L. J. Baier.* Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, Phoenix, AZ.

The high prevalence rates of type 2 diabetes and obesity in modern humans is a major health concern and have disproportionally affected specific populations including the Pima Indians of Southwest Arizona. It has been proposed that lifestyle changes have influenced evolutionary pressure on genes that regulate physiologic pathways which underlie complex genetic disorders including type 2 diabetes and obesity. Therefore, analysis of natural selection signatures may be important to understand the evolutionary processes and population specific differences in prevalence of certain metabolic diseases in different geographical regions. Here, we used genome-wide Affymetrix 6. 0 genotypic data from 73 unrelated full heritage Pima Indians (an estimated kinship coefficient < 0. 03 between any 2 subjects) and integrated these genotypes with genotypic data from 11 HapMap Phase III populations. Markers common to both datasets (N=355,000) were used for population genetic and natural selection analyses. Our principle component (PC) analysis separated Pima Indians from Europeans and East Asians at PC2 (4. 3%) and showed more closeness to East Asians. Population structure analysis using FRAPPE at K =4 (four population model) showed major components identifiable as Amerindian, European, East Asian and African across the populations. In Pima Indians, the major components were Amerindian 0. 91±0. 11 (mean±SD), European 0. 05±0. 10, and East Asian 0. 02±0. 01. The Amerindian component in the Mexican was 0. 39±0. 13. These analyses show that Pima Indian ancestry is majorly Amerindian and there is minimal admixture from other continental populations. Therefore selection signatures that underlie complex diseases common to this population may be unique as compared to other study populations. Additional analysis related to population differences, such as FST and long range haplotype analyses for recent selection events, is ongoing. Genes/regions identified by these methods will be tested for implication in disease processes in Pima Indians.

1532T

A Genomic Map of Positive Selection in Sardinia. *J. H. Marcus[1,9], M. Steri[2,9], M. Floris[3,9], C. WK. Chiang[4], J. Smith[5], F. Busonero[2], A. Maschio[2,6], A. Mulas[2,8], S. Sanna[2], G. Pistis[2], M. Pitzalis[2], M. Zoledziewska[2], A. Angius[2], C. Sidore[2,6], D. Schlessinger[7], G. R. Abecasis[6], J. Novembre[1,5,10], F. Cucca[2,8,10].* 1) Department of Human Genetics, University of Chicago, IL, USA; 2) Istituto di Ricerca Genetica e Biomedica, CNR, Monserrato, Cagliari, Italy; 3) Center for Advanced Studies, Research, and Development in Sardinia (CRS4), AGCT Program, Parco Scientifico e Tecnologico della Sardegna, Pula, Italy; 4) Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA; 5) Department of Ecology and Evolution, University of Chicago, IL, USA; 6) Center for Statistical Genetics, Ann Arbor, University of Michigan, MI, USA; 7) Laboratory of Genetics, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA; 8) Université degli Studi di Sassari, Sassari, Italy; 9) Co-First Authors; 10) Co-Last Authors.

The recent production of population-scale genomic data offers an unprecedented opportunity to understand how natural selection has shaped human phenotypic variation within populations. Sardinia has a rich history of genetic studies driven by its relative isolation and high incidence of malaria, which was endemic there until eradication efforts in the 1940s. To identify signatures of recent positive selection in Sardinia, we use 23 million single nucleotide polymorphisms from low-coverage whole genomes of 3,514 Sardinians along with data from the 1000 Genomes project. Using haplotype (iHS, nSL), cross-population (Fst, PBS, XP-EHH), and site-frequency-spectra (CLR) based statistics we find many genetic variants show evidence of selection. To assess the significance of these selection statistics, we use an empirical null distribution generated from randomly chosen variants matched by minor allele frequency, local recombination rate, and background score. We also evaluate these statistics relative to a null, neutral model using a demographic history inferred from deeply sequenced Sardinian individuals. We show that selection statistics computed for outlier variants cannot be explained by neutral forces alone. By intersecting genome-wide-association study data for hundreds of traits in Sardinia with publicly available functional genomic databases we find that autoimmunity-related genes are significantly enriched for these putatively adaptive variants. Taken together, these results illustrate the importance of characterizing both the demographic history of and phenotypic variation within a population, and especially the utility of whole-genome-sequence data, when proposing and interpreting genetic signatures of positive selection.

1533W

Analyzing Nonsynonymous SNPs in their Protein Structural Context Reveals Novel Signatures of Population Differentiation in Humans. *R. M. Sivley[1,4], J. A. Capra[2,4], W. S. Bush[3,4,5].* 1) Department of Biomedical Informatics, Vanderbilt University, Nashville, TN; 2) Department of Biological Sciences, Vanderbilt University, Nashville, TN; 3) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 4) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 5) Institute for Computational Biology, Case Western Reserve University, Cleveland, OH.

The primary goal of this project is to integrate spatial and functional information from protein structures into population genetic analyses. The increasing availability of experimentally determined and computationally predicted protein structures provides a functionally relevant perspective from which to analyze whole-exome and whole-genome sequencing data. The spatial orientation of variants in 3D structures reflects their functional context and reveals relationships between variants that are distant in sequence, but nearby in structure. To enable this type of analysis, we first mapped 534,947 nonsynonymous SNPs (nsSNPs) from Phase 3 of the 1000 Genomes Project (1KG) to the affected amino acids in 4,965 biological assemblies from the Protein Data Bank. We found that for 59% of 1KG nsSNPs, the nearest nsSNP in structure differed from its nearest nsSNP in the genome. Heteromeric protein complexes accounted for 3. 7% of these nsSNPs, with the nearest nsSNP in structure located in another gene. To demonstrate the utility of incorporating structural information into genetic analyses, we performed a comprehensive scan for population differentiation in human protein structures. We estimated Wright's FST over nsSNPs within 10Å of each amino acid in each structure using the 1KG super-populations. We obtained an empirical p-value by shuffling nsSNP positions and recalculating each FST. We identified 1,791 proteins with structurally defined regions of significant ($p<0.05$) population differentiation, 256 of which were not detected by comparable sequence-based methods. Among the proteins with the most highly differentiated structural regions (FST>0. 05), we identified significant enrichment for several functional categories, most notably immune system response genes (BH adjP=0. 0365), suggesting that differences in pathogen exposure between human populations may have driven differentiation in coherent structural regions. These results highlight the extent to which protein folding affects the spatial proximity of variable amino acids, and demonstrate that considering these structural relationships enables the detection of novel signatures of population differentiation.

1534T

Novel probabilistically interpretable methods for identifying and localizing genomic targets of selective sweeps. *L. A. Sugden, S. Ramachandran.* Department of Ecology and Evolutionary Biology, Brown University, Providence, RI.

Human populations throughout the world have had to adapt to novel pathogens and environments; this adaptive evolution has shaped present-day genomes. Here, we introduce novel frameworks for detecting adaptive sweeps from *de novo* mutations that are easily extensible to detecting adaptive evolution from standing variation. While current methods for detecting adaptive mutations rely on single statistics that probe one of three major signatures of a sweep -- long-range haplotype blocks, changes in the site frequency spectrum, and population differentiation -- recently, composite methods have shown increased power by combining multiple statistics. However, these methods falter when a subset of their component statistics is undefined, as often happens with long-range haplotype statistics, and they yield scores that are fundamentally difficult to interpret. Our approach classifies local targets of selective sweeps within multiple populations in a way that combines multiple statistics, has an easy probabilistic interpretation, and deals naturally with undefined statistics. We introduce two classifiers that infer the probability that a new locus has undergone a sweep, based on distributions learned from demographic simulations. The first is a Naïve Bayes classifier, which assumes independence among component statistics, while the second uses a machine-learning tool called an Averaged One-Dependence Estimator (AODE) to allow for pairwise dependencies. In simulated data, we show that the Naïve Bayes classifier vastly outperforms state-of-the-art methods in detection and localization of sweep signals, in some cases reducing the number of false positive predictions by seven-fold. We show that this classifier performs particularly well when identifying completed sweeps and fast sweeps, which have great biological significance. For a subset of sweep parameters, the AODE further improves classification performance. In data from the 1000 Genomes Project, we show that both classifiers can detect known sweep targets, including the DARC locus in West Africans, the EDAR locus in East Asians, and the SLC24A5 locus in Europeans. We also show that the dependency structure implemented in the AODE is necessary for detection of some signals, including the CD36 locus in West Africans, which harbors malaria resistance alleles. Our methods produce fewer false positives and negatives compared to existing approaches, thus identifying promising targets for experimental validation.

1535W

Adaptation in global human populations has been hard, soft and polygenic. *Z. A. Szpiech[1], R. D. Hernandez[1,2,3].* 1) Department of Bioengineering and Therapeutic Sciences, University of California at San Francisco, San Francisco, CA 94158; 2) Institute for Human Genetics, University of California at San Francisco, San Francisco, CA 94158; 3) Institute for Quantitative Biosciences (QB3), University of California at San Francisco, San Francisco, CA 94158.

There is ample debate about the strength and mode of natural selection that has occurred in recent human evolution. This is particularly so for classical hard sweeps, during which an adaptive allele quickly drags a single haplotype to high frequency. An alternative model of adaptation involves soft sweeps, whereby multiple haplotypes are brought to high frequency (i. e. when a previously segregating neutral or slightly deleterious allele becomes adaptive in a new environment). Yet another alternative model includes polygenic selection, whereby complex phenotypes driven by multiple loci across the genome are selected. Here we develop new statistics designed to identify both hard and soft sweeps, by tracking the decay of homozygosity of the k-most frequent haplotypes away from a core locus. We evaluate our statistics with rigorous simulations under multiple realistic models of human demography and find that they have high power. We then integrate signals of selection across the genome to identify characteristic signals of polygenic selection. We apply our approaches to a large dataset of 1,728 unrelated individuals spanning 20 worldwide human populations from the 1000 Genomes Project. We find that a large number of novel regions consistent with soft sweeps, particularly in African populations, and instances of polygenic selection driving the regulatory architecture of several genes. We then use an Approximate Bayesian Computation framework to infer selection parameters for these regions.

1536T

A 3. 4-kb copy number deletion near *EPAS1* is significantly enriched in high-altitude Tibetans but absent from the Denisovan sequence. *S. Xu[1,2,9], H. Lou[1,10], Y. Lu[1,10], D. Lu[1,10], R. Fu[1,2,10], X. Wang[1], Q. Feng[1], S. Wu[3], Y. Yang[3], S. Li[4], L. Kang[5], Y. Guan[1,6], H. B. Peng[7], Y. J. Chung[7], L. Jin[3], B. Su[8], S. Xu[1,2,9].* 1) Max Planck Independent Research Group on Population Genomics, CAS-MPG Partner Institute for Computational Biology (PICB), Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; 2) School of Life Science and Technology, ShanghaiTec University, Shanghai 200031, China; 3) State Key Laboratory of Genetic Engineering and Ministry of Education (MOE) Key Laboratory of Contemporary Anthropology, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai 200433, China; 4) School of Medicine, Xizang University for Nationalities, Xianyang 712082, Shaanxi, China; 5) Department of Biochemistry and Molecular Biology, Preclinical Medicine College, Xinjiang Medical University, Urumqi 830011, China; 6) Faculty of Medicine and Health Sciences, UCSI University, Kualal Lumpur Campus, Jalan Choo Lip Kung, Taman Taynton Vie, 56000, Cheras, Kuala Lumpur, Malaysia; 7) Integrated Research Center for Genome Polymorphism, Department of Microbiology, The Catholic University Medical College, Socho-gu Seoul 137-701, Korea; 8) State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; 9) Collaborative Innovation Center of Genetics and Development, Shanghai 200438, China; 10) These authors contributed equally to this work.

Tibetan high-altitude adaptation (HAA) has been studied extensively and many candidate genes have been reported. Subsequent efforts targeting HAA functional variants, however, have not been that successful; e. g. , no functional variant has been suggested for the top candidate HAA gene, *EPAS1*. With a method developed in this study, WinXPCNVer, we detected in microarray data a Tibetan-enriched deletion (TED) carried by 90% of Tibetans with 50% being homozygous for the deletion, in contrast to only 3% and 0%, respectively, in 2,792 worldwide samples (P<10-15). Long PCR and Sanger sequencing technologies were employed to determine the exact copy number and breakpoints of the TED in 70 new Tibetan and 182 diverse samples. The TED has identical boundaries (chr2:46694276–46697683; hg19) and is 80 kb downstream of *EPAS1*. Notably, the TED is in strong linkage disequilibrium (LD; $r2$=0. 8) with *EPAS1* variants associated with reduced blood concentrations of hemoglobin. It is also in complete LD with the 5-SNP motif, which was suspected to be introgressed from Denisovan, but the deletion itself is absent from the Denisovan sequence. Correspondingly, we detected footprints of positive selection for the TED occurring 12,803 (95% CI: 12,075–14,725) years ago. We further whole-genome deep sequenced (> 60x) seven Tibetans and verified the TED but failed to identify any other copy number variations with comparable patterns, giving this TED top priority for further study. We speculate that the specific patterns of the TED resulted from its own functionality in HAA of Tibetans or LD with a functional variant of *EPAS1*.

1537W

Purifying selection governs the evolution of the major chaperone sub-network in humans. *K. Hess, J. Ellis, N. Nikolaidis.* Department of Biological Science, Center for Applied Biotechnology Studies, and Center for Computational and Applied Mathematics, California State University, Fullerton, Fullerton, CA.

The cellular stress response (CSR) is one of the most conserved and universal mechanisms present in all species. The CSR allows organisms to adapt to their ever-changing environments and survive. The most important orchestrator of the CSR is a group of proteins called molecular chaperones, which regulate protein homeostasis and cell survival. Although the molecular chaperone networks are critical for cell health and disease, their evolution and function in humans remain largely unknown. Here we investigated the microevolution process of an important subset of the human chaperone network, composed by Hsp70s, Hsp40s, and BAGs, by analyzing single nucleotide polymorphisms (SNPs). The results show that first, in 92% of the genes the number of non-synonymous (ns) SNPs is higher than the number of synonymous (s) SNPs. Second, 65% of the genes had a significantly higher SNP density than the surrounding genes. Third, more than 80% of the genes had significantly higher SNP density within their exonic regions as compared to both intronic and un-translated regions. Fourth, the majority of genes contained more sSNPs than nsSNPs within known functional regions (domains), while the number of sSNPs is similar between domains and non-domain protein regions. Fifth, on average 50% of the genes contained a nsSNP on an amino acid position of known function. Sixth, calculations of synonymous and non-synonymous distances revealed the action of purifying selection, the intensity of which varied dramatically both between and within the gene families. To test the predicted effects of some of these nsSNPs we generated recombinant protein variants corresponding to the wild-type and mutated *HSPA1A*, the major human heat-inducible Hsp70 gene. We then used Isothermal titration calorimetry (ITC) and determined that most mutations only marginally altered the binding affinity to both nucleotide (ATP and ADP) and protein substrates. However, one mutation revealed significant alterations in the way the molecules interact and the reaction products are formed. Collectively, these results suggest that strong purifying selection due to functional constraints shaped the evolution of the major chaperone sub-network in humans. The nsSNPs, which are either predicted or showed to alter the function of a particular protein, may either represent adaptations to particular environmental pressures, disease states, or are mutations that have not been eliminated yet from the population.

1538T

Strong evolutionary constraint on human genes escaping X-inactivation is modulated by their expression breadth and level in both sexes. *A. Slavney[1,2], L. Arbiza[2], A. Clark[1], A. Keinan[2].* 1) Molecular Biology and Genetics, Cornell University, Ithaca, NY; 2) Biological Statistics and Computational Biology, Cornell University, Ithaca, NY.

In eutherian mammals, X-linked gene expression is normalized between XX females and XY males through the process of X chromosome inactivation (XCI). XCI results in silencing of transcription from one ChrX homolog per female cell. However, approximately 25% of human ChrX genes escape XCI and exhibit biallelic expression in female cells. The evolutionary basis of this phenomenon is not entirely clear, but high sequence conservation of XCI escapers across mammals suggests that XCI escape is a direct or indirect result of selection rather than neutral mechanisms (such as inefficient establishment or maintenance of XCI). This signal has been suggested to result from these genes' potential contributions to developmental and physiological sex differences, but XCI escape rarely introduces significant female expression biases, which renders this model a poor fit for most XCI escapers. Instead, we hypothesize that XCI escapers experience selection for high and/or broad gene expression in both sexes, which are strong predictors of reduced nucleotide substitution rates in mammalian genes. To test our hypothesis, we compared the extent of purifying selection and gene expression patterns between annotated human XCI escapers and X-inactivated genes. We used human polymorphism data from the Exome Sequencing Project (ESP) to estimate a mutation burden statistic that captures recent/weak purifying selection, and combined these data with divergences between the human and chimpanzee, orangutan, and macaque reference genomes to estimate the extent of ancient purifying selection using a McDonald-Kreitman framework. We calculated gene expression levels and breadth of ChrX genes using RNA-seq data from the Genotype-Tissue Expression (GTEx) consortium, which included expression data for 23 tissues in females and males. When we accounted for the functionality of each ChrX gene's Y-linked homolog (or "gametolog"), we observed that XCI escapers exhibit greater degrees of both recent and ancient purifying selection in the human lineage than X-inactivated genes. In support of our hypothesis, we also observed that they exhibit higher levels of gene expression, as well as broader expression across tissues, than X-inactivated genes, in both sexes. These results propose a significant role for gene expression in both sexes in driving purifying selection on XCI escapers, and emphasize these genes' potential importance in human disease.

1539W

Evolving ancestry: The shift in individual ancestry composition over time. *D. R. Velez Edwards[1,3], T. L. Edwards[2,3], T. L. McGregor[4].* 1) Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN; 2) Division of Epidemiology, Department of Medicine, Vanderbilt University, Nashville, TN; 3) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 4) Department of Pediatrics, Vanderbilt University, Nashville, TN.

An individual's genetic ancestry can be estimated using ancestry-informative markers and subsequently assigning fractions of ancestry. Applications of this information include admixture mapping, controlling for population structure in association studies, and forensic/identification purposes. In observational studies and in clinical encounters, researchers and care providers frequently assign a single ancestry for each individual. We evaluated the genetic ancestry of individuals in a large cohort spanning over 100 birth years by evaluating ancestry informative markers and estimated proportions of racial ancestry with categorical race, as recorded in the electronic health record. Individuals included in this study are members of the BioVU repository, a cohort of DNA samples linked to deidentified electronic health records. Ancestry informative markers were extracted from the Illumina Infinium Human Exome Beadchip on 35,842 individuals. The data were evaluated using STRUCTURE software with the subjects from the 1000 Genomes Project, and the three ancestral clusters corresponding with African, European, and East Asian ancestry were further evaluated in our analyses. We observed that individuals identified as White in the electronic health record have become more genetically admixed over time. The mean fractions of European ancestry have decreased for all cohorts by decade of birth from over 98% through the 1980's to 95. 7% and 90. 0% for the cohorts born in the 2000's and 2010's, respectively. In addition, the number of individuals with less than 70% attributable to any single ancestry has been increasing over time. These data indicate that younger cohorts of individuals exhibit higher levels of heterogeneity than older cohorts, particularly within those identified as White. We also observed increasing heterozygosity over time, with more rapid increases in younger individuals, implying increasing rates of migration of alleles between racial groups over time. Investigators studying traits in younger cohorts should consider these increasing levels of admixture when developing study designs.

1540T

Estimation of inbreeding and substructure levels in African-derived Brazilian *quilombo* populations. *R. B. Lemes, K. Nunes, D. Meyer, R. C. Mingroni-Netto, P. A. Otto.* Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São Paulo, Sao Paulo, SP, Brazil.

Over 3 million Africans were brought as slaves to Brazil during its colonial times. Many runaway, abandoned and freed slaves founded isolated communities (*quilombos*), that in the state of São Paulo are mostly located in its southern border (Vale do Ribeira). We assessed inbreeding and substructure levels from eight *quilombos* by means of analysis of genealogical and molecular data. In order to circumvent problems related to the paucity of historical records and to the heterogeneity of molecular sampling (performed in different groups of individuals with distinct purposes on different occasions), we developed or adapted methods for obtaining reliable estimates of these parameters, since these problems conferred an additional source of heterogeneity among loci and populations. Considering only individuals with more reliable information on his ascendants, from the genealogical analysis of about 2,000 individuals we got an average inbreeding coefficient **FG** = 0. 0025, a value considered low for isolates, but that is 3 to 4 times higher than the estimates for the overall Brazilian population or for the state of São Paulo. We also obtained estimates of the inbreeding coefficient **F** by analyzing a set of 16 microsatellite and 14 SNP autosomal loci (in about 600 and 300 individuals, respectively). Since on the long run **F**-values estimated from different loci are expected to be normally distributed around the average **F**-value for the same population, we developed a method to exclude outlier values exceeding the usual 95% range. The method averages the **F** values by the reciprocals of their variances, which we determined using a novel approximation. The results of this analysis, performed considering only individuals genotyped as to at least 90% of markers, showed that SNPs provide reliable **F** estimates (**F** = 0. 055 for the whole population), while microsatellites do not. Substructure analysis was performed using Wright's indices **FIT**, **FIS** and **FST**, whose statistical significance we tested using computer bootstrap procedures. The results showed that the levels of substructure of the communities here reported are very small, probably due to their proximity and to the migration levels among them over more than 150 years; therefore, the communities here reported can be considered as a single population aggregate. We also compared the estimates obtained using microsatellites to those using a large panel of SNPs, and discuss the possible causes of differences in estimates.

1541W

The Recent History of Human Population Size. *A. Rogers.* Anthropology, University of Utah, Salt Lake City, UT.

Current methods for estimating the history of population size have poor resolution in the recent past. I introduce here a method that combines information from linkage disequilibrium and from the site frequency spectrum. With samples of about 100 diploid genomes, it outperforms existing methods over the past few hundred generations. I validate the method against simulated data and use it with 1000-Genomes data, controlling statistically for the under-representation of rare variants low-coverage data. Results document post-Neolithic population growth in Europe.

1542T

Y-chromosome diversity suggests southern origin and Paleolithic backwave migration of Austro-Asiatic speakers from eastern Asia to the Indian subcontinent. *XM. Zhang[1], SY. Liao[2], XB. Qi[1], JW. Liu[1,8], J. Kampuansai[5], H. Zhang[1], ZH. Yang[3,4], B. Serey[6], T. Sovannary[6], L. Bunnath[6], H. Seang Aun[6], H. Samnom[7], D. Kangwanpong[5], H. Shi[3,4], B. Su[1,4].* 1) Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, China; 2) School of Life Sciences, Anhui University, Hefei 230039, China; 3) Center for Primate Translational Medical Research, Kunming University of Science and Technology, Kunming 650500, China; 4) Yunnan Key Laboratory of Primate Biomedical Research, Kunming 650500, China; 5) Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand; 6) Department of Geography and Land Management, Royal University of Phnom Penh, Phnom Penh 12000, Cambodia; 7) Capacity Development Facilitator for Handicap International Federation and Freelance Research, Battambang 02358, Cambodia; 8) Kunming College of Life Science, University of Chinese Academy of Sciences, Beijing 100101, China.

Analyses of an Asian-specific Y-chromosome lineage (O2a-M95)—the dominant paternal lineage (60. 65% on average) in Austro-Asiatic (AA) speaking populations, who are found on both sides of the Bay of Bengal—led to two competing hypothesis of this group's geographic origin and migratory routes. One hypothesis posits the origin of the AA speakers in India and an eastward dispersal to Southeast Asia, while the other places an origin in Southeast Asia with westward dispersal to India. Here, we collected samples of AA-speaking populations from mainland Southeast Asia and southern China and then analyzed both the Y-chromosome and mtDNA diversities. Combining our samples with previous data, we generated a comprehensive picture of the O2a-M95 lineage in Asia, including both AA and Daic speaking populations. We demonstrated that the O2a-M95 lineage originated in the southern East Asia among the Daic-speaking populations ~20-40 thousand years ago and then dispersed southward to Southeast Asia after the Last Glacial Maximum before moving westward to the Indian subcontinent. This migration resulted in the current distribution of this Y-chromosome lineage in the AA-speaking populations. Further analysis of mtDNA diversity showed a different pattern, supporting a previously proposed sex-biased admixture of the AA-speaking populations in India.

1543W

Historical migrations and mating patterns affect the genetic landscape of the United States population today. *R. Curtis[1], J. Granka[2], J. Byrnes[2], E. Han[2], P. Carbonetto[2], Y. Wang[2], K. Noto[2], M. Barber[2], A. Kermany[2], N. Myres[1], C. Ball[2], K. Chahine[2].* 1) AncestryDNA, Ancestry, Provo, UT; 2) AncestryDNA, Ancestry, San Francisco, CA.

The population of the United States has been formed through a series of foreign migrations, rapid population expansion, and westward migration. However, not all regions of the US have grown in the same way. For example, according to US Census data, most southern states had a birthrate twice as high as states in the northern US during the 1860s, and foreign migration to the US after 1860 largely congregated in the northern states. To learn more about the genetic imprint of these historical differences, we consider inferred identity-by-descent (IBD) sharing patterns among 850,000 genotyped individuals in the AncestryDNA database who are living in the US today. Leveraging over 1. 5 billion pair-wise estimated IBD relationships, along with deep, annotated pedigrees, we use these IBD relationships to construct an IBD network between individuals. We identify two distinct regional patterns in the IBD network that are consistent with known history. First, we see disparity in the number of DNA connections between individuals living in different regions of the country. For example, a DNA sample from Alabama has, on average, nearly double the number of IBD connections as a DNA sample from New York, even though there are 4x more samples with birth locations in New York. Second, we identify, using associated pedigree data, sub-clusters in the IBD network where many individuals have descended from the same ancestor within the past seven generations. These descendant groups vary in size from 2 to over 120 descendants. Again, we find distinct regional differences in these groups. For example, shared ancestors with many identified descendants tend to hail from more geographically or culturally isolated regions of the country, or regions that have undergone rapid population growth. These patterns differ from what would be expected in a random sample from the database, and could be suggestive of certain founder individuals for some regional communities, i. e. people originating in French Quebec or Utah. In a final analysis, we examine the pedigrees in regions with high rates of IBD sharing. We find that individuals in these regions are more likely to share more than one common ancestral couple with their IBD connections, which might explain some of the regional patterns that we have observed. In summary, our analysis shows that IBD connection patterns in different regions of the US are consistent with historically documented mating biases.

1544T

Mitochondrial DNA analysis suggests a Chibchan migration into Colombia. *MC. Noguera[1,2], C. Anderson[4], D. Uricoechea[3], C. Durán[2], I. Briceño[2,3], J. Bernal[2,5].* 1) Medellín, Antioquia, Facultad de Ciencias de la Salud. Grupo Gisafaco. Corporación Universitaria Remington. Medellín, Antioquia, Colombia. Cll 51 51-27; 2) Pontificia Universidad Javeriana. Instituto de Genética Humana. Facultad de Medicina. Cra7 40-62 ed 32; 3) Grupo de Genética Humana, Facultad de Medicina, Universidad de La Sabana. Chía, Colombia. Km. 7, Autopista Norte de Bogotá. Chía, Cundinamarca, Colombia; 4) Department of Foreign Languages & Cultures, Universidad de La Sabana. Chía, Colombia. Km. 7, Autopista Norte de Bogotá. Chía, Cundinamarca, Colombia; 5) Universidad Tecnológica de Bolívar. Cartagena, Colombia. Carrera 21 #25, Cartagena, Bolívar.

The characterization of mitochondrial DNA (mtDNA) allows the establishment of genetic structures and phylogenetic relationships in human populations, tracing lineages far back in time. We analysed samples of mtDNA from twenty (20) Native American populations (700 individuals) dispersed throughout Colombian territory. Samples were collected during 1989-1993 in the context of the program Expedición Humana ("Human Expedition") and stored in the Biological Bank of the Institute of Human Genetics (IGH) at the Pontificia Universidad Javeriana (Bogotá, Colombia). Haplogroups were determined by analysis of RFLPs. Most frequent was haplogroup A, with 338 individuals (48. 3%). Haplogroup A is also one of the most frequent haplogroups in Mesoamerica, and we interpret our finding as supporting models that propose Chibchan-speaking groups migrated to northern Colombia from Mesoamerica in prehistoric times. Haplogroup C was found in 199 individual s (28. 4%), while less frequent were B and D, with 113 and 41 (16% and 6%) individuals, respectively. The haplogroups of nine (9) individuals (1. 3%) could not be determined due to the low quality of the samples of DNA. Although all the sampled populations had genetic structures that fit broadly into the patterns that might be expected for contemporary Central and South American indigenous groups, it was found that haplogroups A and B were more frequent in northern Colombia, while haplogroups C and D were more frequent in southern and south-western Colombia.

1545W

Ancient mitochondrial genetic diversity across time and space in the Iñupiat populations of North Alaska. *J. Tackney[1], A. M. Jensen[2,3,4], W. S. Watkins[5], E. Fair[1], J. Brenner-Coltrain[1], D. Anderson[6], D. H. O'Rourke[1].* 1) Anthropology, Univ Utah, Salt Lake City, UT; 2) UIC Science LLC, Barrow, AK; 3) Anthropology, Bryn Mawr College, Bryn Mawr, PA; 4) Anthropology, University of Alaska Fairbanks, Fairbanks, AK; 5) Human Genetics, University of Utah, Salt Lake City, UT; 6) Anthropology, Brown University, Providence, RI.

It is believed that Iñupiat/Inuit peoples are the descendants of a recent (~700 ybp) and rapid (<100 years) migration across the North American Arctic by the Neo-Eskimo Thule. Archaeological research, and reduced genetic diversity relative to Siberia, suggests a west-to-east migration. Modern arctic peoples also carry a restricted subset of Native American mitochondrial haplotypes, with clade coalescence dates within the Holocene. This supports origins from later migrations separate from the initial peopling of the Americas. Populations across the North Slope, the likely place of origin for the Neo-Eskimo, have been under sampled. Complete mitochondrial sequences are lacking, so clear links with extant mitochondrial diversity are difficult to make. We investigate two pre-contact archaeological sites with human burials. Nuvuk is a long-term Thule village at Pt. Barrow, AK. Radiocarbon dates on archaeological material suggest remains span from Early Thule to modern Iñupiat Eskimo. Igliqtiqsiugvigruaq, a more recent and interior Alaskan site, is located inside the boundary of Kobuk Valley National Park, adjacent to the Kobuk River. We selected three individuals from Nuvuk previously typed for variants associated with D4b1a2a1a(D3) and A2. The rib fragments were directly dated, with median intercept calibrated dates of 1269, 1313, and 1579 AD. We additionally selected three individuals from Igliqtiqsiugvigruaq previously typed as A2b. Radiocarbon, optically stimulated luminescence, and dendrochronology all indicate an age near the turn of the 19th century. Using a targeted mitochondrial capture method and Ion Proton next-generation sequencing, we sequenced high coverage full genomes. Our bioinformatics pipeline and variant calling with the Torrent Variant Caller is identical to current best-practices approaches. Our results allow phylogenetic refinement of haplotypes A2b and D4b1a2a1a, and respective clade coalescence dates. Our archeological and genetic analyses of ancient Iñupiat from two Alaskan North Slope sites demonstrate continuity between prehistoric Neo-Eskimo Thule populations and extant Arctic populations. This research was supported by NSF grants OPP-0732846, OPP-0637246 to DHO'R, OPP-0820790 to JBC, and PLR-1460387, ARC-0726253 to AMJ. Permissions were granted from the Barrow Senior Advisory Council, the Ukpeagvik Iñupiat Corporation, the Native Village of Barrow, the local community of Kiana, and the National Park Service.

1546T

Mitochondrial DNA Haplogroup C Phylogeny for Altaian Populations and its Implications for the Peopling of Siberia and the Americas. *A. Askapuli[1,3], M. C. Dulik[1], S. I. Zhadanov[1,2], L. P. Osipova[2], T. G. Schurr[1].* 1) Department of Anthropology, University of Pennsylvania, Philadelphia, PA 19104-6398, USA; 2) Institute of Cytology and Genetics, SB RAS, Novosibirsk 630090, Russia; 3) Center for Life Sciences, NLA, Nazarbayev University, Astana 010000, Kazakhstan.

Characterization of mitochondrial DNA at a genomic level is very important since it provides opportunities for more accurately estimating the timing and directionality of prehistoric human migrations from a maternal perspective. The Altai Mountains are located at the geographic center of the Eurasian landmass, and have been a hotspot of human activities since ancient times due to itsgeographic location and rich natural resources. Aiming to contribute to a better understanding of the prehistoric human expansions in Siberia and subsequent colonization of the Americas, we sequenced and characterized eighteen whole mtDNA genomes belonging to haplogroup C from Altaian populations. The sequenced Altaian mtDNAs represent all four subgroups of haplogroup C (C1, C4, C5, and C7), and two of them belong to C1a, the Asian sister branch of Native American C1. The Altaian whole mitochondrial sequences were analyzed together with 313 previously published haplogroup C sequences from different parts of the world. The analyses of whole mitochondrial genomes reveal that haplogroup C lineages in Siberia are distributed without any specific association with geography or language, and suggest northeastern Siberia as a place of origin for haplogroup C and its subbranches C1, C4, C5, and C7. The analyses also indicate that Native American haplogroup C types are distantly related with their Siberian sister branches. Given the distribution pattern of haplogroup C in Eurasia, the timing of expansions could be inferred from the age estimates of the lineages within haplogroup C. Age estimation of haplogroup C sequences in our data set via χstatistics shows that haplogorup C has a TMRCA of 31. 25 kyr (24. 13-38. 56), and its subbranches C1, C4, C5, and C7 have TMRCAs of 21. 64 kyr (16. 83-26. 55), 24. 88 kyr (16. 65-33. 41), 19. 76 kyr (13. 63-26. 08), and 27. 2 kyr (16. 69-38. 17), respectively. Still, it is almost impossible to pinpoint geographic origin of Native Americans and directionality of prehistoric migrations in Siberia with certainty. Based on the results of the current study, the Amur region in northeastern Siberia could be the geographic origin for ancestral Native Americans. In order to obtain clearer picture of human population movements in Siberia and the Americas from a maternal perspective, more mitochondrial genomes need to be sequenced, especially mitochondrial genomes belonging to the relatively diverse haplogroups C and D.

1547W

Structured mating and its genetic consequences in the multi-ethnic Kaiser Permanente (KP) Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Y. Banda[1], T. Hoffmann[1,3], M. Kvale[1], L. Shen[2], D. Ranatunga[2], E. Jorgenson[2], P. Kwok[1], C. Schaefer[2], N. Risch[1,2,3].* 1) Institute Human Genetics, University California San Francisco, San Francisco, CA; 2) Kaiser Permanente Northern California Division of Research, Oakland, CA; 3) Department of Epidemiology and Biostatistics, University of California, San Francisco, CA.

Population stratification is influenced by patterns of racial/ethnic/national endogamy versus exogamy, as well as ancestry-related assortative mating within racial/ethnic groups. We analyzed mating patterns and genetic ancestry between and within the European, East Asian, Latino, African American, and South Asian participants in the KP GERA cohort of 103,000 subjects. Of 7892 total spouse pairs identified from electronic records, 6785 were concordant for race/ethnicity (87 African American, 321 East Asian, 6225 European, 150 Latino, 22 South Asian), while 1087 were discordant. According to random mating, 2066 discordant pairs were expected. The largest excess of concordant pairs occurred within South Asians and African Americans, followed by East Asians and Latinos. In the principal components analysis of genetic ancestry in the concordant pairs, within the Europeans, Ashkenazi subjects formed a separate cluster and induced a modest correlation in PC1 and PC2 between the spouses (0. 36, P<1e-16 for PC1 and 0. 17, P<1e-16 for PC2). After removing the Ashkenazi subjects, the correlation between spouses for PC1 and PC2 fell to 0. 18 (P<1e-16) and 0. 12 (P<1e-16), respectively. The first PC corresponds to a northwest-southeast European cline, and the second PC to a southwest-northeast cline. For the East Asian pairs we observed a correlation of 0. 35 (P<1e-16) in PC1 (representing European versus East Asian ancestry) and a correlation of 0. 77 (P<1e-16) in PC2 (representing a north-south cline in East Asia). The correlation in PC1 largely reflects Filipino spouse pairs, while the high correlation in PC2 reflects extensive endogamy among East Asian nationalities. Among Latinos, we observed spouse correlations of 0. 25 (P=0. 0021) and 0. 24 (P=0. 0038) for the first and second PCs, reflecting European and African versus Native American ancestry, respectively. These correlations reflect both concordance of nationality and ancestry related assortative mating. For the African Americans, the correlation in PC1 was 0. 32 (P=0. 0026), representing African versus European ancestry. Genetic consequences of structured mating observed in this cohort were apparent in that SNPs that loaded most heavily on PCs reflecting genetic ancestry exhibited both an excess of homozygosity and linkage disequilibrium. The very complex nature of structured mating we observed suggests a need for its careful consideration when designing and interpreting population-based genetic studies.

1548T

A comprehensive, diachronic and comparative picture of the mitogenome variation along the Americas. *A. Achilli[1], U. A. Perego[1,2], A. Owings[3,4], H. Lancioni[1], A. Olivieri[5], I. Cardinali[1], M. R. Capodiferro[1], V. Battaglia[5], S. Brandini[5], A. Fichera[5], S. R. Woodward[2], O. Semino[5], J. R. Johnson[6], E. Willerslev[7], M. Stoneking[8], A. Torroni[5], R. S. Malhi[3,4,9].* 1) Department of Chemistry, Biology and Biotechnology, University of Perugia, Perugia, Perugia, Italy; 2) Sorenson Molecular Genealogy Foundation, Salt Lake City, UT 84115, USA; 3) Program in Ecology, Evolution and Conservation Biology, University of Illinois Urbana-Champaign, USA; 4) Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, IL 61801, USA; 5) Department of Biology and Biotechnology "Lazzaro Spallanzani", University of Pavia, Italy; 6) Department of Anthropology, Santa Barbara Museum of Natural History, Santa Barbara, CA 93105; 7) Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, DK-1350 Copenhagen K, Denmark; 8) Max Planck Institute for Evolutionary Anthropology, Department of Evolutionary Genetics, Deutscher Platz 6, D-04103 Leipzig, Germany; 9) Department of Anthropology, University of Illinois Urbana-Champaign, USA.

Native Americans present a remarkable case study in human evolution and population genetics. They belong to one of the few extant human groups whose ancestors entered a vast uninhabited area over a relatively short interval and then apparently remained relatively isolated from other human groups for a considerable period of time before European contact. Their substantial cultural diversity, linguistic complexity, and biological variation has been the subjects of a plethora of studies, whose level of sophistication has consistently grown over the past 40 years, but to date has not yet seen a final consensus among the disciplines involved. After all these years, this debate is still ongoing among scientists regarding major issues such as the number of migratory events, the source populations and arrival timing, and the likely entry routes into the Americas. Even the archaeological evidence is inconsistent as the dating of skeletal remains and Clovis lithic artifacts yielded values of 13-14,000 years before present (YBP), while other archeological studies reported sites with dating to more than 33,000 YBP. The overall picture is gradually emerging much clearer and detailed thanks to the contributions provided by human genetics and genomics studies. The first pieces were provided by analyses of the so-called "classical" genetic markers; subsequent findings came from the phylogenetic surveys of uniparentally transmitted genetic systems, in particular from mitochondrial DNA; then in the last few years, further valuable details were delivered by a new phase characterized by ancient DNA analyses.

Here, we present a comprehensive, diachronic and comparative picture of the mitogenome variation along the double continent. More than 1600 entire mitochondrial DNAs from modern samples (were gathered from literature and unpublished databases and compared with ca. 90 novel ancient pre-European contact mitogenomes (mostly from North America) in order to reconstruct the migratory path(s) of ancient native populations and to evaluate the differential demographic impact of post-European contact in South and North America.

1549W

The relative effective population size of chromosome X and the autosomes along distinct branches of the human population tree. *L. Arbiza, A. Keinan.* Biological Statistics and Computational Biology, Cornell University, Ithaca, NY.

In recent years, many studies have focused on the effective population size of chromosome X relative to the autosomes. This comparison can be useful to reveal past demographic processes, differences in the histories of males and females, and the action of natural selection. We have recently shown how the ratio of nucleotide diversity between the two (X-to-Autosome ratio; X/A), when compared between pairs of populations (relative X/A), can be used to uncover sex-biased processes in human history. While this strategy serves to alleviate the response of genetic diversity to the influence of events in a time range that largely predates the split of the studied populations, a different and more natural approach to capture recent changes occurring after populations split can be formulated based on the differentiation of allele frequencies between populations, as commonly summarized by the FST statistic. Here, we consider population differentiation in humans, and extend beyond simple pairwise comparisons, using allele frequency differences across several populations to learn about the ratio of X-to-autosomal effective population size along distinct branches in the tree of human populations. We then test these for differences from the expectation of equal female-to-male breeding ratios, as well as differences between different branches. Using coalescent simulations of a variety of previously published human demographic models, we show that our approach is able to capture the ratio of interest and is more accurate than estimates based only on pairwise FST across all pairs of populations. We then turn to the latest data from the 1000 Genomes Project, controlling for the effect of uncertainty associated with low coverage sequencing, as well as the influence of linked selection (background selection or hitchhiking), all of which differentially affect the X chromosome and the autosomes. Estimating the X-to-autosomal effective population size ratio for branches leading to different 1000 Genomes populations, as well as for internal branches in the population tree, points to a higher female effective population size in African-specific population history, but not in non-Africans. More interestingly, we localize previously-debated observations to a significant increase in male effective population size on the branch leading to all non-African populations, suggesting male-biased processes associated to the Out-of-Africa event.

1550T

Accurate non-parametric estimation of recent effective population size from segments of identity by descent. *S. R. Browning[1], B. L. Browning[2].* 1) Dept Biostatistics, Univ Washington, Seattle, WA; 2) Dept Medicine, Medical Genetics division, Univ Washington, Seattle, WA.

Existing methods for estimating historical effective population size from genetic data have been unable to accurately estimate effective population size during the most recent past. We present a non-parametric method for accurately estimating recent effective population size using inferred long segments of identity by descent (IBD). We find that inferred segments of IBD contain information about effective population size from around 4 generations ago to around 50 generations for SNP array data and to over 200 generations for sequence data. In human populations that we examine, the estimates of effective size are approximately one-third of the census size. We estimate the effective population size of European-ancestry individuals in the UK 4 generations ago to be 8 million and the effective population size of Finland 4 generations ago to be 0. 7 million. Our method is implemented in the open-source IBDNe software package.

1551W

Past demography revealed from rare variants with population specific F_{ST} values. *J. Goudet[1,3], B. S. Weir[2].* 1) Department of Ecology and Evolution, Biophore, University of Lausanne, Lausanne, Vaud, Switzerland; 2) Department of Biostatistics, University of Washington, Seattle, WA; 3) SIB-Swiss Institute of Bioinformatics, Unil-Sorge, Lausanne, Vaud, Switzerland.

Statement of PurposeUnderstanding the past demography of humans is essential in many areas of human genetics, in particular because many demographic processes leave genomic signatures identical to those that can be left by selection. Computer intensive methods such as approximate Bayesian computation of exact likelihoods have been proposed to this end, but they have difficulty with the magnitude of current data. Instead, we propose the use of population specific FST values, and show that, in conjunction with the frequencies of derived alleles, these measures allow the recovery of several aspects of the past demography. Methods usedWe estimate population specific FST from the proportion of pairs of alleles within a population that are identical in state (ibs) relative to the unweighted proportion of ibs pairs of alleles between all pairs of populations within a study. For SNPs, we show the parametric form of population-specific FST as a function of the derived allele frequencies under classical models of population structure and we use computer simulation to verify our predictions. We use 1000 Genome Phase 3 data for African and South Asian samples to illustrate how the past demography can be recovered. Summary of resultsWe show, both in simulations and with 1000 genomes data, that population specific FST as a function of the derived allele frequencies shows the signature of past demographic events: a past population split such as South Asian splitting from African, followed by population expansion results in strongly negative FST for Africa, for SNPs with a low frequency of derived alleles, mirrored by strongly positive values for Asia. The negative FST values for rare variants reflects the lower matching across alleles within the African sample than between the African and South Asian samples, and this signal is missed by the usual population average FST. The following demographic estimates result: assuming an effective population size for Africa of 10,000 individuals, dispersal of 1,780 individuals out of Africa occurred 44,000 years ago (25 years/ generation); South Asia grew to 10,900 individuals 4,700 years ago, and Africa and South Asia kept exchanging seven migrants per generation during this time. These estimates, while in broad agreement with those already published, point to a more recent expansion from Africa.

1552T

Historical mating patterns in the U. S. revealed through admixture and IBD patterns from genome-wide data from over 800,000 individuals. *J. M. Granka[1], Y. Wang[1], E. Han[1], J. K. Byrnes[1], A. Kermany[1], R. E. Curtis[2], P. Carbonetto[1], K. Noto[1], M. J. Barber[1], N. M. Myres[2], C. A. Ball[1], K. G. Chahine[2].* 1) AncestryDNA, San Francisco, CA; 2) AncestryDNA, Provo, UT.

Within a diverse population like the United States, many individuals are admixed, with ancestry from many worldwide regions. Non-random mating and migration can result in non-random combinations of ancestries within admixed individuals (i. e. , certain sets of ancestries may be common, and others may be rare); such dynamics can also affect patterns of identity-by-descent (IBD) among admixed and non-admixed individuals. To shed insight into historical mating and migration, we study genome-wide genotype data of over 800,000 AncestryDNA customers, as well as a subset of over 400,000 born in the US. First, we use a supervised algorithm to estimate individuals' genetic admixture proportions across 26 global regions. We measure correlations between the estimated ancestries, and find certain sets of ancestries to frequently co-occur in individuals' estimates. Such relationships may reflect historical events; e. g. , the association between ancestry from the Americas and the Iberian Peninsula could reflect Colonial Era admixture. In addition to historical mating patterns, however, the admixture inference procedure and the delineation of global regions could also impact such correlations. To disentangle whether these trends could reflect mating patterns and preferences, we examine associations between the estimated ancestries of the parents of over 10,000 trios. Observed correlations agree with many of those identified within individuals, and potentially reflect more recent historical trends. Thirdly, we extend our study to IBD patterns in an inferred IBD network among genotyped individuals. Sub-clusters of the IBD network, which can often be annotated by ethnicity or historical US migration, are often inter-connected by bridging IBD connections; we highlight several connected sub-clusters in light of findings from genetic ancestry. Finally, we corroborate findings from these three analyses, as well as their potential timescales, by examining over 500,000 AncestryDNA customer pedigrees. Associations of country-level birth locations between pairs of couples support many of the non-random associations of ethnicities and IBD connections identified using genetic data. Many of the associations we observe reflect historical phenomena, and while not conclusive about their cause, suggest that many individuals with admixed ancestry, including those in the US, have present-day genetic signatures reflecting the migration and subsequent non-random mating of their ancestors.

1553W

Discovery of a previously unknown ancestral origin of the modern Taiwanese population that is distinct from the north-south gradient seen in other Han Chinese populations using the Taiwan Biobank. *C. W. Lin[1,2], C. H. CHEN[1,2,3,4], J. H. YANG[3], P. E. WU[1,2], C. N. HSIUNG[1,2], L. C. CHANG[3], J. CHANG[2], I. W. SONG[2], S. L. YANG[2], F. T. LIU[1,2], C. Y. SHEN[1,2,5].* 1) Taiwan Biobank, Academia Sinica, Taipei, Taiwan; 2) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 3) National Center for Genome Medicine, Academia Sinica, Taipei, Taiwan; 4) School of Chinese Medicine, China Medical University, Taichung, Taiwan; 5) College of Public Health, China Medical University, Taichung, Taiwan.

The aim of the Taiwan BioBank is to build a nationwide biomedical research database that integrates genomic profiles, lifestyle patterns, dietary habits, environmental exposure histories, and long-term health outcomes of 300,000 Taiwanese residents (representing almost 1. 5% of the Taiwanese population). We describe here results from 8265 samples that were genotyped using the Taiwan BioBank array, which was specifically designed for the Taiwanese population. After data quality control, genotype data for 589,016 single-nucleotide polymorphisms (SNPs) in 7203 unrelated individuals were denoted as TWB7203 and further analyzed. The 7203 individuals were clustered into three cline subgroups: 4. 5% were of northern Han Chinese ancestry, 77. 6% were of southern Han Chinese ancestry, and 17. 8% were an admixture of Han Chinese and a previously unknown ethnic group. This unknown group was genetically distinct from neighboring southeast Asian groups and Austronesian tribes, but was similar to the southern Han Chinese. Long-range linkage disequilibrium and flips of major alleles at about 400 SNPs across the major histocompatibility complex region suggested that the previously unknown group may have experienced evolutionary events different from those of the other southern Han Chinese. The difference was further supported by the unique pattern of body figures measures of this unknown group. Genome-wide summary statistics for the ethnic subgroups of TWB7203 were released through a publicly accessible web-based calculation platform, Taiwan View (http://taiwanview. twbiobank. org. tw/taiwanview/twbinfo. do), on which genome-wide association analyses can be performed using TWB7203 as the reference. The release of this large-scale population-level and subpopulation-level genomic information will greatly benefit human genetic research.

1554T

Estimation of growth rates for populations and haplogroups using full Y chromosome sequences. *F. L. Mendez, G. D. Poznik, C. D. Bustamante, 1000 Genomes Project Consortium.* Department of Genetics, Stanford University, Stanford, CA.

Evolutionary processes affecting a population influence gene genealogies across the genome. Coalescent theory provides the mathematical framework to connect realized genealogies to the underlying evolutionary processes. However, in most cases, information about the genealogies is obtained only indirectly through the observation of genetic variation. Therefore, in general, very limited information about any individual locus is available. As the longest non-recombining portion of the human genome, the Y chromosome accumulates mutations relatively quickly. When large amounts of sequence are used, the Y chromosome provides an unparalleled ability to resolve the structure and coalescence times of its genealogy. Because patterns of variation in the Y chromosome are only influenced by processes affecting men, they can be used to study both demographic and social phenomena. The 1000 Genomes Project includes whole Y-chromosome data from more than 1000 men and has an extensive representation of most lineages that have experienced recent massive expansions in size. Though the dynamics of population growth have likely changed over time, we are more interested in the growth rates at the times of these rapid expansions than on an average effect. To study this, we have developed a new method that takes advantage of the temporal resolution provided by Y-chromosome data and of historical data, while accounting for the uncertainties associated with the coalescent and mutational processes. We estimate the growth rates for several branches of the Y-chromosome tree, including those in Europe, sub-Saharan Africa and South Asia. We estimate that several lineages within the European R1b, sub-Saharan African E1b, and South Asian R1a haplogroups experienced growth rates of at least 20-60% per generation at the onset of their massive expansions, some 3-5 thousand years ago. These high growth rates are comparable to those experienced by human populations during the 20th century. However, we find that most observed genealogies are unlikely to be the result of whole population expansion or of natural selection.

1555W

Bayesian Nonparametric Inference of Population Size Changes from Sequential Genealogies. *J. Palacios[1,2,3], J. Wakeley[1], S. Ramachandran[2,3].* 1) Organismic and Evolutionary Biology, Harvard University, Cambridge, MA; 2) Ecology and Evolutionary Biology, Brown University, Providence, RI; 3) Center for Computational Molecular Biology, Brown University, Providence, RI.

Sophisticated inferential tools coupled with the coalescent model have recently emerged for estimating past population sizes from genomic data. Accurate methods are available for data from a single locus or from independent loci. Recent methods that model recombination require small sample sizes, make constraining assumptions about population size changes, and do not report measures of uncertainty for estimates. Here, we develop a Gaussian process-based Bayesian nonparametric method coupled with a sequentially Markov coalescent model which allows accurate inference of population sizes over time from a set of genealogies. In contrast to current methods, our approach considers a broad class of recombination events, including those that do not change local genealogies and uses all sufficient statistics needed to estimate population size changes over time. First, we show that under the sequentially Markov coalescent model, the set of local Tajima's genealogies consisting of ranked tree shapes and coalescent times, are sufficient statistics for inferring the population size trajectory. Second, we show that our method outperforms recent likelihood-based methods that rely on discretization of the parameter space. We illustrate the application of our method to multiple demographic histories, including population bottlenecks and exponential growth. In simulation, our Bayesian approach produces point estimates four times more accurate than maximum likelihood estimation (based on the sum of absolute differences between the truth and the estimated values). Further, our method's credible intervals for population size as a function of time cover 90 percent of true values across multiple demographic scenarios, enabling formal hypothesis testing about population size differences over time. Using genealogies estimated with ARGweaver, we apply our method to European and Yoruban samples from the 1000 Genomes Project and confirm key known aspects of population size history over the past 150,000 years.

1556T

Modeling SNP array ascertainment with Approximate Bayesian Computation to improve demographic inference. *C. Quinto[1], A. Woerner[1], K. Veeramah[2], M. Hammer[3].* 1) Genetics GIDP, University of Arizona, Tucson, AZ; 2) Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY; 3) Arizona Research Laboratories, Division of Biotechnology, University of Arizona, Tucson, AZ.

Single nucleotide polymorphisms (SNPs) in commercial genotyping arrays have often been discovered by sampling a small number of chromosomes from a group of selected populations. This form of non-random ascertainment skews patterns of nucleotide diversity, and can affect population genetic inferences. In addition, the effect of ascertainment bias is particularly strong when the populations of interest are genetically distinct from the ascertainment panel. Although different methods have been proposed to take into account this bias, the challenge remains because the exact discovery protocols are not known for most of the commercial arrays.

In this work, we propose a demographic inference pipeline that explicitly models the SNP discovery protocol in an Approximate Bayesian Computation (ABC) framework. We simulate genomic regions according to a demographic model that includes Yoruba, Han Chinese and European individuals, and recreate the SNP distribution of a commercial chip (Affymetrix Axiom Genome-Wide Human Array) by varying the number of sampled chromosomes and the allele frequency cut-off in the given regions. We then calculate site frequency spectrum and haplotype based summary statistics obtained from both the ascertained and genomic data and infer the ascertainment and demographic parameters of the model using ABC. To validate this method, we simulate summary statistics with known parameter values and treat them as observed data. We are able to successfully recover the true parameter values from the simulations in the 95% High Posterior Density Interval. Finally, as proof of concept, we implement our pipeline to study a real example of admixture, likely affected by ascertainment bias. We model the admixture process in the Mexican population, replicating the SNP distribution given by the inferred ascertainment protocol. Ancestry blocks are identified in admixed whole genomes and its distribution is used to estimate the time and proportion of admixture in an ABC analysis. Our inference framework is applicable to studies when only SNP chip data of the populations of interest are available and will improve to accuracy of the demographic model inference.

1557W

Selective constraint and sex-biased demography of human populations from X chromosome-autosome comparisons. *M. H. Quiver[1], J. Lachance[1], K. Mullen[2], M. E. B. Hansen[2], M. A. Chen[2], P. H. Hsieh[3], K. R. Veeramah[4], S. A. Tishkoff[2].* 1) School of Biology, Georgia Institute of Technology, Atlanta, GA; 2) Departments of Biology and Genetics, University of Pennsylvania, Philadelphia, PA; 3) Department of Molecular and Cell Biology, University of Arizona, Tucson, AZ; 4) Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY.

Because the number of X chromosomes differs for men and women, comparisons between sex-linked and autosomal genetic loci reveal sex-biased patterns of human demography. Using 44 high-coverage whole genomes from a diverse global set of 11 human populations we quantified the strength of selective constraint on different chromosomes, found evidence of sex-biased colonization, and determined whether recent migrations are matrilocal or patrilocal. Relative amounts of genic and intergenic diversity were similar across all studied populations regardless of subsistence pattern or geography. The strength of selective constraint on genes was greater for X-linked loci compared to autosomal loci - a pattern that is consistent with selection against deleterious recessive alleles. The ratio of X chromosome to autosome diversity (Q) was greater than the null expectation of 0. 75 for African populations and less than 0. 75 for non-African populations, with lower values of Q for populations located farther from Africa. This pattern is consistent with a male-biased serial founder effect model, and computer simulations suggest a plausible out-of-Africa bottleneck size of 320-340 males and 60-70 females. Using PSMC, we found evidence of large historic population sizes for West African Pygmies, but not Hadza or Sandawe populations. Genetic distances revealed female-biased gene flow between Hadza and Sandawe hunter-gatherers, between Maasai pastoralists and African farmers, and between Chinese and Japanese populations. We found evidence of male-biased gene flow between African farmers and hunter-gatherers, and between different African farmer populations. This calls into question the idea that patrilocality is coupled with the emergence of agriculture.

1558T

Genetic, Geographic and Cultural Reconstruction of an Ancient Endogamous Community. *D. K. Sanghera[1], A. Raina[2], C. E. Aston[1], D. D. Mascarenhas[3].* 1) Pediatrics, University of Oklahoma HSC, Oklahoma City, OK ,USA; 2) All India Institute of Medical Sciences, New Delhi, India; 3) Mayflower Organization for Research and Education, Sunnyvale, CA, USA.

The provenance of a rare R1a1 Y-haplogroup (Y-HG) subtype designated as 657A lies in proximity to anancient migration route running through Afghanistan but is largely absent from other geographic locations. A clan of 657A Brahmin “founder” family lineages within the Goud Saraswat community (GSB) in a town in Western India wasidentified in which 15 of 16 males from nine families were R1a1 Y-HG, including 10 who were 657A. TMRCAcalculations using pairwise comparisons to control cohorts suggested a probable migration history for this priestlysubgroup. To support this genetic narrative we present archeological, toponymic, numismatic, linguistic,iconographic, architectural, sociological and literary data. Specifically, in this study we test two main hypothesesregarding these 657A families: (1) Using Y-HG centroid analysis, chi-square analysis of TMRCA distributions andarcheological find-spots, and discriminant function analysis we show that the parental Z93 L342. 2 sub-clade in which657A occurs originated in West Asia and that 657A individuals migrated toward the southeast by a Bolan Pass routedistinct from the traditionally presumed route of “Vedic” ingress into the Indian subcontinent; and (2) Priestly 657Alineages in Western India retain distinct family practices with respect to literacy, religious practice and migration notshared by other more orthodox Brahmins of canonical geographic origin within the same community, despiteintermarriage. Long-term transmission of differentiated family practices within a single patrilineal endogamouscommunity has rarely been documented.

1559W

Unique autosomal STR heterogeneity among three traditional populations from historical Tibet. *T. Tsering[1], T. Gayden[2], S. Chennakrishnaiah[3], A. Bukhari[4], R. Herrera[5].* 1) Tata Institute of Fundamental Research, Mumbai, India, 400 005; 2) Department of Human Genetics, McGill University, Montreal, Canada; 3) Department of Experimental Medicine, McGill University, Montreal, Quebec, Canada; 4) College of Medicine, Florida International University, Miami, Florida, USA 33199; 5) Department of Molecular Biology, Colorado College, Colorado Springs, Colorado, USA.

Reaching an average altitude of more than 4000 meters above sea level, the Tibetan plateau is the highest and largest highland on the planet and is rightly called the ‘roof of the world’. It is bordered on three sides by huge cordilleras: the Kunlun, the Karakoram and the Himalayas. These three massifs along with high elevation have kept Tibet relatively isolated throughout its long history and prehistory. However, given its strategic location in Eurasia, the region also served as a corridor for human migrations, including the ancient Silk Road. While several studies on Tibetan populations have been previously reported, none of them have comprehensively characterized the genetic population structures of the three major culturally defined regions of historical Tibet: Amdo, Kham and U-Tsang. Thus, in this study, we genotyped 15 highly polymorphic autosomal STR loci using the AmpFLSTR Identifiler System in 338 unrelated individuals from Amdo (86), Kham (101) and U-Tsang (151). All of the analyzed loci were in agreement with Hardy-Weinberg equilibrium expectations, with the exception of the locus D19S433 in the Kham province. FGA is the most informative marker with the highest observed heterozygosity and gene diversity values across the three Tibetan populations. Conversely, TH01 is the least variable locus displaying the lowest values for the same two parameters. In addition, FGA displays four 2 bp microvariant alleles (20. 2, 21. 2, 22. 2 and 25. 2) that are distributed between U-Tsang and Kham populations but are completely absent in the Amdo people. U-Tsang displays the highest total numbers of alleles (139) trailed by Kham (130) and Amdo (128) groups. The allele frequency data from this study were compared to relevant global reference populations. Our results suggest that although these three Tibetan groups segregate within the cluster of the Himalayan populations in both the Correspondence Analysis (CA) plot and the Neighbor Joining (NJ) tree, they exhibit some degree of genetic differentiation among themselves. The genetic uniqueness of the three historically Tibetan populations is congruent with their characteristic dialect, culture and tradition. In conclusion, the 15 autosomal STR loci examined provides a significant dataset for forensic applications, parentage testing and population genetic studies.

1560T

Geographic patterns of Identify-by-descent recapitulate fine-scale migration history of the African Americans. *Y. Wang, M. Barber, J. Byrnes, P. Carbonetto, R. Curtis, J. Granka, E. Han, A. Kermany, N. Myres, K. Noto, C. Ball, K. Chahine.* Ancestry. com, San Francisco, CA.

African Americans constitute the third largest ethnic group in the United States. The population structure and migration history of this group has drawn great interest. Recently, Simon Gravel (Gravel et al. , ASHG 2014) presented a population genetics study using 6,000 genotyped African American samples, in which he and colleagues identified signature of recent south-to-north migrations, possibly due to an event known as The Great Migration. Unfortunately the signal they found was only marginally significant, likely due to a limited sample set size. Here we present a large-scale study of the African American migration history, using genotyping data from over 770,000 consenting AncestryDNA customers. First we predict the genetic ancestry of all samples using ADMIXTURE (Alexander*et al. ,*2009) with a high-quality reference panel of 3,000 samples. We select samples with estimates of more than 50% sub-Sahara African ancestry for further analysis. This leads to a subset of over 50,000 African Americans representing all 50 U. S. states. In a similar approach, we also select more than 600,000 samples with estimates of over 90% European ancestry. Examining the identity-by-descent (IBD) patterns among these African American individuals, as well as self-reported birth locations, we identify a clear south-to-north migration pattern, distinct from the east-to-west pattern present in individuals with primarily European ancestry. Furthermore, our analysis reveals three distinct migration routes by which African Americans left the South. The first route goes along the Atlantic coast from South Carolina to Pennsylvania and New England. The second route starts from Alabama and Mississippi and disperses into the Great Lakes region. And the last route extends from Louisiana and Texas to the west. These routes agree with historical records revealing three separate migratory patterns during the Great Migration. In contrast, we observe different patterns when we look at IBD shared between African Americans and European Americans. We find that African Americans tend to share more IBD with European Americans born in states, such as South Carolina, Georgia, Virginia and Louisiana, where the major transatlantic slave trade ports were located. Our results suggest a large part of the European-African admixture dates back long before the Great Migration.

1561W

RECONSTRUCTING GENETIC HISTORY OF SIBERIAN AND NORTHEASTERN EUROPEAN POPULATIONS. *E. Wong[1], A. Khrunin[2], L. Nichols[2], D. Pushkarev[3], D. Khokhrin[2], D. Verbenko[2], O. Evgradov[4], J. Knowles[4], J. Novembre[5], S. Limborska[2], A. Valouev[1].* 1) Department of Preventive Medicine, Keck School of Medicine of USC, Los Angeles, CA; 2) 2. Department of Molecular Bases of Human Genetics, Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, Russian Federation; 3) Illumina, Inc. , Advanced Research Group, San Diego, CA, USA; 4) 4. Department of Psychiatry and Behavioral Sciences, Keck School of Medicine, Zilkha Neurogenetic Institute, University of Southern California, CA, USA; 5) Department of Human Genetics, University of Chicago, Chicago, IL, USA.

Siberia and Western Russia are home to some of the least studied ethnic groups in the world, and their genetic history holds keys to understanding peopling of the world. We present whole-genome sequencing data from 28 individuals belonging to 14 distinct indigenous populations from that region. We used these datasets together with an additional 32 modern-day and 15 ancient human genomes to build and compare autosomal, Y-DNA and mtDNA trees and delineate genetic history. Our analyses uncover complex migratory processes that shaped the genetic landscapes in Asia and Europe. Admixture events between ancient Siberian groups resulted in distinct ancestries of nowadays Western and Eastern Siberians. Western Siberians share genetic affinity with modern Europeans. Both can trace their ancestry to the lineage of a 24,000-year-old Siberian Mal'ta boy. For Eastern Siberians, they have much weaker genetic affinity with Europeans and their ancestor separated from East Asians much later (approximately 10,000 years ago). Major migration wave from Eastern Siberians into Western Siberian groups occurred approximately 7,000 years ago, and it extended into Northeastern Europe. This is based on the admixtures we observed between Siberians and lineages represented by the 5,000-year-old hunter-gatherer Ire8 from Pitted Ware Culture excavated in Sweden, the 2,900-year-old Iron age Hungarian IR1 from the Mezocsat Culture, and modern-day northeastern Europeans. Our whole-genome data based on a broad sample of populations in Siberia and Western Russia provides new insights at a high-resolution into the genetic history of Eurasians.

1562T

Ages of mitochondrial DNA lineages coincides with the agriculture spread in Finland. *S. Översti[1], P. Onkamo[1], J. Palo[2].* 1) Department of Biosciences, PO Box 56 (Viikinkaari 5) FI-00014 University of Helsinki, FINLAND; 2) Laboratory of Forensic Biology, Department of Forensic Medicine, Hjelt Institute, PO Box 40 (Kytösuontie 11) FI-00014 University of Helsinki, FINLAND.

The current inhabitants of Finland in the Northeastern Europe are quite unique in terms of their genetic composition. Based on Y chromosomal and genome wide studies Finns differ from other European populations: especially the Y chromosomal diversity is reduced and distinctive. In contrast, Finnish mitochondrial DNA (mtDNA) haplogroup distribution is similar to other European populations. Mitochondrial genepool in modern Europeans is a mixture of Mesolithic hunter-gatherer associated haplogroups (U and V) and Neolithic associated farmer haplogroups (H, J, K and T). The frequency of hunter-associated haplogroup U in Finland is one of the highest in Western Eurasia. Also, it is more common in Eastern and Northern parts of the country while farmer haplogroups are more frequent in Southern and Western Finland. In this study we compiled a comprehensive data set of 833 modern Finnish complete mtDNA sequences from the public databases and utilized coalescent based Bayesian phylogenetic inference (BEAST v. 1. 8. 1) to perform fine resolution phylogenetic analyses on the sequences. We also exploited previously published radiocarbon dated ancient complete mtDNA sequences from Western Eurasia in our analysis as calibration points to the phylogenetic trees, enhancing their accuracy. Our results demonstrate that among Finns, many typically “European” haplogroups, both hunter-gatherer and farmer associated, actually comprise lineages specific for Finns. Several of these lineages, despite being rather common in present Finnish population, are virtually absent from other populations. Oldest of these haplogroups date back over 7,000 years, though most appear to be around 3,000-5,000 years old. This period temporally coincidences with the arrival and especially the spreading of the agriculture and Corded Ware culture in Finland. Age estimates are also concurrent with the arrival of another culture, the textile ceramics, into Finland from Volga region (main period of textile ceramics lies between 1,700-1,000 BC). According to these results there is distinct evidence that arrival of these cultural entities also influenced Finnish mitochondrial DNA pool and this impact is still visible in modern day Finns.

1563W

Fine scale population structure of Spain and the genetic impact of historical invasions and migrations. *C. Bycroft[1], C. Fernandez-Rozadilla[1,2], A. Carracedo[2], C. Ruiz-Ponte[2], I. Quintela-García[3], P. Donnelly[1,4], S. Myers[1,4].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford; 2) Galician Public Foundation of Genomic Medicine (FPG-MX)-Grupo de Medicina Xenómica-Centro de Investigación BioMédica en Red de Enfermedades Raras (CIBERer)-Universiy of Santiago de Compostela, Spain; 3) Grupo de Medicina Xenomica, Universidade de Santiago de Compostela, Centro Nacional de Genotipado - Plataforma de Recursos Biomoleculares y Bioinformaticos - Instituto de Salud Carlos III (CeGen-PRB2-ISCIII); 4) Department of Statistics, University of Oxford, 1 South Parks Road, Oxford, OX1 3TG, UK.

As well as being linguistically and culturally diverse, the Iberian Peninsula is unusual among European regions in that its demographic history includes a prolonged and large-scale occupation by people of predominately north-west African origin. Therefore, the Iberian Peninsula provides a unique opportunity for studying fine-scale population structure and admixture, and to test cutting-edge methods of detecting complex or subtle population genetic patterns. Previous studies using Y-chromosome, mtDNA as well as autosomal data have detected limited genetic structure in Iberia. However, powerful new methods and larger datasets mean it has recently become possible to detect and characterise genetic differentiation at a sub-national level. We performed the largest and most comprehensive study of Spanish population structure to date by analysing a dataset of ~1,400 Spanish individuals typed at ~700,000 SNPs. Using the fineSTRUCTURE method we detected striking and rich patterns of population differentiation within Spain, at scales down to tens of kilometres. Strikingly, the major axis of genetic differentiation in Spain runs from west to east, while conversely there is remarkable genetic similarity in the north-south direction. To infer details of historical population movements into Spain, we analysed Spain alongside a sample of ~6,000 individuals from Europe, North Africa, and sub-Saharan Africa. Across Spanish groups, we identify varying genetic contributions from north-west African ancestral populations, at times that all fall within the period of Islamic occupation. We also identify Basque-like admixture within Spanish groups to the south of the Basque-speaking region, implying southerly gene flow from this region. This analysis has revealed details of the strengths and weaknesses of different approaches to investigating population genetic history, as well as providing important new insights into the complex genetic history of Spain.

1564T

Polygenic Adaptation Regression Analysis. *Y. Field[1,2], N. Telis[3], E. A. Boyle[1,4], D. Golan[1,5], J. K. Pritchard[1,2,6].* 1) Genetics, Stanford University; 2) Howard Hughes Medical Institute; 3) Biomedical Informatics, Stanford University; 4) Stanford School of Medicine; 5) Statistics, Stanford University; 6) Biology, Stanford University.

Understanding how natural selection had shaped the existing genetic variation within humans is a major goal in population genetics. With the growing understanding that many human diseases and complex traits have a polygenic genetic architecture, it has been hypothesized that adaptation in recent human history might be largely polygenic as well. The increased frequency of many alleles associated with genetic basis for tall stature in northern Europe, has been the major supporting example for the polygenic adaptation model. However, beyond this outstanding example, the nature and extent of polygenic adaptation in recent human history is still poorly understood. Current methods for testing for polygenic adaptation, based on allele frequency differences between populations, do not account for the linkage disequilibrium between loci. In turn, there is no general framework available for testing for adaptation over one set of functionally related loci, while controlling for possible causal effects (on allele frequency differences) by other genetically-linked genomic features. For example, one would like to test for adaptation among known GWAS hits, controlling for the selection for height; or to test for selection within regulatory regions, controlling for possible selection on non-synonymous sites; or to control for admixture effects on allele frequency differences, etc. To address this need, we have developed POLARIS, a novel and general method for POLygenic Adaptation Regression analysIS. Our method is based on a multivariate normal model for the frequency differences between populations, which is structured to explicitly represent linkage disequilibrium, drift and annotation-dependent polygenic adaptation. The method allows to test, and control, for annotation-dependent effects on both the mean and variance of allele frequency difference, giving it a great flexibility to mix directed and undirected hypotheses. As we demonstrate with an initial analysis of publically available datasets, POLARIS opens the road for a richer and more extensive characterization of the nature and extent of polygenic adaptation in recent human history.

1565W

Patterns of IBD (identity-by-descent) sharing among 780,000 present-day Americans reveal geography and recent settlement history in the United States. *E. Han[1], R. E. Curtis[1], M. J. Barber[1], J. K. Byrnes[1], P. Carbonetto[1], J. M. Granka[1], A. R. Kermany[1], N. M. Natalie[1], K. Noto[1], Y. Wang[1], E. R. Battat[2], C. A. Ball[1], K. G. Chahine[1].* 1) AncestryDNA, San Francisco, CA; 2) History and Literature, Harvard University, Cambridge, MA.

The United States is often thought of as a new homogeneous population formed by mixture of different nationalities, ethnicities, and cultures. However, recent research shows that there is a distinct geographic distribution for ethnicity and isolation-by-distance models hold in the United States. We take over 780,000 genotyped individuals from the United States and identify over 600,000,000 segments of identity-by-descent (IBD). We recursively run the *Louvain community detection* method on the resulting IBD network (vertices are individuals and weighted edges are IBD connections among them). Finally, we annotate the network vertices with information from the corresponding individuals' pedigree data, which include the surnames, birth dates and locations of ancestors dating back hundreds of years. The result is a striking set of genetic clusters that reveal migrations and settlements over time, show the movement of groups within the United States, and identify groups of people connected to each other through ethnicity or culture. We are able to identify Asian, Jewish, and Latin American clusters spread over the Americas. We detect French Americans in Maine and Louisiana, Scandinavian Americans in Midwest, Irish and Italian American in urban areas, German Americans in Pennsylvania and Ohio, and English Americans in Northeast. We detect a Mormon migration to Utah and an Amish group in Pennsylvania. We identify more geographically isolated populations in the Cumberland Mountains. In the south, we identify African Americans and agricultural populations. Often, the groups we identify are associated with known migrations, and their movements are associated with known history of economic changes in various parts of the United States. For example, a large wave of German immigrants are known to have settled in Pennsylvania before 1780-1810, and tended not to move once they settled there. We detect not one but three distinct subgroups that settled in slightly different parts of Pennsylvania. We can trace their ancestry from different parts of Germany and Switzerland at that time and confirm it through aggregating estimates of their ancestry based on their genomes and common surnames found in their pedigrees. In summary, our analysis provides a significant insight into population structure in the United States which still remains distinct enough to identify only using patterns of IBD sharing among present-day Americans.

1566T

Identity by descent analysis reveals fine-scale population structure in Crete. *A. M. Plantinga[1], F. Tsetsos[2], P. Paschou[2], P. Drineas[3], B. L. Browning[1,4], G. Stamatoyannopoulos[5].* 1) Department of Biostatistics, University of Washington, Seattle, WA; 2) Department of Molecular Biology and Genetics, Democritus University of Thrace, 68100 Alexandroupolis, Greece; 3) Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY; 4) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 5) Departments of Medicine and Genome Sciences, University of Washington, Seattle, WA.

The island of Crete is 160 miles long and 37 miles at its widest point. Despite its small area, Crete contains populations that historically have been relatively isolated due to Crete's mountainous topography. We hypothesized that this isolation would create detectable fine-scale population structure across Crete. To study this structure, SNP array data was collected for 129 subjects sampled from 17 regions of Crete. All individuals were at least 70 years old at the time of sampling and reported having maternal and paternal grandparents from the same region as the subject. We show that identity by descent (IBD) analysis, which detects chromosomal segments that are shared between individuals due to inheritance from recent common ancestors, reveals fine-scale regional relationships that are consistent with the geography and history of Crete. For instance, the amount of genetic material shared IBD between two populations decreases both with separation by mountains or other natural barriers and with increasing east-west distance. Expansion of the IBD analysis to include samples from other parts of Greece reveals close relationships between samples from West Crete and samples from the Peloponnese, which is northwest of Crete, and similarly close relationships between samples from East Crete and samples from the Dodecanese, which are northeast of Crete. These results indicate that IBD is a powerful way to detect genetic relationships between populations on a fine geographic scale.

1567W

The fine-scale genetic structure of the Japanese population: estimation of genetic materials derived from Asians and the basis of genetic differentiation. *F. Takeuchi[1], T. Katsuya[2], R. Kimura[3], T. Nabika[4], T. Ohkubo[5], Y. Tabara[6], K. Yamamoto[7], M. Yokota[8], WH. Saw[9], YY. Teo[9], N. Kato[1].* 1) Dept of Gene Diagnostics and Therapeutics, National Center for Global Health and Medicine, Tokyo, Japan; 2) Dept of Clinical Gene Therapy, Osaka University Graduate School of Medicine, Suita, Japan; 3) Department of Human Biology and Anatomy, Graduate School of Medicine, University of the Ryukyus, Nishihara-cho, Japan; 4) Department of Functional Pathology, Shimane University School of Medicine, Izumo, Japan; 5) Department of Hygiene and Public Health, Teikyo University School of Medicine, Tokyo, Japan; 6) Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; 7) Department of Medical Chemistry, Kurume University School of Medicine, Kurume, Japan; 8) Department of Genome Science, School of Dentistry, Aichi Gakuin University, Nagoya, Japan; 9) Saw Swee Hock School of Public Health, National University of Singapore, Singapore.

BackgroundJapan has often been viewed as a genetically homogenous country among Asians. Based on archaeological evidence, it is suggested that a modern Japanese population was formed by admixture of the Jomon and Yayoi peoples around 3000 years ago, yet genetic data supporting this is limited. In the recent methodological transition from GWAS to genome-wide sequencing, knowledge about the fine-scale genetic structure of Japanese is important as a signature of historical demographic events to avoid its potential for confounding disease association studies. **Methods**For 1,600 individuals comprising 200 each from eight geographical regions in Japan (seven regions in or close to the mainland of Japan plus Okinawa), we performed SNP genotyping with GWAS arrays and applied the fineSTRUCTURE algorithm that could capture fine-scale population structure based on haplotype coancestry to compute genetic clusters in the Japanese populations. Then, we estimated the ancestry of Japanese by referring to genome-wide SNP data for 25 Asian (non-Japanese) populations derived from the Asian Diversity Project. Moreover, we examined whether genetic differentiation between the regional clusters of Japanese could result from genetic drift or positive selection. **Results**We found that the genetic clusters computed solely based on SNP data could distinguish the eight geographical regions in Japan; and that the ancestry of Japanese was a mixture of Korean, Chinese and Southeast Asian populations, while the composition of three ancestries differed modestly between the regional clusters. Genetic diversity was most prominent between the Okinawa population and any of the mainland Japanese populations, in accordance with previous studies. Geographical genetic differences, although relatively weak in signal, were detectable across the chromosomes, and in a few instances they appeared to be under some positive selection. **Conclusion**Our data show that the fineSTRUCTURE algorithm allows for capturing fine-scale population structure among seven regions in or close to the mainland of Japan, which have not been distinguishable in previous studies using principal component analysis. The mainland Japanese populations are grouped into two super-clusters in addition to individual regional clusters. However, such geographical genetic differentiation is found to principally result from genetic drift and it may not appreciably confound disease association studies involving the mainland Japanese populations.

1568T

Recent polygenic adaptation in Europe. *N. Telis[1], E. A. Boyle[2,5], Y. Field[3,4], J. K. Pritchard[3,2,4].* 1) Biomedical Informatics, Stanford University; 2) Biology, Stanford University; 3) Genetics, Stanford University; 4) Howard Hughes Medical Institute; 5) Stanford School of Medicine.

Adaptive evolution in recent human history remains poorly characterized. Human population genetics has focused on strong selective sweeps. However, our understanding of other selective patterns and their effects on patterns of human genetic diversity is still limited. Although there is compelling evidence for recent selection pushing increased height in northern Europe, the literature is devoid of other strong notable examples of recent polygenic selective events. We develop a non-comparative scoring method for individual polymorphisms to detect signals of recent adaptation, based on using singleton density to approximate haplotype age. Simulations suggest that this method preferentially detects recent evolutionary events with several different sweep patterns. We apply this method to 1,600 individuals from the ALSPAC cohort and confirm known selection signals in Northern Europeans, as well as broader signals of polygenic selection. We investigate associations with these signals and demonstrate that these signals are robust to population allele frequency differences in Europeans. We use this method in combination with population allele frequency differences to identify novel signals of polygenic adaptation in modern Europeans.

1569W

Insights on modern human migration and rare variation spread using ten of millions of genealogical profiles. *M. Wahl[1], J. Kaplanis[2], A. Gordon[3], Y. Erlich[3, 4].* 1) Molecular and Cellular Biology, Harvard University, Cambridge, MA; 2) Whitehead Institute for Biomedical Research, Cambridge, MA; 3) New York Genome Center, New York, NY; 4) Department of Computer Science, Columbia University, New York, NY.

Investigation of recent human migration patterns via genealogical records provides a low-cost, medium-accuracy, and high-volume complement to genomic approaches. National and church population registers are renowned for their reliability regarding persons who live and die near their birthplace, but rarely track persons who undergo long-distance migrations. Familial records maintained by hobbyist genealogists are more likely to maintain this information, but may be partial or incorrect: these disadvantages are allayed by the rising popularity of tools to digitize, share, and compare family trees, increasing the number and accuracy of available data. Here we employ public records on over 43 million persons from the 1700s through the present day, obtained from the genealogical social network Geni. com, to study patterns of human migration from birth to childbirth and death. Fathers appear both more likely than mothers to raise children in their hometown and more likely to migrate unusually long distances before raising children. We find evidence for an east-west bias in migration both locally and at long distances. Dispersal of descendants of a common ancestor is examined as a function of average generations elapsed, ancestor's birth year, and proximity to the frontier of a human range expansion (viz. , colonization of the American West). We further simulate inheritance of hypothetical novel autosomal, Y chromosome, or mtDNA mutations from given ancestors to investigate typical distributions of carriers in the present day.

1570T

'Human Knockout Project' in a Pakistani population with high levels of consanguinity. *P. Natarajan[1,2], W. Zhao[3], A. Rasheed[4], S. Khetarpal[5], P. Frossard[4], E. S. Lander[2], S. Gabriel[2], M. J. Daly[2,6], J. Danesh[7], D. J. Rader[5,8], S. Kathiresan[1,2], D. Saleheen[3,4] on behalf of the PROMIS Investigators.* 1) Center for Human Genetic Research and Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA, USA; 2) Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Center for Clinical Epidemiology and Biostatistics and Department of Biostatistics and Epidemiology, University of Pennsylvania, USA; 4) Center for Non-Communicable Diseases, Karachi, Pakistan; 5) Division of Translational Medicine and Human Genetics, Department of Medicine, University of Pennsylvania, USA; 6) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 7) Department of Public Health and Primary Care, University of Cambridge, UK; 8) Department of Genetics, University of Pennsylvania, USA.

Experimental disruption of both gene copies ("knockout") in model organisms has proven useful to understand gene function. Here, we leverage naturally-occurring null mutations and consanguinity in the human population in order to: 1) identify humans knocked out for genes; and 2) characterize the phenotypic consequences of complete gene disruption. We performed whole-exome sequencing in 7,078 individuals living in Pakistan, a region of the world with high levels of consanguinity, and tested whether knockouts differed from wild-type participants across a range of over 200 cardiometabolic traits. Among Pakistanis, the median excess homozygosity, as determined by the F inbreeding coefficient, was 1. 6%. 47,656 mutations were predicted to be null (nonsense, frameshift, or canonical splice-site) and we further filtered variants based on minor allele frequency, transcript position, splice sites, and conservation to identify 36,850 "high-confidence" null mutations across 12,131 autosomal genes. Across all participants, 961 distinct genes were completely disrupted by homozygous high-confidence null mutations. 1,306 participants (18. 4%) had at least one gene knocked out. Simulations using the excess observed autozygosity suggest that with the sequencing of ~200,000 Pakistanis, we can expect about 8,000 genes to be completed disrupted; this number is six-times larger than if a similar number of participants from an outbred population were sequenced. In a phenotypic screen, homozygosity for null mutations at *PLAG27* was associated with absent enzymatic activity of lipoprotein-associated phospholipase A2 in the blood; at *CYP2F1*, with higher plasma interleukin-8 concentrations; and at either *A3GALT2* or *NRG4*, with markedly reduced plasma insulin C-peptide concentrations. These observations provide a roadmap to understand the consequences of complete disruption of a large fraction of genes in the human genome.

1571W

Homozygous loss-of-function variants in European cosmopolitan and isolate populations. *J. Wilson[1,2], V. Svinti[2], J. G. Prendergast[3], Y. -Y. Chau[2], A. Campbell[4], I. Patarcic[5], I. Barroso[6], P. K. Joshi[1], N. D. Hastie[2], A. Miljkovic[5], M. S. Taylor[2], G. eneration Scotland[4], U. K. 10K[7], S. Enroth[8], Y. Memari[6], A. Kolb-Kokocinski[6], A. F. Wright[2], U. Gyllensten[8], R. Durbin[6], I. Rudan[1], H. Campbell[1], O. Polasek[1,5], A. Johansson[8], S. Sauer[9], D. J. Porteous[4], C. Drake[2], V. Vitart[2], C. Hayward[2], C. A. Semple[2], V. B. Kaiser[2].* 1) Centre for Global Health Research, Usher Institute for Population Health Sciences & Informatics, University of Edinburgh, Edinburgh, UK; 2) MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK; 3) The Roslin Institute, The University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, UK; 4) Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK; 5) Medical School, University of Split, Soltanska 2, Split, 21000, Croatia; 6) Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 1SA, UK; 7) members of the UK10K consortium are listed in www. uk10k. org; 8) Immunology, Genetics & Pathology, Uppsala University, Husargatan 3, Box 815, Uppsala, SE-751 08, Sweden; 9) Max-Planck-Institute for Molecular Genetics, Otto-Warburg-Laboratory, Berlin, Germany.

Homozygous Loss of Function (HLOF) variants provide a valuable window on gene function in humans, as well as an inventory of the human genes that are not essential for survival and reproduction. All humans carry at least a few HLOF variants, but the exact number of inactivated genes that can be tolerated is currently unknown - as are the phenotypic effects of losing function for most human genes. Here, we make use of 1,432 whole exome sequences from five European populations to expand the catalogue of known human HLOF mutations; after stringent filtering of variants in our dataset, we identify a total of 173 HLOF mutations, 76 (44%) of which have not been observed previously. We find that population isolates are particularly well suited to surveys of novel HLOF genes because individuals in such populations carry extensive runs of homozygosity, which we show are enriched for novel, rare HLOF variants. Further, we make use of extensive phenotypic data to show that most HLOFs, ascertained in population-based samples, appear to have little detectable effect on the phenotype. On the contrary, we document several genes directly implicated in disease that seem to tolerate HLOF variants. Overall HLOF genes are enriched for olfactory receptor function and are expressed in testes more often than expected, consistent with reduced purifying selection and incipient pseudogenisation.

1572T

Harnessing the power of exome sequencing and isolated populations to identify risk factors for inflammatory bowel disease. *M. A. Rivas[1,2], C. Stevens[1], B. Glaser[3], I. Peter[4], G. Atzmon[5], A. P. Levine[6], E. Schiff[6], N. Pontikos[6], V. Plagnol[7], A. W. Segal[6], S. Targan[8], D. Turner[9], P. Saavalainen[10], M. Farkkila[11], K. Kontula[12], A. Palotie[1,13,14], A. Pulver[15], J. H. Cho[16], D. McGovern[8], R. J. Xavier[1,17], M. J. Daly[1,2], T2D-GENES and NIDDK IBD Genetics consortia.* 1) Medical and Population Genetics, Broad Institute, Cambridge; 2) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 3) 3Hadassah-Hebrew University Medical Center, Endocrinology and Metabolism Service Department of Internal Medicine, Jerusalem, Israel; 4) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA;; 5) Department of Genetics and Medicine, Albert Einstein College of Medicine, Bronx, NY, USA; 6) Division of Medicine, University College London, London, United Kingdom; 7) UCL Genetics Institute, University College London, London, UK; 8) F. Widjaja Foundation Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA; 9) Juliet Keidan Institute of Pediatric Gastroenterology and Nutrition, Shaare Zedek Medical Center, The Hebrew University of Jerusalem, Jerusalem, Israel; 10) Department of Medical Genetics, Biomedicum Helsinki, University of Helsinki, Helsinki, Finland; 11) Department of Medicine, Division of Gastroenterology, Helsinki University Hospital, Helsinki, Finland; 12) Department of Medicine, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland; 13) Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland; 14) Department of Neurology, Massachusetts General Hospital, Boston, MA, USA; 15) Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA; 16) Icahn School of Medicine at Mount Sinai, Dr Henry D. Janowitz Division of Gastroenterology, New York, New York, USA; 17) Gastrointestinal Unit and Center for the Study of Inflammatory Bowel Disease and Center for Computational and Integrative Biology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA.

As part of a broader collaborative network of exome sequencing studies, we developed a jointly called data set of more than 3500 Ashkenazi Jewish exomes and make publicly available a resource of site and allele frequencies, which should serve as a valuable reference for medical genetics in the Ashkenazi Jewish population. We developed a novel statistical framework for the analysis of isolated populations in order to identify and characterize alleles that are enriched through the bottleneck event. We apply this to this resource and previously described Finnish reference from the SISu Project and estimate that 13% and 19% of alleles present in Ashkenazi Jewish and Finnish samples are uniquely enriched in these populations by roughly 10x compared to the maximum frequency seen in all other population exome sequencing resources. Among these enriched alleles, in case-control analysis of Crohn's disease we find an unusual aggregation of novel distinct Ashkenazi Jewish associated alleles in *NOD2* (p. A612T, p=5. 5x10-9; p. A755V, p=0. 002; p. D357A, p=0. 003; p. P427L, p=0. 005) in addition to those previously reported as associated in *NOD2*. A major focus of the program is to identify protein truncating alleles that protect against the inflammatory bowel diseases. Protective loss-of-function variants hold particular promise for identifying potential targets of inhibition withhigher confidence in both safety and efficacy. We used the sequencing data to seek such variants in genes strongly implicated by GWAS. We discovered a novel protective association at a predicted protein-truncating mutation (p. Arg179Ter) in*RNF186*, a single-exon ring finger E3 ligase with strong colonic expression. This study presents an unprecedented opportunity to examine the role of rare variants and its contribution to inflammatory bowel disease predisposition. Initial findings from these studies will enable studies of the allelic associations across multiple phenotypes in large population cohorts, powerful in-depth characterization of the functional consequences, and deeper understanding of the allelic architecture in directly implicated genes from ongoing fine-mapping studies of common variant associations. .

1573W

Rare coding and regulatory noncoding variants in the Saguenay-Lac-Saint-Jean founder population. *A. Morin[1,2], T. Kwan[1,2], L. Létourneau[2], AM. Madore[3], K. Tandre[4], ML. Eloranta[4], M. Ban[5], M. Caron[1,2], G. Bourque[1,2], S. Sawcer[6], AC. Syvänen[6], L. RonnBlom[4], MG. Lathrop[1,2], S. Gravel[1,2], C. Laprise[3], T. Pastinen[1,2].* 1) Human Genetics, McGill University, Montréal, Quebec, Canada; 2) McGill University and Genome Québec Innovation Centre, Montréal, Quebec, Canada; 3) Département des sciences fondamentales, Université du Québec àChicoutimi Saguenay, Quebec, Canada; 4) Department of Medical Sciences, Section of Rheumatology, Uppsala University, Uppsala, Sweden; 5) Department of Clinical Neurosciences, University of Cambridge, Cambridge; 6) Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden.

In the past few years, research has started to focus on delineating the whole genetic architecture of complex traits including very rare and singleton variations. We have developed ImmunoSeq: a custom capture panel that targets the exome region as well as noncoding regulatory regions of immune cells in order to study coding and regulatory noncoding variants in the context of complex traits that have an important immune or inflammatory component. The noncoding regulatory regions were selected base on DNase I hypersensitive sites (DHSs) mapping from a set of 12 different cell types available from ENCODE and NIH Roadmap epigenomics Projects. Asthma is one of the traits being examined using the Saguenay-Lac-Saint-Jean (SLSJ) asthma familial collection. The SLSJ region is located in northeastern Quebec and is known for its unique demographic history and founder effect, characterized by several population bottlenecks followed by rapid expansion. Since founder populations have smaller spectrum of rare variations but are enriched in rare and unique Loss-of-Function (LoF) variants, we aim to show the advantages of using the SLSJ population in the study of rare coding and noncoding regulatory variants in a complex trait. Rare variant distribution and characteristics of the SLSJ population (n=298) were compared to three other European populations from Sweden (n= 369), United Kingdom (n=264), and France (n= 96). We observed a lower number of low frequency and rare variants (MAF<5%) in the SLSJ compared to the other populations, which reflects the multiple bottlenecks in its history. A high proportion of rare variants unique to the SLSJ were observed (40%) that is increased when looking at selectively constrained (GERP++>=1) variants (55%). We also observed a larger proportion of variants that are common in the SLSJ but rare in the other populations. This is already known on the basis of rare disorders. As an example, the the *LRPPRC* mutation causing the Leigh syndrome French Canadian type has a carrier rate of 1 in 21 among inhabitants of the region, whereas it is extremely rare in other French populations. In the future, we will interrogate differences between coding and noncoding variants in the different populations as well as determine how much power we can gain using the SLSJ founder population in the study of rare variants in a complex trait.

1574T

Whole genome view of the Finnish bottleneck effects using 2926 whole genome sequences from Finland and UK. *H. Chheda[1], P. Palta[1], M. Pirinen[1], S. McCarthy[2], V. Salomaa[3], R. Durbin[2], T. Aittokallio[1], A. Palotie[1,4,5], S. Ripatti[1,2,6].* 1) Institution for Molecular Medicine Finland (FIMM), Helsinki, Finland; 2) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, United Kingdom; 3) National Institute for Health and Welfare, Helsinki, Finland; 4) Program in Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of MIT and Harvard, Cambridge, MA, USA; 5) Department of Medical Genetics, University of Helsinki and University Central Hospital, Helsinki, Finland; 6) Public Health, University of Helsinki, Helsinki, Finland.

Lim et al (Plos Genetics 2014) showed recently that loss-of-function (LoF) and missense variants in 0. 5-5% frequency are enriched in Finnish population compared to Non-Finnish Europeans, providing an opportunity to study downstream effects of these variants in Finns. However this change in the frequencies may not be confined only to the coding region. To this extent we have studied the enrichment of variants in the Finnish population across the whole genome. To study the bottleneck effects across the whole genome, we analyzed single nucleotide variants (SNVs) from 1463 low coverage whole genome sequences both from Finland (~4. 6x) and UK (6x). These samples were processed together by the Haplotype Reference Consortium to harmonize the variant calls and minimize the batch effects. As observed previously, we see a 1. 34x enrichment of the LoF variants (p-valueLoF = 0. 056) in the 2-5% minor allele frequency (MAF) range and 1. 1x enrichment in the missense variants (p-valuemissense = 2. 95e-05). Further, we studied the enrichment of variants across the whole genome. We found significant enrichment in Finns in the MAF range from 0. 5-5%, with maximum enrichment in the MAF range of 2-5% (p-value = 6. 4e-323). We also see enrichment across different functional sub-categories in Finns with the highest enrichment observed for conserved regions (p-valueconserved_regions=9. 36e-24, p-valueTFBS=6. 02e-46, p-valuepromoter=1. 67e-11, p-valueenhancers=0. 001), although not as considerable as for the LoF variants. Furthermore, in the regulatory regions, rare and low frequency variants (MAF <= 2%) are enriched beyond expected bottleneck effects. When limiting the analysis to the 23,441 variants that were enriched at least 100x in Finns, genes in pathways related to neuron development, signal transductions and cation transport channels were observed to be significantly over represented after correcting for multiple testing. These results show that the enrichment of low frequency variants in founder populations is not limited to coding loss-of-function and missense variants, but are also observed in conserved regions and regulatory elements. This finding provides opportunities to study downstream health effects of these variants in founder populations with multiple bottleneck effects such as Finns outside of the coding regions.

1575W

A genetic population isolate in The Netherlands showing extensive haplotype sharing and long regions of homozygosity. *L. Olde Loohuis[1], M. Somers[2], M. Aukes[2], B. Pasaniuc[1,4], K. de Visser[5], R. Kahn[2], I. Sommer[2], R. Ophoff[1,2,3].* 1) Center for Neurobehavioral Genetics, University California Los Angeles, Los Angeles, CA; 2) Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands; 3) Department of Human Genetics, David Geffen School of Medicine at UCLA, University of California Los Angeles, Los Angeles, CA, USA; 4) Department of Pathology and Laboratory Medicine, University of California Los Angeles, Los Angeles, CA, USA; 5) Department of General Practice, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands.

Background Genetic isolated populations have features that may facilitate genetic analyses and can be leveraged to improve power of mapping genes to complex traits. Our aim was to test the extent a population with a former history of geographic isolation, and currently with one of the highest fertility rates in The Netherlands, shows signs of genetic isolation. **Methods** Genome-wide genotype data was collected of 72 self-reported unrelated individuals from this population as well as in sample of 104 random control subjects from The Netherlands. Additional reference data from different populations and population isolates was available through HapMap and the Human Genome Diversity Project (HGDP). We performed a number of analyses to compare the genetic structure between these populations. We calculated the pairwise genetic distance (Fst) between populations, examined the extent of identical-by-descent (IBD) sharing and estimated the effective population size. **Results** Genetic analysis of this population showed consistent patterns of a population isolate at all levels tested. We confirmed that this population is most closely related to the Dutch control subjects, and detected high levels of IBD sharing and runs of homozygosity at equal or even higher levels than observed in previously described population isolates. The effective population size of this population was estimated to be several orders of magnitude smaller than that of the Dutch control sample. **Conclusions** We conclude that the geographic isolation of this population combined with rapid population growth has resulted in a genetic isolate with great potential value for future genetic studies. .

1576T

Recessive disease gene mapping in India: extraordinary opportunities for understanding health and disease. *N. J. Nakatsuka[1], K. Thangaraj[2], P. Moorjani[1,3,4], A. Tandon[1,3], N. Patterson[3], L. Singh[2], D. Reich[1,3,5].* 1) Department of Genetics, Harvard Medical School, Boston, MA USA; 2) Centre for Cellular and Molecular Biology, Hyderabad, India; 3) Broad Institute of MIT and Harvard, Cambridge, MA USA; 4) Department of Biological Sciences, Columbia University, New York, NY USA; 5) Howard Hughes Medical Institute, Harvard Medical School, Boston, MA USA.

Modern India is a region of remarkable cultural, linguistic, and genetic diversity with over 4,500 anthropologically well-defined groups. Large genetic differentiation has been observed between many of these groups, reflecting strong founder events with effects that have been preserved in some cases for thousands of years due to low genetic exchange between groups. We undertook a systematic survey to assess the strength of founder events in over 1200 individuals from over 230 Indian groups genotyped on Affymetrix (6. 0 and Human Origins) and Illumina (650K) arrays. These groups include tribes, castes, and religious groups with a wide-range of census sizes and spanned every state in India. We also analyzed Ashkenazi Jews and Finns, two groups known to have high rates of recessive diseases due to strong founder events. To determine the severity of founder events, we measured the total length of the genome inherited identical-by-descent (IBD) in each group. The data were phased with Beagle 3. 3. 2, and detection of IBD fragments was performed using FastIBD and GERMLINE. The HaploScore algorithm was used to filter out false positive fragments. To reduce the influence of recent consanguinity, we excluded closely related individuals detected by the presence of very long IBD segments. We quantified the IBD score for a group as the combined length of IBD segments between 3 to 20cM long, averaged over all pairwise comparisons within the group. We find that over 100 Indian groups in our dataset have founder effects stronger than in Ashkenazi Jews and Finns, including many groups with large census sizes (>1 million). This represents an extraordinary opportunity for biological discovery and potential reduction of genetic disease burden through mapping of recessive disease genes and prenatal counseling. Future work should focus on better characterization of the history and relationships amongst the founder events, as well as mapping variants associated with genetic diseases in the groups with the strongest founder events.

1577W

Using Y chromosome diversity to investigate the origin and the formation of the Taiwanese. *J. Loo, T. Wang, Y. Lai, M. Lin.* Medical Research, MacKay Memorial Hospital, New Taipei City, Taiwan, Taiwan.

Taiwanese comprise Aborigines and Non-Aborigines. The Aborigines are further divided into many Mountain tribes and Plain tribes. The Non-Aborigines are generally categorized as “Taiwan Han”, and are represented by two major populations: Minnan and Hakka. Actually, these groups can be specifically divided further into separate groups each having well characterized but independent profiles. From the article of Trejaut et al. 2014, the multidimentional scaling plot of Y-SNP frequencies displayed the relative position of Taiwanese relationship between themselves and other Asian populations (Figure 5). Interestingly, the Taiwan populations were located on a straight line, with Taiwan Mountain Aborigines on the one end, the Taiwan Plain Tribes in the middle, and Minnan, Hakka and Taiwan Coastal people on the other end. In a first stage we used BEAST package to compare the relationship of all Taiwan aboriginal Y haplogroup O1a1* (M214, M175, M119, P203) haplotypes with non aboriginal (Han) Taiwanese within the same haplogroup. Regrettably only five Y-STR loci could be used due to the vast diversity of typing methods between different laboratories. Several points are noticed:1) There are clades which are typically composed of haplotypes belonging to Taiwan Aborigines (*P*=0. 006);2) These clades are not unique as they can also be found among other Island Southeast Asian populations;3) Clades that are poor in aboriginal haplotypes represent Taiwan Han (*P* =0. 0029), Taiwan Coastal people (*P* <0. 0001), Northeast Asian and Northern Han (*P* =0. 0011);4) The distribution of Taiwan Plain Tribes haplotypes shows no difference between aboriginal and non aboriginal clades (p=0. 3629) which Beast indicate it might be recent admixture;5) While “Han” are represented by 4 points of divergence within the O1a1* Tree, the Taiwan aboriginal groups is represented by 29 divergence pointsConclusion: Based on the Y O1a1* haplogroup, the landscape of Taiwan Aborigines and Taiwan Non-Aborigines are different. And the formation of Taiwan Non-Aborigines is related to more recent migration of Northeast Asia. The second stage of this study will be to analyze other haplogroups (O2 and O3).

1578T

Characterization of 20,000 clinically relevant variants in 50,000 non-European individuals. *E. E. Kenny[1], C. R. Gignoux[2], S. A. Bien[3], N. Zubair[3], A. E. Justice[4], L. Fernandez-Rhodes[4], M. Graff[4], G. L. Wojcik[2], Y. Vaydylevich[1], G. M. Belbin[1], A. A Seyerle[4], J. M. Kocarnik[3], R. D. Jackson[5], J. Xing[6], F. Schumarcher[7], E. Stahl[1], J. L. Ambite[7], S. Buyske[6], C. Wassel[8], I. Cheng[9], L. A. Hindorff[10], T. C. Matise[6], R. J. F. Loos[1], C. Kooperberg[3], K. E. North[4], C. A. Haiman[7], J. Romm[11], C. D. Bustamante[2], C. S. Carlson[3], N. S. Abul-Husn[1].* 1) Icahn School of Medicine at Mt. Sinai, New York, NY; 2) Stanford University, Stanford, CA; 3) Fred Hutchinson Cancer Research Center, Seattle, WA; 4) University of North Carolina, Chapel Hill, NC; 5) Ohio State University, Columbus, OH; 6) Rutgers University, Piscataway, NJ; 7) University of Southern California, Los Angeles, CA; 8) University of Pittsburgh, Pittsburgh, PA; 9) Cancer Prevention Institute of California, Fremont, CA; 10) National Human Genome Research Institute, National Institute of Health, Bethesda, MA; 11) Johns Hopkins University, Bethesda, MA.

During the past decade major advances in deciphering the genetic basis of human disease have resulted in >6000 disorders that are now understood at a genetic level. However, the implementation of these findings into clinical care poses a difficult challenge. Vast quantities of rare variants in human populations make questions of causality, penetrance and expressivity for any particular variant difficult to ascertain. This problem is compounded by the recent observations from population genetics that rare variants are typically restricted to one or a few closely related populations, with implications that causal variants discovered in one population are unlikely to translate to another. Therefore, the characterization of clinically relevant variants (CRVs) in multi-ethnic cohorts is a critical step in advancing global genomic medicine efforts.

To address this, we generated a panel of 2,720 clinically relevant genes, including monogenic and polygenic disease genes, curated from ClinVar, HGMD, and OMIM in consultation with domain experts, and pharmacogenes from PharmGKB. Within these genes, we selected 15,000 variants previously classified as pathogenic, and 5,000 loss-of-function variants we classified as likely pathogenic extracted from a multi-ethnic panel of approximately 40,000 whole exome sequenced individuals. These variants were included on the Illumina Infinium Multi-Ethnic Genotype Array (MEGA) which was used to genotype 50,000 individuals, including African American, Afro-Caribbean, North, Central and South American, Japanese, and Native Hawaiian Islanders, as part of the Population Architecture using Genomics and Epidemiology (PAGE) Study. These data are linked to broad epidemiological trait and disease information, and a subset of 12,500 are linked to longitudinal electronic medical records (EMR).

We will present data on the frequencies of known and likely pathogenic variants, and will report phenotypic evidence to support causality and penetrance for a subset of CRVs segregating in PAGE. For example, we have already uncovered a skeletal dysplasia variant in *COL27A1* that is observed only in Puerto Ricans, and is in homozygosity in five individuals with extreme short stature (mean 4’ 2” females and 4’ 10” males). EMR data supports features of skeletal dysplasia despite a lack of diagnosis in those individuals. Follow up work will focus on characterizing other rare disease variants and uncovering monogenic forms of complex disease in this cohort.

.

1579W

Understanding germline mutation from multi-sibling families. *R. Rahbari[1], A. Wuster[1], S. Lindsay[1], R. Hardwick[1], L. Alexandrov[1], A. Dominiczak[2], A. Morris[3], D. Porteous[4], B. Smith[3], M. Stratton[1], M. Hurles[1], UK10K Consortium.* 1) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK, CB10 1SA; 2) nstitute of Cardiovascular and Medical Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, United Kingdom; 3) Medical Research Institute, University of Dundee, Dundee, United Kingdom; 4) Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, Scotland, United Kingdom.

Germline mutation is a driving force behind genome evolution, genetic variation and genetic disease. Despite recent focus on the germline mutation rate, there are some fundamental aspects of germline mutation that remain elusive. We investigated mutation rates, spectra, and timing by sequencing the genomes of multi-sibling families. We found that while the number of mutations in the child increases with paternal age in each family, the rate of this increase differs more than two-fold between families. We meta-analyzed the mutation spectra of 6,570 high confidence *de novo* mutations and found remarkable consistency of germline mutation spectra between the sexes, between families and at different paternal ages. Moreover, we noted that 3. 8% of de novo mutations per genome are germline mosaic in the parents, resulting in 1. 3% of mutations being shared between two or more siblings. The rate of sibling recurrence of *de novo* mutations varied significantly between families. Our data suggest that the mutation rate per cell division is higher during early embryogenesis and during proliferation of primordial germ cells, but is reduced three-fold during post-pubertal spermatogenesis. These findings have important consequences for the prediction of recurrence risks of genetic disorders caused by *de novo* mutations.

1580T

Most Rare and Common Variants in the *MAP3K1* Gene Have Neutral Phenotypic Effects. *M. Groden, J. Loke, K. Upadhyay, A. Pearlman, H. Ostrer.* Albert Einstein College of Medicine, Bronx, NY.

Gain-of-function (GOF) missense and splicing mutations in the *MAP3K1* gene cause 46,XY partial and complete gonadal dysgenesis by tilting the balance from the male to the female sex-determining pathways (Loke et. al. , Hum Mol Genet. 23:1073, 2014). Loss-of-function (LOF) mutations in *MAP3K1* while observed commonly in breast cancer (and other cancers) are rarely present in the germline state (only 1 stopgain, 1 frameshift and 1 inframe deletion and no splice or indels inframe insertions observed in the 1000 Genomes Project (1kG). Here, we used flow variant analysis (FVAs) to assess the phenotypic consequences of this stopgain and 14 rare and common missense variants of 24 observed in 1kG. Lymphoblastoid cells (LCLs) for these individuals (obtained from the Coriell NHGRI Catalog) were compared to LCLs with 5 known GOF mutations for MAP3K1 protein binding to RHOA, MAP3K4 and FRAT1 and phosphorylation of downstream targets, p38 and ERK1/2. The accuracy of the functional predictive algorithms (SIFT, Polyphen-2) was also compared to the observed molecular phenotypes. FVAs for the GOF variants showed increased phosphorylation of p38 and ERK1/2 and binding of cofactors (FRAT1, MAP3K4 and RHOA), and decreased binding of RAC1. The stopgain variant increased phosphorylation of ERK1/2, but not p38, and increased binding of RAC1 and RHOA. None of those effects was observed for the common or rare population ascertained variants. These phenotypic effects were not predicted by SIFT and Polyphen-2. We have extended the use of FVAs to demonstrate the phenotypic effects of LOF mutations in *MAP3K1*. Despite the common inference that most rare variants are deleterious, FVAs of rare missense 1kG *MAP3K1* variants demonstrated neither loss nor gain of functional effects. Computational classifiers over predicted the prevalence of deleterious variants in 20% of our sample demonstrating the utility of FVAs over these tools. Although physical phenotypes have not been associated with *MAP3K1* LOF variants, they may be subject to diminished reproductive fitness (negative selection) as occurs for GOF variants in this gene.

1581W

Rare variants – the key to decipher cryptic relatedness in diverse population structures. *A. C. Shetty[1], T. D. O'Connor[1,2] On behalf of the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA).* 1) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD; 2) Program in Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD.

Background: Common variants have primarily been used to infer the structure of the human population. Using common variants, we have only been able to successfully discern the broad-scale patterns of population structure. Recent dramatic events of population expansion have resulted in humans harboring a vast number of rare variants that play an influential role in delineating fine-scale patterns of population structure, which when left unrecognized could lead to spurious results from association studies. **Hypothesis:** Rare variants are concentrated within regions of low nucleotide variability supporting the premise that they tag recent haplotypes in the population. **Methods:** Our study analyzed the genetic variation data from the latest phase III release of the 1000 Genomes Project (1KG) and 952 high coverage (~30x depth) whole genomes sequenced as part of the **C**onsortium on **A**sthma among **A**frican-ancestry **P**opulations in the **A**mericas (CAAPA) project. The genetic variants were categorized in different bins based on their derived allele frequency (DAF). We computed the average nucleotide diversity (ϖ) for each 10kb window in a 1Mb region flanking each variant. Separate ϖ values were computed for individuals possessing the variant (carriers) and individuals without the variant (non-carriers). The ratio of ϖ in carriers to non-carriers was computed for each 10kb window. **Results:** For rare variants (DAF $\leq$ 0. 05), ϖ values were lower in carriers than non-carriers. However, this lower ratio was only observed in the immediate flanking regions (~200kb to 350kb) from the variant. The ratio of ϖ in carriers to non-carriers gradually increased till no difference in ϖ values were seen between carriers and non-carriers as we moved further upstream or downstream from the variant. In addition, as the DAF increased from rare to more common variants, the difference in ϖ between carriers and non-carriers diminished. These results were consistent within the populations from 1KG as well as within the African-admixed CAAPA dataset. **Conclusions:** We were able to demonstrate that rare variants resided within genomic loci that posses lower nucleotide diversity in carriers than non-carriers. This correlation of derived allele frequency of genetic variation to genomic variability could help identify recently diverged haplotypes giving us insights into cryptic kinship in population samples and reveal signatures of fine-scale population structure in association analyses.

1582T

Path Analysis and Structural Equation Modeling for traditional cardiovascular risk traits in Punjabi adolescents. *S. K. Brar[1], B. D. Badaruddoza[2].* 1) Deptt of Human Genetics, Guru Nanak Dev University, Amritsar, India; 2) Director, Aligarh Muslim University Murshidabad Centre, Murshidabad, WB.

The accuracy of different anthropometric and socio-economic lifestyle indices to predict cardiovascular diseases is still debated in adolescent age group and few data are available in state of Punjab in India. Keeping in mind, that the complex nature of all studied anthropometric and socio-economic lifestyle variables to evaluate the determinants of pre-hypertension and hypertension, the path analysis of structural equation modeling using Analysis of Moments Structures (AMOS 21. 0) model fitting program has been applied in the study consisting a total of 1530 adolescent boys and 1530 adolescent girls aged 10 to 18 years old. The study first tested the adequacy of hypothesized path model which contain the inter-correlated variables (such as BMI, waist circumference, WHR, sum of skinfolds, food habits, exercise, screen time and family status). Next, the study tested the relative strength of association of independent variables with SBP and DBP on the basis of the magnitude of path coefficients of regression, covariances and variance. The study adopted maximum likelihood estimation, 95% confidence limit and critical ratio in generating the estimate of path coefficients and its significance. The whole analysis confirmed that the waist circumference, BMI and family income status have significant contribution to increase blood pressure. The findings of the present study regarding effect of BMI, waist circumference, WHR, sum of skinfolds, food habits, exercise, screen time and family status on blood pressure are in an agreement with few other studies It is further established from estimates of covariances from path analysis that combination of BMI with waist circumference, waist circumference with sum of skinfolds, BMI with sum of skinfolds, exercise with family status and screen time with family status have significant association with elevated SBP and DBP. Whereas, impact of variances on SBP and DBP of all the studied variables have been found significant for path analysis model among both adolescent boys and girls.

1583W

Privacy leaks from genomic data-sharing beacons. *S. Shringarpure, C. Bustamante.* Department of Genetics, Stanford University, Stanford, CA.

A major goal of the human genomics community is to enable efficient sharing and analysis of genomic and phenotypic data to understand the genetic contributors of health and disease. The Beacon Project aims to simplify data sharing through a web service (“beacon”) that provides only allele presence information. Users can query institutional beacons for information about genomic data available at the institution. Queries are of the form ``Do you have a genome that has a specific nucleotide (e. g. ,`A’) at a specific genomic position (e. g. , position 1127 on chromosome 1)” and the server can answer “Yes” or “No”. We hypothesize that beacons are susceptible to re-identification attacks, i. e. , given a query individual’s genome, we can predict whether this individual was among the set of genomes used to create the beacon. This is troubling since beacons often summarize data characterized by having a particular disease. For instance, identifying that a genome is present in the beacon at a cancer research institute suggests that the subject has cancer. Thus, beacons can leak not only membership information but also phenotype information. We have developed a statistical test to predict whether a given individual genome is present in the beacon using only allele presence queries. For a simulated beacon with 1000 individuals, 5000 SNP queries are enough to detect membership in the beacon with 95% power and 5% false positive rate. For empirical evaluation, we set up a beacon using 65 CEU individuals from Phase 1 of the 1000 Genomes project. With just 250 SNPs, beacon membership can be detected with 95% power and 5% false positive rate. In existing beacons, we were also able to identify an individual from the Personal Genome Project using just 1000 SNPs. We used our framework to show that relatives can be detected in a beacon (in the simulated beacon, 40,000 SNPs allow detection of parents/siblings of the query individual). Our results are robust to sequencing errors and differences in variant calling. Our test is independent of allele frequencies. Including multiple populations in the beacon does not reduce the power of the test. Our results show that beacons can leak unintended phenotypic information about subjects if care is not taken to ensure privacy. We suggest ways of reducing re-identification risks, such as increasing beacon size or sharing only common variation. Our solutions will allow data sharing while simultaneously decreasing the risk of re-identification.

1584T

Using DNA methylation data to test heritability-based predictions of evolutionary models of aging. *C. Robins[1,2], A. McRae[3,4], J. Powell[3,5], P. Visscher[3,4], G. Montgomery[6], D. Cutler[1], K. Conneely[1].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA; 2) Graduate Program in Population Biology, Ecology, and Evolution, Laney Graduate School, Emory University, Atlanta, GA, USA; 3) The Queensland Brain Institute, University of Queensland, Brisbane, QLD, Australia; 4) University of Queensland Diamantina Institute, University of Queensland, Translational Research Institute (TRI), Brisbane, QLD, Australia; 5) The Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD, Australia; 6) QIMR Berghofer Medical Research Institute, Brisbane, Australia.

Evolutionary theories have proposed both adaptive and non-adaptive explanations for the existence of aging. Major hypotheses include the theories of: 1) mutation accumulation (MA), and 2) antagonistic pleiotropy (AP), with disposable soma (DS) as a special case of AP. MA suggests aging to be caused by deleterious late-acting mutations that have accumulated in the population over many generations as a result of a declining strength of selection. In contrast, DS suggests aging to be caused by random somatic damages that accumulate over the lifetime of an individual due to imperfect somatic maintenance and repair mechanisms resulting from the evolutionary optimization of energy allocation tradeoffs. Past empirical tests have supported aspects of each theory, but no consensus has been reached on their relative importance in explaining human aging. Here, we use DNA methylation as a biomarker of aging to test contrasting predictions of MA and DS at sites across the genome. Many widespread and robust DNA methylation changes occur with age. That is, numerous sites across the genome show consistently variable patterns of methylation between younger and older ages. Sites exhibiting this pattern of changing methylation with age are called age-differentially methylated (aDM). Many of these sites also show heritable patterns of methylation, where closely related individuals are more correlated in their methylation levels than unrelated individuals. These facts allowed us to devise a test that differentiates MA from DS. MA predicts that aDM sites will show increasing heritability of DNA methylation levels with age, while DS predicts the opposite. To test these contrasting predictions in genome-wide DNA methylation data we used variance components models in SOLAR to test if the heritability of methylation level at aDM CpG sites increased or decreased with age. DNA methylation was measured at >480K CpG sites in blood sampled from 610 individuals from 176 families from the Brisbane Systems Genetics Study (age range=10-75 years). After exclusion of probes containing SNPs, 51% of CpG sites were aDM (FDR<. 05) and 42% had heritable methylation levels. Of the ~85K aDM and heritable sites, <5 showed significant increasing heritability, while >2500 showed significant decreasing heritability with age. These results suggest the heritability of age-associated DNA methylation changes may be overwhelmingly consistent with DS.

1585W

The NHGRI Sample Repository: A Collection of DNA and Cell Lines for Human Genetic Research. *A. M. Resch, F. Rahimov, M. Berkowitz, L. Brown.* Coriell Institute for Medical Research, Camden, NJ.

The knowledge of the entire spectrum of genetic variation across ethnically diverse human populations is important for reconstructing human evolutionary history and understanding the genetic basis of human disease. Established in 2006, the mission of the NHGRI Sample Repository for Human Genetic Research is to provide high-quality human biospecimens for researchers and clinicians investigating human population genetics and disease. The NHGRI Repository offers DNA samples and cell lines from twenty-seven distinct populations that include samples used for the International HapMap Project, 1000 Genomes Project, Human Microbiome Project (HMP), and the Fosmid Library Collection for Analysis of Structural Variation. Currently the NHGRI Repository hosts over 3,700 unique cell cultures and DNA samples for the HapMap and 1000 Genomes collections, 298 unique DNA samples for the HMP collection, and 11 fosmid libraries. Whole blood samples were collected from anonymized and de-identified donors with broad consent for various studies, including DNA sequencing and genotyping, gene expression and proteomics. Epstein-Barr virus-transformed lymphoblastoid cell lines (LCLs) were derived from peripheral blood mononuclear cells (PBMCs) and genomic DNA was isolated from LCLs. Cell lines and DNA may be used by academic and commercial entities. In addition, the NHGRI Repository utilizes secure laboratory information management systems to monitor inventory and distribution of samples. Thus, by establishing and maintaining a centralized collection of human biospecimens, the NHGRI Repository provides a unique resource for human genetic studies. The full catalog can be accessed at https://catalog. coriell. org/1/NHGRI.

1586T

The effect of SNP array ascertainment bias on the distribution of runs of homozygosity lengths. *A. Gladstein[1], M. Hammer[2].* 1) Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ; 2) Arizona Research Laboratories, Division of Biotechnology, University of Arizona, Tucson, AZ.

Background: Runs of homozygosity (ROH) are chromosomal segments of continuous homozygous sites that reflect the relatedness of an individual's ancestors. ROH are most often found using SNP genotyping arrays. ROH must be inferred indirectly since SNP arrays do not contain complete genomic information. Importantly, ROH can be incorrectly identified because of the underlying SNP array ascertainment bias. Ascertainment bias is the systematic deviation of population genetic statistics from theoretical expectations caused by sampling a nonrandom set of individuals or by biased SNP discovery protocols. Although SNP array ROH have been used in many demographic and disease studies, no work has assessed the impact of ascertainment bias on ROH identification.

Purpose: We develop a correction for the effect of ascertainment bias on the distribution of ROH lengths identified from SNP array data. We hypothesize that if a SNP array introduces ascertainment bias, on average, ROH identified from the SNP array will be longer than the true genomic ROH. Moreover, the difference between SNP array and genome ROH lengths will be a function of the genetic distance from populations used to discover the SNPs.

Methods: We simulate genomes using a two population split model and create pseudo SNP arrays from the simulated genomes. We find the ROH truth set from the simulated genomes and identify ROH from the pseudo array using PLINK. We also determine the most accurate PLINK parameters by comparing the pseudo array PLINK ROH to the truth set. We compare the distribution of the simulated genome and pseudo SNP array ROH lengths under increasing population divergence times. We determine a transformation, as a function of divergence time, that makes the distribution of pseudo SNP array ROH lengths closer to the true genome distribution.

Results: Under a null model of no ascertainment bias we found many PLINK parameter sets that produced similar distributions of ROH lengths to the simulated genome distribution. If we defined ROH as having a minimum length of less than 500kb, the default PLINK parameters were among the worst. Compared to the truth set, fewer and longer ROH were found from the pseudo array. We provide recommendations for the best PLINK parameters to accurately identify ROH from SNP array data and corrections to the distribution of ROH lengths based on the genetic distance between the study population and those used to discover the SNPs.

1587W

Anthropometric and cardiovascular trait variation among sub-Saharan African populations: the impact of subsistence practice and genetic ancestry. *M. E. B. Hansen[1], J. Lachance[2], S. Soi[1], L. Scheinfeldt[1], A. Ranciaro[1], S. Thompson[1], J. Hibro[1], S. A. Tishkoff[1].* 1) Genetics Dept, University of Pennsylvania, Philadelphia, PA; 2) School of Biology, Georgia Institute of Technology, Atlanta, GA.

The African continent is home to a diverse range of indigenous peoples that have adapted to a wide range of ecological environments and subsistence lifestyles. Many complex traits are expected to display variation between populations due to demographic history and/or adaptation to these diverse environments. In an effort to survey phenotypic variation in Africa and begin to understand the genetic and environmental factors that contribute to this variation, we have collected trait measurements on height (N~5000), BMI (N~5000), grip strength (N~2000), systolic and diastolic blood pressure (N~2000), and pulse (N~2000) from agricultural, pastoral, and hunter-gatherer communities across eastern and western sub-Saharan Africa. We present the observed variation in these traits between genders, across populations, and across subsistence practices. We find significant differences in trait values among these categories. A subset of 697 individuals were genotyped on the Illumina 1M-Duo SNP array. We performed a GWAS using a linear mixed model approach, and find that only height replicated GWAS top hits from non-African cohorts (p-value enrichment). To assess the impact of genetic ancestry and subsistence lifestyle on trait variation, we performed STRUCTURE analysis to determine ancestral cluster proportions, and fit both the mean effects and variance components for ancestry and subsistence category using linear mixed models. The fraction of variance explained by genetic ancestry is discussed, as well as the implications for future genotype/phenotype analyses within sub-Saharan Africa for these and related traits.

1588W

Splice site mutation in the key cytoskeletal gene *TLN1* causes systemic capillary leak syndrome in a first reported familial case. *D. Oz-Levi[1], V. Papadaki[2], R. Sion-Sarid[3], P. Nikolopoulou[2], T. Olender[1], S. Houri[3], V. Kostourou[2], D. Lancet[1].* 1) molecular genetics, weizmann institute of science, Rehovot, Rehovot, Israel; 2) Vascular Adhesion Lab, BSRC Alexander Fleming, 34 Fleming Str, Vari, 166 72 Athens, Greece; 3) Pediatric Intensive Care Unit, Wolfson Medical Center, Holon, Israel.

Systemic Capillary Leak syndrome (SCLS) is a rare disease characterized by life-threatening attacks of marked increase in capillary permeability, leading to plasma extravasation, vascular collapse and hypotensive shock, with roughly 150 isolated cases reported worldwide. We identified a first familial SCLS case, an Israeli family with ten affected individuals, showing suspected autosomal dominant inheritance with incomplete penetrance (Sion-Sarid R. et al 2010). The proband is a 14-year old boy with multiple attacks since the age of 5 months. His sister died of an acute attack at infancy. Exome sequencing on the proband and on an affected father's cousin yielded 37 candidate variants. We used our newly established variant interpretation tool VarElect (http://varelect. genecards. org/), to attain phenotype-dependent variant prioritization, leveraging the comprehensive information within the GeneCards Suite (http://www. lifemapsc. com/). Employing "capillary leak" as keyword, VarElect produced a prioritized list of contextually annotated genes, scored according to their likelihood to be disease related. The top candidate gene was *TLN1*, encoding Talin1, a cytoskeletal protein playing a key role in focal adhesion mechanosensory function, mediating the coupling of integrin b to actin filaments. The variant was A>G heterozygous substitution at a splice donor site (+2) of intron 54 (c. 7543+2A>G). cDNAs harboring the *TLN1* mutant allele show both skipping of exon 54, leading to an in-frame deletion of 21 residues, and retention of intron 54, that creates a premature stop codon. Talin1 coordinates endothelial dynamics, mediating microvascular remodeling, capillary contractility, and controlling of endothelial cell growth via VEGF and TGF-beta. These factors were found significantly elevated in SCLS serum of patients during an acute crisis. Such augmentation may provoke vascular leak symptoms by eliciting endothelial hyperpermeability. Transcriptome and proteome analyses on patient skin fibroblasts identified down-regulation of several immune genes, echoing previous reports on a connection of capillary leak to immunity. We hypothesize that the mutated protein could lead, perhaps via a dominant-negative model, to aberrant initiation of adhesion site assembly and endothelial dysfunction. Together with the observed increase in VEGF and TGF-beta, this suggests a model for SCLS pathogenesis, underlining a plausible molecular mechanism for this rare and precarious disorder.

1589T

Splice-EMASE: Quantifying allele-specific alternative splicing in diploid genomes. *N. Raghupathy, K. Choi, S. C. Munger, G. A. Churchill.* The Jackson Laboratory, Bar Harbor, ME.

Current RNA-seq analyses employ three separate pipelines to quantify gene expression abundance, Allele-specific gene expression (ASE), and alternative splicing from RNA-seq data. Gene-level abundance is estimated from alignment of all reads genome-wide, ASE is assessed by analyzing only reads that overlap known SNP locations; and alternative splicing is estimated by analyzing reads overlapping splice junction or exon-specific reads. Earlier, we developed computational tools; Seqnature and EMASE, to build individualized diploid genome using known phased genetic variations, perform personalized RNA-seq analysis using diploid transcriptome, and estimate allele-specific expression and gene expression simultaneously. Here we present Splice-EMASE, an extension of EMASE with splice-aware EM algorithm on diploid exons and junction alignment profiles. Splice-EMASE quantifies allele-specific alternative splicing in addition to ASE and total gene expression simultaneously. The EM algorithm probabilistically allocates both allelic and genic multi-mapping reads to estimate effective read counts at exons and splice-junctions. These effective exon and splice-junction read counts could be readily used to get allele/isoform/gene expression estimates. We demonstrate the utility of the approach by using simulated data and real RNA-seq data.

1590F

A system for gene ranking through variant annotation in Autism Spectrum Disorders. *E. Larsen[1], I. Menashe[2], M. Ziats[3], W. Pereanu[1], S. Banerjee-Basu[1].* 1) MindSpec Inc. , McLean, VA; 2) Department of Public Health, Faculty of Health Sciences, Ben Gurion University of the Negev, Israel; 3) Baylor College of Medicine, Houston, TX.

The search for genetic factors underlying autism spectrum disorders (ASD) has led to the identification of hundreds of genes containing thousands of variants that differ in mode of inheritance, frequency and function. These data are summarized in our Autism Database (AutDB; http://autism. mindspec. org/autdb/Welcome. do), a publicly available, curated, web-based, and searchable database for genetic variations associated with ASD. With the rapid pace of variant discovery in recent years, a major challenge in the field of ASD biology involves assessing the collective genetic evidence in an unbiased, systematic manner. Therefore, we set out to develop a scoring algorithm for prioritization of candidate genes based on the cumulative strength of evidence from each ASD-associated variant in AutDB. A total of 928 research articles from AutDB were analyzed to generate a dataset of 2187 rare variants and 711 common variants distributed across 461 ASD-associated genes. Each individual variant was manually annotated with 17 descriptors extracted from the original report, followed by score assignment using a set of standardized scoring parameters that are summed up to yield a single score for each gene in the database. There were remarkable variations in gene scores resulting in a log-normal distribution of scores with a mean gene score of 16. 65 ± 29. 57. Interestingly, there were 12 genes with scores deviating more than two standard deviations (SDs) from the mean score of all genes, with three high confidence genes (*SHANK3*, *CHD8*, and *ADNP*) with extremely high scores. Importantly, the gene scores generated by our approach were significantly correlated with the scores in the SFARI Gene scoring module (Spearman r = -0. 63; *P<0. 0001*) indicating a strong agreement between gene prioritization using our approach and the expert-mediated SFARI Gene scoring initiative. In this study, we provide a framework for assessment of diverse types of genetic variants associated with ASD. Using this model, we prioritize a set of genes with cumulative evidence that are likely to be important for defining the genetic risk architecture in ASD.

1591W

Typing of PRDM9 genotypic class from Next Generation Sequencing data in early onset childhood cancer. *A. Ang Houle[2], J. C. Grenier[2], V. Bruat[2], P. Awadalla[1,2,3].* 1) OICR, Toronto, Canada; 2) UHC Sainte-Justine Research Center, Montreal, Canada; 3) CARTaGENE, Montreal, Canada.

Recombination is the exchange of genetic information between homologous chromosomes during meiosis. The positions of recombination events cluster tightly together in recombination hotspots, which are determined in part by the DNA binding zinc finger domain of the rapidly evolving protein PRDM9. The repetitive microsatellite-like structure of the coding sequence of the ZnF array drives a constant generation of multiple new alleles in populations. Alleles of PRDM9 containing the *k*-ZnF have previously been associated with patients affected with childhood acute lymphoblastic leukaemia. PRDM9 alleles are notoriously difficult to type due to the repetitive nature of the ZnF arrays. Here, we propose a method to characterize alleles of PRDM9 from next-generation sequencing samples, by identifying the number of alleles containing a specific ZnF type. Our method is based on the correlation of profiles representing the counts of nucleotide sequences unique to each ZnF from ideal sets of short reads representing an allele pair and from the sequencing short reads of the sample. A scoring weight function is then applied to the top scoring correlations: the highest scoring genotypic class is deemed as predictive for the sample. We conduct a simulation analysis to examine the validity of the predictions obtained by our method with all pairs of known alleles, to compensate for the rarity of genuine samples containing some of the alleles, at varying coverage of next-generation sequencing data. We also confirm that the method can accurately genotype previously unobserved PRDM9 alleles by masking the ideal profiles containing corresponding alleles in the a priori given to the software. We also conducted a preliminary analysis to identify the PRDM9 *k*-ZnF genotype in cohorts of childhood cancers, including paediatric glioblastoma multiforme, which is characterized by the recurrent mutations in the histone H3, the target of PRDM9 epigenetic mark. This method opens the possibility of identifying associations between characteristics of PRDM9 alleles in other types of early onset childhood cancers, through a data-mining effort in public cancer databases.

1592T

Genomic Crowdsourcing: Allele Frequency Community Provides Expansive, Ethnically Diverse, Freely Available Community Resource for Allele Frequency Annotation. *D. Bassett[1], K. Boycott[2], C. Bustamante[3], D. Cooper[4], G. Eley[5], L. Furmanski[1], G. Glusman[6], D. Goldstein[7], M. Hegde[8], P. Hieter[9], A. Joecker[1], T. Kaminski[10], K. Kernohan[2], A. Krämer[1], S. Letovsky[10], S. Levy[11], T. Love[1], C. E. Mason[12], N. Pearson[13], H. Rehm[14], D. Richards[1], H. Rienhoff[15], E. Schadt[16], S. Shah[1], J. Shendure[17], B. Solomon[5], P. van der Spek[18], J. G. Vockley[5], R. Yip[1], X. Zhu[1], hundreds of Allele Frequency Community members worldwide.* 1) QIAGEN Bioinformatics, 1700 Seaport Blvd, 3rd Floor, Redwood city, CA USA; 2) Children's Hospital of Eastern Ontario (CHEO), 401 Smyth Road, Ottawa, Ontario K1H 8L1, Canada; 3) Stanford School of Medicine, 291 Campus Drive, Stanford, CA 94305-5101 USA; 4) Cardiff University School of Medicine, Cardiff CF14 4XN, UK; 5) Inova Translational Medicine Institute, 8110 Gatehouse Road, Falls Church, VA, 22042, USA; 6) Institute for Systems Biology, 401 Terry Avenue North, Seattle, WA, 98109, USA; 7) Columbia University Institute for Genomic Medicine, 630 West 168th St New York, NY 10032, USA; 8) Emory Genetics Laboratory, 2165 N Decatur Rd, Decatur, GA , 30033, USA; 9) University of British Columbia, 2329 W Mall, Vancouver, BC, V6T 1Z4, Canada; 10) Enlighten Health Genomics, a business of Laboratory Corporation of America® Holdings (LabCorp®), T W Alexander Drive, Research Triangle Park, North Carolina 27709, USA; 11) HudsonAlpha, 601 Genome Way, Huntsville, AL 35806, USA; 12) Weill Cornell Medical College, 1305 York Ave, New York, NY 10021, USA; 13) New York Genome Center, 101 Avenue of the Americas, New York, NY 10013, USA; 14) Harvard Medical School and Brigham & Women's Hospital, 65 Landsdowne Street, Cambridge, MA , 02139, USA; 15) Imago Biosciences, San Francisco, California, USA; 16) Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY, 10029, USA; 17) University of Washington, 3720 15th Ave NE, Seattle, WA, 98195, USA; 18) Erasmus University Medical Center, 3000 CA, Rotterdam, The Netherlands.

A key challenge in genome interpretation and precision medicine is the lack of an extensive, high quality, ethnically-diverse collection of human genomes as a reference set. A prospective disease-causing variant that appears to be "rare" based on publicly available sequence may in fact be a polymorphism in an ethnic population under-represented in public databases. Resources such as the Exome Variant Server, the 1000 Genomes Project, and the Exome Aggregation Consortium have been immensely valuable to the community, and Kaviar combines such datasets into integrated allele frequencies, but public databases have not been funded to provide broad and deep ethnic representation. QIAGEN's Ingenuity Variant Analysis™ genome interpretation solution has been used to interpret hundreds of thousands of ethnically diverse human sequencing samples. However, these NGS datasets are private and most are never publicly released. The Allele Frequency Community (www. allelefrequencycommunity. org) has been formed to address this interpretation need. Community members have pooled extensive human exome- and genome-wide variant call datasets in a secure, anonymized, pooled fashion to create the largest integrated, freely-accessible, hosted community database of allele frequencies ever available. More than 100,000 human exome- and genome-wide variant call datasets, including over 13,500 whole genomes, are already included in the Allele Frequency Community. The database is richly ethnically diverse, representing over 100 countries of origin and has been shown in benchmarking studies to significantly decrease the false positive rate in disease causing variants identification. To enable this community resource to grow over time, users have the opportunity to opt-in to join the Allele Frequency Community and benefit from the extensive database, agreeing in return to allow their sequences to contribute to the database. Only anonymous, pooled allele frequencies are provided, protecting patient privacy. QIAGEN Bioinformatics agreed to host the content and make it available free of charge via its HIPAA and Safe Harbor compliant genome interpretation ecosystem, which includes QIAGEN's Ingenuity Variant Analysis, CLC Biomedical Research Workbench and Clinical Insight offerings. Application of this new resource to clinical sequencing cases will be presented.

1593F

Robust Classification of Protein Variation Using StructuralModeling and Large-Scale Data Integration. *E. H. Baugh[1,4], R. Simmons-Edler[1,4], C. L. Mueller[2,4,5], R. F. Alford[6], N. Volfovsky[3], A. E. Lash[3], R. Bonneau[1,2,3,4,5].* 1) Department of Biology, New York University, Manhattan (New York City), NY; 2) Computer Sciences Department, New York University, Manhattan (New York City), NY; 3) Simons Foundation, Manhattan (New York City), NY; 4) New York University Center for Genomics and Systems Biology, Manhattan (New York University), NY; 5) Simons Center for Data Analysis, Simons Foundation, Manhattan (New York University); 6) Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA.

Existing methods for interpreting protein variation focuson annotating mutations as deleterious or neutral ratherthan detailed interpretation of variant effects and frequentlyuse only sequence-based or structure-based information. We present VIPUR, a computational framework seamlesslyintegrating sequence analysis and structural modeling(with the Rosetta protein modeling suite) to identify andinterpret deleterious protein variants. To train VIPUR,we collected 9,477 protein variants with known effectson protein function from multiple organisms and curatedstructural models for each variant from crystal structuresand homology models. VIPUR can be applied to mutationsin any organism's proteome with improved generalizedaccuracy and interpretability compared to other methods. We apply VIPUR to inflammation associated variantsand de novo mutations associated with autism spectrumdisorders, demonstrating the structural diversity of disruptedfunctional sites and improved interpretation of mutationsassociated with human diseases.

1594W

Big Data Warehouse for Large Scale Genomics. *B. Bernard[1], J. Slagel[1], S. Elasady[1], A. Black[2], G. Glusman[1], G. Eley[2], J. Vockley[2], I. Shmulevich[1], J. Niederhuber[2].* 1) Institute for Systems Biology, Seattle, WA; 2) Inova Translational Medicine Institute, Fairfax, VA.

Large-scale clinical genomics studies consisting of thousands of whole genome sequences require a system with the ability to analyze billions of unique variants over terabytes of data, generated with different technologies by disparate laboratories. Modern big data/cloud technologies, rather than traditional tab-delimited file-based approaches, are advantageous and increasingly necessary for scalable, high access genome analysis workflows and tools. Additionally, large collections of standardized annotations enable the evaluation and interpretation of genomic variation and improve the quality of analysis results. To this end we have developed a cross-platform, Hadoop-based system for distributed storage and query of annotations and large collections of genome sequences.

The system has been designed to integrate whole genome sequences from Illumina and Complete Genomics, with an extensible model for genomic data integration derived from multiple sequencing platforms. The genomic data model normalizes technical details such as variant call quality and no-calls; supports versioning and provenance of evolving annotation sources, genome builds, and sequences originating from different studies; and scales seamlessly to handle increases in content with minimal impact on performance. Over twenty annotation sources have been standardized, normalized, and integrated for deployment in several cloud-based or in-house platforms. These annotation sources are joined with sequencing data to support a variety of in-database analysis workflows and ad hoc queries, rather than static tab-delimited vcf files. Incremental and modular updates to the system are possible as new genome sequences are added to continuously growing collections and annotation sources are modified. Data replication of sequences and annotations, as well as indexing/partitioning schemes, are employed to enhance performance for scenarios ranging from deep analysis of family pedigree to association studies of large collections of genomes. Distributed queries can be coupled with parallel computation, for example with Python or R through industry standard APIs (e. g. , ODBC or GA4GH). As the use of genomics in the clinic grows, large scale, distributed, normalized data systems such as the one outlined here, will be imperative to continuously improve the accuracy and quality of personalized medicine.

1595T

Increased Statistical Power Using Informed Conditioning in Case-Control Studies. *M. Bilow[1], S. Eyheramendy[2], E. Eskin[1].* 1) UCLA, Los Angeles, CA; 2) Pontificia Universidad Católica de Chile, Santiago, Chile.

Genome-wide association studies (GWAS) have found thousands of associations between genetic variants and disease. Many of these studies have collected clinical phenotypes in addition to genetic data. Many of these clinical phenotypes correlate with disease status, and some genetic variants that affect disease have also been found to affect clinical phenotypes. Including multiple phenotypes in a single association test should make it easier to discover genes that affect clinical phenotypes and disease. As such, there is great interest in studying multiple phenotypes in a single GWAS. Unfortunately, most association tests used in GWAS in the literature analyze a single phenotype at a time. This is an inherently less powerful technique than joint analysis of multiple phenotypes when the genetic variant affects both case-control disease status and clinical phenotypes. A class of multivariate methods that combine multiple continuous phenotypes in a single association test have recently become more acknowledged within GWAS. However, it was shown recently that multivariate models can lose power in studies integrating continuous clinical phenotypes with discrete case-control data. Part of this loss of power is caused by multivariate models not fully Part of the reason for the loss of power is because individuals are selected to the study on the basis of their disease status. This results in a phenomenon known as selection bias, as selection of equal numbers of individuals affected by a rare disease and unaffected by the disease will oversample all of the factors that influence the disease's ontogeny. We present a novel technique that combines discrete case-control data with a continuous trait in a single multivariate model. This model, based on the liability threshold model, succinctly expresses an individual's likelihood of disease using a latent continuous variable and increases statistical power. We demonstrate the model increases statistical power in simulated data as well as data derived from the North Finland Birth Cohort dataset.

1596F

A resampling-based method for comparison of location parameters in multivariate datasets. *M. Borges[1,2], B. Carvalho[1,2], I. Lopes-Cendes[1,2].* 1) Department of Medical Genetics, State University of Campinas, Campinas, Sao Paulo, Brazil; 2) Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas, São Paulo, Brazil.

Experiments involve analyzing differences between groups. Despite the simplicity of concept, such evaluation includes a series of hypothesis assessments. The most common characteristic under scrutiny is the theoretical distribution of the data. However, other technical details, particularly high dimensionality, play a significant role in hypothesis testing. Ideal statistical procedures for group comparison must be robust to departures of the standard normality assumption and provide means of handling highly dimensional datasets. Nonparametric methods make no assumptions about probability distribution of the data and are natural candidates for robust strategies. Permutation tests comprise a powerful class of nonparametric strategies called resampling-based methods. Using this method, we estimate the distribution of the statistic of interest under the null hypothesis (i. e. , no differences between the groups) by randomly shuffling the group memberships of the observations. This allows the researcher to perform hypothesis testing without distributional assumptions. Projection methods are natural choices to address high dimensionality and to provide simpler data visualization. We developed our own comparison method by combining multidimensional scaling (MDS) with permutation tests. Our strategy uses MDS to project the data onto lower dimensions, while preserving the distance between the data points. Using these projections, we determine the statistic of interest: Mahalanobis distance between the groups. The method empirically estimates the null distribution of the distances between the groups using permutation. The strategy uses this distribution to assess the evidences of differences between the groups. This method is available as an R function at https://goo. gl/tEftzz.

1597W

Multilevel Regression Approach for Cross-Platform Transformation of Gene Expression Data. *Y. Cheung[1], N. Dimitrova[1], W. Verhaegh[2].* 1) Clinical Informatics Services and Solutions Depart, Philips Research, Briarcliff Manor, New York, USA; 2) Precision and Decentralized Diagnostics Department, Philips Research, Eindhoven, Netherlands.

The dynamic ranges of gene expressions can vary considerably depending on the choice of profiling platform. As a result, prognostic gene signatures are usually platform-specific, and expression data generated by heterogeneous platforms in general cannot be directly combined for computational analysis, thus limiting the scope of use of legacy data and hindering the adoption of new profiling technologies. Specifically, tremendous resources have been spent on microarray studies, and it is desirable to transfer the knowledge and insights onto the new platform such as next generation sequencing. Towards the goal of cross-platform compatibility of gene expression data, we propose a generalized multilevel regression framework that efficiently transforms gene expression data from the original to the designated platforms over a wide dynamic range. It can be flexibly adapted for any platforms, and supports out-of-sample data transformation after model training. As a first implementation, we developed a two-level framework that transforms gene expressions from Affymetrix HT-HG-U133A microarray (log2RMA) to Illumina HiSeq 2000 RNA-Seq (log2 RSEM) platforms based on 545 TCGA samples that have data for both platforms. For each sample, on average the correlation between the two platforms is $r = 0.71$ and the fractional root-mean-square (rms) error is $erms = 0.36$. Using 100 samples for training, after two levels of regression, both average correlation and fractional rms error are significantly improved: $r = 0.966$ and $erms = 0.092$. Moreover, principal component analysis showed that the transformed samples merge very well with the true RNA-Seq samples without systematic bias.

1598T

An automatic next-generation sequencing analysis pipeline for family-based disease studies. *R. -H. Chung, W. -Y. Tsai, C. Y. Kang, P. J. Yao.* National Health Research Institutes, Zhunan, Taiwan.

For disease studies using next-generation sequencing (NGS) data, family design has become an attractive approach to identifying rare mutations that are enriched in families. Several family-based algorithms and tools are available for analyzing sequencing data in families with disease. Integrating these resources into an automatic analysis pipeline will significantly facilitate the analysis. However, current analysis pipelines for family data mostly focus on a specific function. Therefore, a pipeline which can accommodate different analysis strategies for family-based analysis is desirable. We developed an analysis pipeline, FamPipe, which can be applied to the analysis for Mendelian disorders or complex diseases. Identity-by-descent (IBD) sharing statistics or linkage LOD scores calculated using Merlin are incorporated in FamPipe to identify linkage regions. For Mendelian disorders, segregation scores, which account for sequencing errors, are calculated in FamPipe. A disease model identification (DMI) algorithm is developed to classify variants into different disease models based on the segregation scores. For complex disease studies, family-based association tests can be performed in the linkage regions or across the genome. Furthermore, two family-based imputation tools, Merlin and GIGI, are integrated in FamPipe. We used simulation studies to evaluate the sensitivity and specificity for the DMI algorithm, IBD sharing statistics and linkage LOD scores. We also used simulations to compare the performance between the two family-based imputation tools. The simulation results showed that the DMI algorithm achieved specificity and sensitivity of 100% in the presence of genotyping errors when analyzing 10 nuclear or extended pedigrees consisting of both affected and unaffected individuals. Simulation results also showed that the IBD sharing statistics can have higher sensitivity of identifying linkage regions than linkage LOD scores when there are only a few families available, but linkage LOD scores can be more powerful when 10 families were analyzed. Finally, GIGI had higher imputation quality scores (IQS) than Merlin for large pedigrees consisting of 69 individuals in each pedigree. Similar IQS were observed for GIGI and Merlin for imputing pedigrees each with 12 individuals. In conclusion, an NGS analysis pipeline with commonly used functions for family-based disease studies is developed, which will expedite the analyses of NGS data in families.

1599F

An automated, differences-based model for retrospectively applying dynamic annotation data to static whole exome sequencing result sets. *D. Corsmeier[1], P. White[1,2].* 1) The Biomedical Genomics Core & The Center for Microbial Pathogenesis, The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) Department of Pediatrics, College of Medicine, The Ohio State University, Columbus, OH.

The vast majority of causative variants in Mendelian diseases are believed to be in the coding regions of the genome. However, whole exome sequencing studies typically have a molecular diagnosis rate of 25%, suggesting that many genes underlying these diseases have yet to be found and demonstrating that novel disease-causing gene discovery is still dynamic and rapidly expanding. While the underlying genomic data is static, literature and databases are constantly growing and periodic review is likely to help diagnose numerous additional cases. Current approaches to variant annotation create a snapshot and make the process of retrospectively examining existing data cumbersome. As such, an effort to periodically update annotations and apply them to these data sets requires complete reanalysis, a method lacking efficiency and practicality. In an effort to overcome this problem, we have implemented an automated system to apply evolving variant annotation to an existing repository of human genomic data. As new pathogenic variants are discovered, the corresponding annotation data are extracted and queried against the genomic data with the goal of identifying previously undiagnosed patients whose genetic etiology is revealed by the new discovery. This differences-based approach minimizes both compute resources and the effort required to review the new results. The report constructed from these queries includes primary findings and incidental, medically actionable variants as defined by the American College of Medical Genetics and Genomics. To demonstrate the utility of our system we used exome data sets that included variants for which annotation had changed in recent updated annotation database releases. Highly optimized queries were utilized to examine genomic loci across a sample data repository for cases in which annotation data changed and a report was generated accordingly. This process ensures that novel gene discovery information relevant to an existing patient's health or treatment will not be delayed or overlooked entirely.

1600W

Gene and pathogenic variant discovery for Mendelian and Complex Familial Traits. *H. Dai[1], G. T. Wang[1], B. Peng[2], R. L. P. Santos-Cortez[1], S. M. Leal[1].* 1) Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Bioinformatics and Computational Biology, The University of Texas - MD Anderson Cancer Center, Houston, TX.

We provide methods and implementation of best practices to identify rare variants involved in the etiology of Mendelian and familial complex traits using next-generation sequence (NGS) data. We demonstrate through case examples, bioinformatics protocols to analyze exome and genome sequence data to elucidate pathogenic variants that are either *de novo*, underlie Mendelian phenotypes or complex traits with familial aggregation. We feature four types of commonly adopted study designs analyzing either genomes or exomes from (1) a single affected individual (2) multiple family members, (3) multiple families and (4) multiple unrelated individuals with a family history of disease. For each design we illustrate the procedures to integrate data from different sources, performing variant annotations, and selecting potentially pathogenic variants from single or multiple exomes based on several parameters, including but not limited to mode of inheritance, variant sharing among pedigree members, population minor allele frequency, functional annotation and prediction, linkage mapping data and variants/genes previously implicated in disease etiology. Case studies were novel pathogenic variants have been discovered for nonsyndromic hearing impairment, thoracic aortic aneurysms and dissections, autism, Moyamoya disease, otitis media and rare autosomal recessive traits such as achromatopsia and trichothiodystrophy are used to illustrate the protocols and best practices. . We have also developed an easy-to-use bioinformatics software, Variant Mendelian Tools (VMT), to implements the protocols which makes it possible for any researcher with NGS data to efficiently hunt down and identify pathogenic variants. VMT is designed to be flexible to accommodate regular updates from annotation databases and incorporation of new information from a variety of sources including public databases and in-house data, e. g. linkage regions. The analysis protocols we developed are distributed under the VMT platform which can be readily adapted and shared for a variety of projects, owing to the compact, human-readable syntax that VMT adopts. Our work is highly beneficial to clinicians and researchers who aim to identify pathogenic variants from NGS data but have minimal knowledge and experience in the use of Linux and programming languages, and/or in annotations and variant discovery using family data.

1601T

The ClinGen Interface for Curating the Clinical Validity of Gene-Disease Associations: Specifications and Implementation. *S. Dwight[1], T. P. Sneddon[1], K. Liu[1], K. Dalton[1], F. Tanaka[1], B. Hitz[1], E. Riggs[2], O. Birsoy[3], A. Buchanan[2], R. Ghosh[4], N. Strande[5], S. Plon[4], H. Rehm[3], C. Martin[2], J. Berg[5], J. Michael Cherry[1], ClinGen Resource.* 1) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 2) Geisinger Health System, Danville, PA; 3) Laboratory for Molecular Medicine, Partners Healthcare Personalized Medicine, Cambridge, MA; 4) Pediatrics-Oncology, Baylor College of Medicine, Houston, TX; 5) University of North Carolina, Chapel Hill, NC.

The Clinical Genome Resource (ClinGen) is a NIH-funded program dedicated to creating a curated database of clinically relevant genes and genomic variants to inform genome interpretation in a variety of clinical contexts. It is anticipated that clinicians, researchers, laboratories and patients will benefit from this resource. A key element of ClinGen is the development of a suite of user interfaces that will facilitate expert curation of genes and variants, their association with disease and their clinical actionability. Here we report the development of one such tool in this suite - the gene curation interface. This interface will facilitate the curation process and guide the curator in determining the strength of a gene-disease association using the clinical validity classifications developed by the ClinGen Gene Curation Working Group (GCWG). This standardized classification scheme for clinically relevant genes optimizes their inclusion on clinical genetic testing panels and clinical interpretation of test results. Evidence supporting or contradicting a gene's association with disease occurs through manual literature curation of clinical and functional evidence. The specifications for the gene curation interface were developed iteratively in close working collaboration with members of the GCWG. The interface supports consistent application of the clinical validity classifications via careful design of the curator workflow, meaningful relationships between captured evidence, ontologies and controlled vocabularies, and free text descriptions that allow curator expressivity. The design supports gene curation by both "community curators" and ClinGen approved curators, with final review by an Expert Reviewer associated with a specific ClinGen Clinical Domain Working Group. As such, the completed gene curation interface is designed to provide ClinGen supported curation of gene-disease relationships by the larger genetics community. Additionally, the evidence fields captured by the interface will support the clinical validity classification and the assimilation of supplementary knowledge associated with the gene or disease, such as specific phenotypes involved, demographics, segregation, methodologies, and implicated variants. All of this curated information will be made available and searchable through a public web portal at www. clinicalgenome. org.

1602F

Efficient testing of multiple phenotypes in genome-wide association studies. *L. Gai[1], E. E. Eskin[1,2].* 1) Computer Science, UCLA, Los Angeles, CA; 2) Human Genetics, UCLA, Los Angeles, CA.

Expression quantitative trait loci (eQTL) mapping has emerged as a powerful tool for finding genetic variation that affects gene regulation. However, the wealth of information available through eQTL mapping comes with a great computational cost, especially when in combination with more sophisticated statistical models as such as mixed models. Unlike traditional genome-wide association studies (GWAS) where only a few clinical phenotypes (such as disease status or a biomarker) are studied at a time, a single eQTL study may involve tens of thousands of gene expression phenotypes. Since eQTL studies are generally conducted by running an independent GWAS on each expression phenotype, this results in billions of tests. Some of these tests are redundant, as some subsets of the phenotypes are not actually independent. Here we present we present Rapid Association-testing for Multiple Phenotypes (RAMP), a novel method to reduce the number of tests required in eQTL studies by leveraging the correlation between phenotypes. We show that RAMP greatly reduces the number of tests required for eQTL studies while maintaining statistical power. Finally, we apply RAMP combined with Genome-wide Rapid Association Testing (GRAT), a method to reduce the number of SNPs sampled per phenotype, to discover biologically relevant markers in an eQTL dataset. RAMP will be made available in a future release of pyLMM, a linear mixed model solver.

1603W

In silico Construction of Personalized DNA Methylation Biomarkers. *E. Gatev[1,2], M. Park[2,3], L. McEwen[2], J. MacIsaac[2], M. Kobor[2,4].* 1) Bioinformatics, University of British Columbia, Vancouver, Canada; 2) Centres for Molecular Medicine and Therapeutics, University of British Columbia, Vancouver, Canada; 3) Population and Public Health, University of British Columbia, Vancouver, Canada; 4) Medical Genetics, University of British Columbia, Vancouver, Canada.

DNA Methylation variability across individuals can be exploited to construct personalized biomarkers for disease and risk. This requires the integration of genetic and DNA methylation data. The typical scenario of a limited number of observation poses several challenges, including over-fitting, perormance assessment and interpretability of the results. We offer a novel approach to the construction of personalized DNA methylation biomarkers, that addresses these challenges. Our approach is based on constructions of informative reduced representations of the multiple big “Omics” data, and their integration in a novel way. We use a three stage pipeline with genomic feature construction and selection, relevant DNA-methylation probe selection and data integration to demonstrate a proof-of-concept cortisol personalized biomarker in a small cohort of 187 individuals.

1604T

Leveraging tumor lineage trees to predict and genotype somatic structural variations using paired-end sequencing. *I. Hajirasouliha[1], F. Hormozdiari[2], V. Popic[1], Z. Weng[1], R. B. West[1], A. Sidow[1], S. Batzoglou[1].* 1) Stanford University, Stanford, CA; 2) University of Washington, Seattle, WA.

We present LAPSI (Lineage-Aided Maximum-ParsimonyStructural VariationInference), a novel method for detecting and genotyping somatic structural variations (SVs) in multiple whole-genome sequencing tumor samples, taken from a patient. In contrast to standard SV discovery approaches, which do not leverage phylogenetic information, LASPI makes use of the multi-sample lineagetree structurereconstructedfrom deep sequencing somatic SNV datasets using a specialized method suchas LICHeE [1]. We assume that lineage trees based on different somatic variant classes are almost identical, and demonstrate that leveraging lineage trees reduces false negatives in detecting and genotyping of SVs. Our method effectively pools samples that share a common ancestor in the tree and finds clusters of discordant reads that suggest the same SV breakpoint across these samples. Placement of SVs onto specific branches of the lineage tree resultsin amorecomprehensive roadmap of the tumor’s genome evolution that begins at the zygote. LAPSI utilizes a combinatorial optimization formulation with the objectiveof assigning optimal SV clusterstothe branches of the tree,such that only a small fraction of discordant alignments are not assigned to any SV clusters, while sub-clonal SVs are favorably selected. LAPSI allows mixed-lineage samples to be present in the model, which makes the problem computationally NP-hard, and employs an effective approximation algorithm,based on the greedy *maximum coverage* problem. We tested LAPSI on multi-samplebreast cancerdatasetsand their corresponding lineage trees [2, 3]. Our initial investigation of thesedatasetsdemonstratesthe great potential of this method for detecting and genotyping somatic SVs in cancer genomes. For example, by investigating potential SVs in one of the patients, we have found a set of ~300 novel somatic deletions which were not present in the Lymph sample of the same patient. This set contains several *subclonal* deletions (i. e. those only present in a subset of samples). Furthermore, we wereable to identify a small somatic inversion with a relatively high support, whichwasnot seen in Lymph. **References**[1] Popic et al. , Fast and scalable inference of multi-sample cancer lineages. *Genome Biol.* 2015. [2] Newburger et al. Genome evolution during progression to breast cancer. *Genome Res.* 2014. [3] Weng et al. Cell-lineage heterogeneity and driver mutation recurrence in pre-invasive breast neoplasia,*Genome Med.* 2015.

1605F

***Bior_annotate* and *VCF-Miner*: An annotation and visualization framework designed for a constantly evolving genomic landscape.** *S. Hart[1], P. Duffy[2], M. Meiners[2], D. Quest[2], A. Hossain[2], M. Kalmbach[2], J-P. Kocher[1].* 1) Health Sci Res, Mayo Clinic, Rochester, MN; 2) Department of Research IT, Mayo Clinic, Rochester, MN.

Keeping up with the rapid pace of development in genomics is a non-trivial task. New algorithms to get more data out of sequencing experiments are published every week and existing tools can be updated at any time. Annotating the variants requires combining data from disparate external and in-house annotation sources, or command-line tools. While some applications have been developed to automate the annotation and filtering of genomics variants, these systems are highly restrictive. Maintaining existing annotations or adding new annotations is dependent on the authors’ availability or willingness to do so. Many times annotation and filtering are often integrated into toolkits, but lose track of provenance for where the data actually originated and how it has been modified. Because the Variant Call Format (VCF) is the standard reporting output format for many next generation sequencing experiments, we used the specification to build a flexible framework for adding new annotation and tool sets with minimal coding effort. The new *bior_annotate* application automatically splits large VCF files into smaller chunks for array processing and has several built-in checks for data consistency. Configuration files allow users to point to the specific catalogs (or versions of catalogs) that are needed. Because it is written in BASH, new command-line tools are also easy to add as optional processing operations. The inputs are a set of configuration files and a VCF, and the output is an annotated VCF file. This VCF File can be uploaded to *VCF-Miner*, a graphical user interface that sorts, filters, queries, and exports tables from VCF files. This allows rapid automatic integration of new and existing annotations from the VCF to be available to the end users, who many times are not computational biologists. When combined, these two tools make the process of adapting to an ever changing environment a much more tractable and reproducible task.

1606W

EpiCenter2: a powerful tool for RNA-seq and ChIP-seq data analysis. *W. Huang, L. Li.* Biostatistics and Computational Biology, NIEHS, Research Triangle Park, NC.

Cost-effective RNA-seq has quickly become a popular method for studying gene expression and their regulation mechanisms. The development of analysis methods and tools, however, still lags far behind the need. In particular, sequencing is now being used for more complicated or more challenging studies, and existing tools are no longer sufficient for these new and more advanced applications. For example, many early tools, while effective when dealing with data with few replicates, cannot take full advantage of more recent sequencing data with more replicates and/or more complex experimental designs. To meet such computational challenges, we introduce a new tool, EpiCenter2, for analyzing both RNA-seq and ChIP-seq data from the latest sequencing technologies. EpiCenter2, while based on our previous work EpiCenter, uses a new statistical framework that can take advantages of more biological replicates and gives more accurate predication of differentially expression genes or epigenetic changes. Specifically, EpiCenter2 can employ permutation method to assess statistical significance of difference between two groups. In addition, EpiCenter2 supports visualization of read coverage maps of any specific region or gene, and can be used for three-group comparison analysis. EpiCenter2 has been extensively tested in-house for real data analysis. Tests on simulated sequencing data also suggest that EpiCenter2 is accurate and efficient. EpiCenter2 is implemented in both C++ and R codes, and is computationally efficient by supporting multiple-thread computing. EpiCenter2 will be publicly available upon it is released.

1607T

FOAM: avoiding misleading variants in genomic regions Frequently Observed As Mutated. *C. Humphries[1], T. Ferrah[1], V. Dhankani[1], B. Bernard[1], D. Mauldin[1], L. Hood[1], M. Robinson[1], N. Clegg[1], J. Roach[1], P. Kothiyal[2], G. Eley[2], J. Vockley[2], J. Niederhuber[2], I. Shmulevich[1], G. Glussman[1].* 1) Institute for Systems Biology, Seattle, WA; 2) Inova Translational Medicine Institute, Fairfax, Virginia.

A well-known problem in genome analysis is that certain genes recurringly appear to be mutated - regardless of the phenotype. These include genes with many similar pseudogenes, large gene families such as olfactory receptors or immune receptors, poorly annotated genes, long intercellular matrix proteins that are very tolerant of missense mutations, and genes with extremely long coding sequences. It is common practice to discount variants that occur within these 'usual suspect' genes. However, this approach may result in discarding true causative variants. We developed a new method to identify error-prone genomic regions rather than whole genes. These regions are Frequently Observed As Mutated (FOAM) in individual genomes; variants within them should thus be treated with special caution. We analyzed high quality (>40x) whole-genome sequence (WGS) data from 7,879 individuals, sequenced on the CGI and Illumina platforms for the Inova Translational Medicine Institute (inova. org/itmi) and the Institute for Systems Biology (familygenomics. systemsbiology. net). We collected a joint set of 24M singleton variants (each observed in just one genome) and defined FOAM regions by clustering the variants based on distance and variant density. The FOAM regions span ~8% of the genome and overlap ~3. 5% of UCSC annotated exons. As expected, telomeric and centromeric satellite sequences are enriched in FOAM regions, as are highly similar gene families like olfactory receptors. Likewise, genes with internal sequence repeats overlap FOAM regions. For example, a FOAM segment overlaps the last three exons of the *KANK1* gene (i. e. , the ankyrin repeat domains) whereas it did not overlap the other, non-repeat exons. We identified a set of SNPs listed in the GWAS catalog that correlate to overlap with FOAM regions. Additionally, a review of the literature revealed that GWAS SNPs in FOAM regions tend not to be replicable in independent follow-up studies. We further enumerated pathogenic variants in ClinVar that overlap FOAM regions and may require further evaluation. Filtering observed genomic variants in FOAM regions is a more precise method for studying common false-positive genes, and informs about problematic regions in otherwise trustworthy genes. Fine-grained annotation of variants in FOAM regions leads to fewer spurious findings — a particular concern in clinical settings.

1608F

FIRE: functional inference of genetic variants that regulate gene expression. *N. M. Ioannidis[1], J. Davis[1], N. B. Larson[2], A. French[2], S. K. McDonnell[2], A. J. Battle[3], S. B. Montgomery[1], S. N. Thibodeau[2], W. Sieh[1], C. D. Bustamante[1], A. S. Whittemore[1].* 1) Stanford University School of Medicine, Stanford, CA; 2) Mayo Clinic, Rochester, MN; 3) Johns Hopkins University, Baltimore, MD.

Whole-genome sequencing technologies are increasingly enabling studies of genetic variation in large numbers of healthy and diseased individuals. However, interpreting the clinical significance of variants identified in these studies remains a critical challenge that has motivated the development of a variety of computational tools to aid in pathogenicity prediction. Current pathogenicity prediction tools for noncoding variants are not designed to provide insight into the biological functions of these variants upstream of their effect on pathogenicity. One of the major potential functional effects of noncoding variants is regulation of the expression and splicing of nearby genes. Here we present FIRE (Functional Inference of variants that Regulate Expression), a new tool trained to recognize expression quantitative trait loci (eQTLs) that enables prioritization of noncoding variants by predicting their potential to regulate the expression of nearby genes. We developed FIRE by training a random forest to recognize cis-eQTL variants using data from the Geuvadis consortium study of gene expression in the lymphoblastoid cell lines (LCLs) of 373 individuals of European descent. We find that FIRE achieves significant separation of scores for eQTL and non-eQTL variants using predictive variant features such as sequence conservation, overlapping functional elements from ENCODE and Ensembl, and position relative to the nearest transcription start site and splice site. We also show that FIRE scores for eQTLs identified in Yoruban individuals are similar to those for eQTLs in individuals of European descent, suggesting that FIRE scores are predictive across populations. Finally, we show that FIRE scores are only slightly lower for eQTLs identified in an independent study of 471 prostate tissue samples than for eQTLs in LCLs, suggesting that FIRE scores are predictive across tissue types despite a small degree of tissue specificity in the genomic features that characterize regulatory variants. We plan to deposit FIRE scores for all variants into the Clinical Genome Resource (ClinGen) knowledge base, and we anticipate that these predictions will be useful in future studies of regulatory variation and for prioritizing the likely clinical significance of rare genetic variants identified in large-scale clinical sequencing studies.

1609W

Turning Publicly Available Gene Expression Data into Discoveries Using Gene Set Context Analysis. *Z. Ji[1], S. A. Vokes[2,3], C. V. Dang[4], H. Ji[1].* 1) Department of Biostatistics, Johns Hopkins University, Baltimore, MD; 2) Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX; 3) Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX; 4) Abramson Cancer Center, University of Pennsylvania, Philadelphia, PA.

Gene Set Context Analysis (GSCA) is an open source software package to help researchers use massive amounts of publicly available gene expression data to make discoveries. Users can interactively visualize and explore gene and gene set activities in 25,000+ consistently normalized human and mouse gene expression samples representing diverse biological contexts (e. g. , different cells, tissues and disease types, etc.). By providing one or multiple genes or gene sets as input and specifying a gene set activity pattern of interest, users can query the expression compendium to systematically identify biological contexts associated with the specified gene set activity pattern. In this way, researchers with new gene sets from their own experiments may discover previously unknown contexts of gene set functions and hence increase the value of their experiments by making more discoveries. GSCA has a graphical user interface (GUI). The GUI makes the analysis convenient and customizable. Analysis results can be conveniently exported as publication quality figures and tables. GSCA is available at https://github. com/zji90/GSCA. This software significantly lowers the bar for biomedical investigators to use PED in their daily research for generating and screening hypotheses, which was previously difficult because of the complexity, heterogeneity and size of the data.

1610T

An automatic end-to-end solution for disease-causing variant detection in rare and hereditary diseases with a high case solve rate and a much reduced false positive rate. *A. Joecker[1], S. Shah[2], D. Basset[2], R. Yip[2], D. Richards[2], B. Oester[1], G. Eley[3], B. Solomon[3], J. G. Vockley[3].* 1) QIAGEN Aarhus, Aarhus, Denmark; 2) QIAGEN Silicon Valley, Redwood City, USA; 3) Inova Translational Medicine Institute, VA, USA.

Next generation sequencing technology has enabled identification of causal genetic variants underlying rare inherited diseases in a short time. However, data analysis and interpretation of the large amounts of data is still a bottleneck, as causal variants are often not detected or, when detected, may include false positives that need to be ruled out through additional experiments. We present here a new streamlined automatic solution workflow, which enables the detection and interpretation of causal variants from whole genome and whole exome data with a higher case solve rate and a much reduced false positive rate. The new solution combines new functionaltity of Biomedical Genomics Workbench (BxWB) with Ingenuity Variant Analysis to provide a streamlined end-to-end workflow from FASTQ files to disease-causing variants. In addition to the normal variant calling from data of mother, father and child, the new functionality in BxWB includes a back-check to create a list of all variants found in all three individuals to look up the mapped sequencing reads and identify missing variants. This approach reduces the false positive rate of the end result significantly as many potential *de novo* variants can be excluded as inherited. It also increases sensitivity particularly for variants with low coverage. When used together with QIAGEN's Ingenuity Variant Analysis, that integrates the Ingenuity Knowledge Base and HGMD, in an automatic workflow, the workflow not only identifies in almost all cases the disease-causing variants in the tested cases, but also significantly reduces the false positive rate. In this study, we analyzed 15 TRIOs from the INOVA clinic (www. qiagenbioinformatics. com/inovagenomes) using the new solution and will show the results of this experiment in detail.

1611F

HITseqClust: Analysis of HIV-1 Integration Targeting-sequencing (HIT-seq) data to identify gene clusters targeted by specific transcription factors. *R. C. Johnson[1], G. W. Nelson[1], M. C. Cam[2], X. Wu[3], S. H. Hughes[4].* 1) CCR Collaborative Bioinformatics Resource, Frederick National Laboratory, Frederick, MD; 2) CCR Collaborative Bioinformatics Resource, National Cancer Institute, Bethesda, MD; 3) Leidos Biomedical Research, Inc, Frederick National Laboratory for Cancer Research, Frederick, MD; 4) HIV Drug Resistance Program, National Cancer Institute, Frederick, MD.

Lens epithelium-derived growth factor (LEDGF) is a transcription coactivator that assists HIV integration into the host genome. It has been shown that HIV integration can be redirected by replacing the N terminus, which binds chromatin, with chromatin binding domains from other transcription factors. The use of HIV Integration Targeting (HIT-seq) can shed valuable light into the function of these factors. After HIV integration sites have been sampled by next generation sequencing the relative density of integration sites is calculated using a gaussian kernel density estimator, and genes with high a density of integrations sites are flagged for each transcription factor variant. Genes targeted by individual transcription factor variants or groups of transcription factor variants (clustered using a principal components analysis of integration density among all genes) are compared using a pathway analysis, with a focus on both gene sets similarly targeted and those differentially targeted by different transcription factor variants. We revisit a published data set looking at H3K4me3 interactions with TAF3 to identify groups of genes differentially targeted by wild type TAF3 or the M880A mutant of TAF3. We also present a subset of genes that are not differentially targeted by only one variant, but are strongly associated with both. HITseq is a highly specific method for identifying genes targeted by various transcription factors. The software presented here identifies genes targeted by various transcription factors and compares gene clusters targeted by different variants of these transcription factors as well as differences and similarities among cell lines treated with identical variants.

1612W

Model selection via iterative hard thresholding for genome-wide association studies. *K. L. Keys[1], G. K. Chen[2], K. L. Lange[1,3].* 1) Biomathematics, UCLA, Los Angeles, CA; 2) Biostatistics, USC, Los Angeles, CA; 3) Human Genetics, UCLA, Los Angeles, CA.

Genome-wide association studies (GWASes) examine genetic variation between two groups of patients distinguished by a measurable phenotype. Each patient is genotyped at several places on the genome, and the genetic variation between cases and controls is compared. Since GWASes were introduced in 2005, researchers have performed GWASes for hundreds of traits on thousands of individuals. GWASes produce massive quantities of data that present computational and model selection challenges to their analysis. Current statistical software performs penalized regression, such as LASSO or MCP, to select the handful of informative SNPs from the thousands of predictors. Unfortunately, these methods introduce SNPs whose spurrious relationships to the trait obfuscate its genomic underpinning. We exploit iterative hard thresholding (IHT) to effectively select with great confidence the genetic markers most informative for the trait. Our parallel implementation exploits GPUs and PLINK data compression to fit GWAS data in computer memory. By leveraging commodity laptop computer hardware, we enable GWAS analysis on desktop machines and discard the need for expensive supercomputing resources. We demonstrate algorithmic performance on both simulated and real GWAS data. Our tests suggest that IHT effectively controls statistical type I errors better than LASSO and MCP regression.

1613T

Automated Identification, Prioritization and Visualization of Large-Scale Genomic Data. *M. J. Kiel[1,2], N. Patel[1,2], R. W. Peng[3], S. Schwartz[2,3], M. S. Lim[1,2,4], K. S. J. Elenitoba-Johnson[1,2,5].* 1) Department of Pathology, University of Michigan, Ann Arbor MI; 2) GENOMENON, Ann Arbor MI; 3) AlfaJango, Ann Arbor MI; 4) Department of Hematopathology, University of Pennsylvania, Philadelphia PA; 5) Center for Personalized Diagnostics, University of Pennsylvania, Philadelphia PA.

The tertiary component of genome sequence analysis requires identification of informative disease-gene-mutation associations for extraction of clinicobiological meaning from patient data. The accuracy and efficiency of tertiary analysis is limited by inaccessibility and non-uniformity of this information in both public databases and primary articles. To overcome these limitations, we have developed MASTERMIND - asuite of novel analytic tools that dramatically reduces the time and effort required to organize and integrate genomic information from any data source including millions of full-text scientific articles and dozens of heterogeneous variant databases. To comprehensively interrogate genomic data from both structured and unstructured databases, an automated querying architecture was designed using customized open-source analytics engines and a combination of publicly available and custom-developed APIs. Curated lists of diseases and gene transcriptswith synonyms comprising 11. 7K and 50. 9K total entrieswere used as initial query parameters. Custom-designed algorithms were used to generate comprehensive mutation query lists comprising602M total entriessorted by biological outcomeand used assecond-tier queries. Using titles andabstracts of 24M primary articles, we identified 909K putative disease-gene associations which we then confirmed by automated scanning of5. 8M full-text articles to comprehensively identify all disease-mutation citations within each article. Integrated metadata for each finding was used to prioritize disease-gene-mutation associations in accordance with the abundance and quality of supportingevidence. To facilitate rapid comprehension, theseassociationswere then organized withina singular graphical UIwith display of all relevantinformation from the primary source material used to drive data prioritizationincluding interactive access to annotated full-text articles. This method of rapidly assimilatingdisease-gene-mutation-associations for displaywas reproduced for several additional structured databases including ClinVar. In summary, MASTERMIND rapidly and comprehensively interrogates, organizes and displaysgenome data and has promising applications in expediting tertiary analysis of human genome sequencing data in clinical assays of individual patients.

1614F

SoloDel: A probabilistic model for detecting low-frequent somatic deletions from unmatched sequencing data. *S. Kim[1], J. Kim[2], H. Nam[3], S. Kim[2], D. Lee[4].* 1) SMRI, Rockville, MD; 2) Yonsei University College of Medicine, Seoul 120-752, Korea; 3) Gwangju Institute of Science and Technology, Gwangju 500-712, Korea; 4) KAIST, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Korea.

Finding somatic mutations from massively parallel sequencing data is becoming a standard process in genome-based biomedical studies. There are a number of robust methods developed for detecting somatic single nucleotide variations (SNVs). However, detection of somatic copy number alteration (SCNAs) has been substantially less explored and remains vulnerable to frequently raised sampling issues: low frequency in cell population and absence of the matched control samples**.** We developed a novel computational method SoloDel that accurately classifies low-frequent somatic deletions from germline ones with or without matched control samples. We first constructed a probabilistic, somatic mutation progression model that describes the occurrence and propagation of the event in the cellular lineage of the sample. We then built a Gaussian mixture model to represent the mixed population of somatic and germline deletions. Parameters of the mixture model could be estimated using the expectation-maximization (EM) algorithm with the observed distribution of read-depth ratios at the points of discordant-read based initial deletion calls. Combined with conventional structural variation caller, SoloDel greatly increased the accuracy in classifying somatic mutations. Even without control, SoloDel maintained a comparable performance in a wide range of mutated subpopulation size (10% to 70%). SoloDel could also successfully recall experimentally validated somatic deletions from previously reported neuropsychiatric whole genome sequencing data.

1615W

FMAP: Functional Mapping Analysis Pipeline for Comparative Metagenomics and Metatranscriptomics. *J. Kim[1,2], M. Kim[1,2], A. Koh[2,4,5], Y. Xie[1,2,3], X. Zhan[1,3,6].* 1) Quantitative Biomedical Research Center, University of Texas Southwestern Medical Center, Dallas, TX; 2) Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, TX; 3) Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX; 4) Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX; 5) Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, TX; 6) Center for Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX.

FMAP is an open-sourced stand-alone functional analysis pipeline for analyzing whole metagenomic and metatranscriptomic sequencing data. FMAP assembles a publically available UniProt reference database, which improves the sequence alignment outcome compared with other existing pipelines. Integrated statistical analysis procedures can identify differentially abundant genes and enriched functional pathways between sample groups. These results can be easily visualized through heatmaps and functional pathway diagrams. In addition, FMAP provides novel operon-level analysis in order to better interpret microbial functions.

1616T

Improving haplotype phasing accuracy using many short IBD segments. *A. Kleinman, E. Durand, C. McLean.* 23andMe, 899 West Evelyn Ave. , Mountain View, CA 94041.

Haplotypes are required for many genetic applications, such as imputation and fine mapping. Here we present Origin, a novel method for leveraging IBD data to improve phasing accuracy. Origin calculates the most likely haplotype matches for every IBD segment spanning a genomic region and integrates switch error predictions from all IBD segmentsusing a weighted voting method. We extracted over 350,000 individualsfrom the 23andMe dataset. We phased all individuals with Finch, a modification of the Beagle phasing algorithm that separates the haplotype graphconstruction and phasing steps. We then computed IBD across all pairs of individuals using GERMLINE, used HaploScore to filter out false-positive segments, and then ran Origin using this data to rephase. We assessed phasing accuracy by comparing Finch andOriginphasing to the gold-standard trio phasing in 3,000trio children. Origin decreased the global switch-error rate by 25-30%, with increased IBD coverage corresponding to better performance gains. The algorithm can be easily modified to run on genotypes phased and IBD calculated by other methods, and so is widely applicable to population-scale datasets.

1617F

coalescentSTR: a statistical approach for short tandem repeat number estimation based on coalescent theory from high-throughput sequencing data. *K. Kojima[1], Y. Kawai[1], N. Nariai[2], T. Mimori[1], T. Hasegawa[1], M. Nagasaki[1].* 1) Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan; 2) Institute for Genomic Medicine, University of California San Diego, San Diego, United States.

In the estimation of repeat numbers in a short tandem repeat (STR) region from high-throughput sequencing data, two types of approaches are usually considered: approaches counting repeat patterns in sequence reads aligned to the region and paired-end read based approaches that consider the difference between the insert size inferred from each aligned read pair and its actual size. Although the former approaches can estimate repeat numbers accurately, STR regions with the size more than the sequence read length cannot be handled. On the other hand, the latter approaches can handle STR regions longer than the length of sequence reads. However, the accuracy of repeat numbers estimated with the latter approaches is not high, compared with those with the former approaches. Thus, we propose a statistical approach named coalescentSTR that considers the unobserved genealogy of multiple individuals and estimates their repeat numbers from paired-end read distances. The genealogy is represented by coalescent trees sampled from phased genotypes around a target STR region with Markov chain Monte Carlo. By propagating belief of repeat numbers from insert size data in the sampled coalescent trees, more accurate repeat number estimation is expected for STR regions longer than the length of sequence reads. For the estimation of repeat numbers, a new belief propagation algorithm is proposed to search repeat numbers maximizing the model likelihood, while handling the coalescent trees as hidden variables. The effectiveness of our approach was verified from the comparison with lobSTR, RepeatSeq, and STRViper on simulation datasets and real exome data for 33 JPT individuals in the 1000 Genomes Project.

1618W

Bayesian learning of risk variants and functional enrichments in pleiotropic traits. *Y. Li.* MIT, Cambridge, MA.

Genome wide association studies have provided valuable information to decipher the underlying causal variants across many common complex human diseases. However, we are often underpowered due to sample size, rare variants, linkage disequilibrium, etc. One promising direction is to integrate facets of functional genomic data with GWAS summary statistics across several related dieases to (1) re-prioritize GWAS hits; (2) detect functional enrichment of risk variants; (3) identify pleiotropic causal variants in multiple related disorders. In this regard, we present a novel Bayesian framework to jointly learn the underlying disease-associated signals across multiple disorders. Our model demonstrates promising performance on both simulation and applications to real GWAS datasets.

1619T

Improving Specificity in Ion Proton Data. *D. S. Lieber[1], N. Chennagiri[1], E. White[1], D. Brinza[2], J. Veitch[2], T. Yu[1], J. F. Thompson[1].* 1) Claritas Genomics, Cambridge, MA; 2) Thermo Fisher Scientific, San Francisco, CA.

The Ion Proton instrument by Life Technologies is a chip-based sequencing technology that can produce whole-exome scale next generation sequencing data within hours. A critical part in development of a sequencing platform involves improving specificity via data filtering and thus we have worked to improve Ion Proton variant calls using a dual strategy. We first created and applied a strand-specific error (SSE) filter, whereby variants with strand bias across multiple samples were removed. We additionally created a filter using data derived from over 8,000 exomes sequenced at Claritas Genomics and compared the allele frequencies to the ExAc database. Any variant with high allele frequency in the Claritas database but low allele frequency in ExAc was deemed to be a likely false positive. Using NA12878 NIST data to benchmark performance, these methods allowed us to reduce the number of false positive SNPs by 3. 8-fold and indels by 5. 6-fold, with a 0. 1% loss in sensitivity for both SNPs and indels. These results demonstrate that Ion Proton variant call accuracy can be significantly improved through downstream bioinformatics processing.

1620F

Low-cost ancestry inference using targeted sequencing. *B. Liu[1], S. Shringarpure[1], C. Gignoux[1], R. Zhang[2], K. S. Smith[1], C. D. Bustamante[1], S. B. Montgomery[1], D. Vollrath[1].* 1) Pathology, Stanford University, Palo Alto, CA; 2) Sun Yat-Sen University, Guangzhou, P. R. China.

To control for population structure in genetic studies, one can preselect specific individuals based on self-reported ancestry. However, such information may not be available at the time of study, or may be inaccurate in cases where subjects are from admixed populations. Under such circumstances, it is desirable to genotype study participants for the purpose of ancestry inference. Current platforms for ancestry determination mostly use array technology. Such technology is cost-efficient on a production scale, where thousands of samples need to be genotyped. However, microarray platforms become costly on a laboratory-scale, where less than 100 individuals need to be genotyped for a more targeted set of ancestry informative markers. For example, Illumina's lowest-cost microarray platform, HumanCore-24, charges $88 – $98 per sample for the chip alone. If the project requires more variants than packaged on the chip, custom markers can be added at a minimum of 1,152 samples and 1,000 markers at an additional cost. This is often greater than required when fewer than 100 individuals need ancestry inferred. We have developed a targeted sequencing pipeline to determine ancestry proportion using ancestry informative markers (AIMs). Our approach consists of the following steps. We first ascertain the ancestral common variants from the 1000 Genomes Project Phase 3 catalog. We iteratively add markers to each population of interest to ensure a set of AIMs with a balanced cumulative locus specific branch length (LSBL) for each population. We designed primers pools for the AIMS sites iteratively using yamPCR. Targeted amplifications are performed on the Fluidigm Access Array platform. Amplicons are then barcoded by sample and sequenced to a depth of 100-1000 reads per site. We infer ancestry proportions using the likelihood frameworkof ADMIXTURE. As a proof of principle, we used this approach to select 123 SNPs that explain more than 99% of ancestry proportions between African and European populations. We performed targeted amplification on 12 admixed African American samples genotyped by the 1000 genomes consortium. The true ancestry proportions are recapitulated at a high proportion using the 123 AIMs sites alone (R2 = 0. 99). Compared to commercially available microarray platforms, this pipeline can be completed under comparable time scale, is several-fold cheaper ($17 - $55, depending on the sample size), and offers competitive accuracy in ancestry inference.

1621W

Improved Algorithm for Amplicon Sequencing Assay Designs. *G. Liu, M. Manivannan, C. Lou, A. Atkins, K. Rhodes, A. Broomer, F. Hyland, M. Andersen.* Thermo Fisher Scientific, South San Francisco, CA.

Ion AmpliSeq™ sequencing is one of the most promising applications of the Ion Torrent NGS platform. It involves multiplex PCR for target enrichment. Thermo Fisher offers online Ion AmpliSeq Designer to customers to assist assay designs. While more and more people are adopting Ion AmpliSeq technologies, challenges for assay designs started to emerge. Here we present bioinformatics approaches to improve the following areas of assay design: 1) assay specificity; 2) primer quality control; 3) SNP under primer; and 4) flexibility to adapt to different applications of Ion AmpliSeq sequencing including variant calling, copy number variation detection, and metagenomics. Design algorithms are developed to ensure high coverage with controlled risk of amplification efficiency, off-target reads and SNP effects. With the optimized design algorithm, numerous community and customer panels have been created, including the Ion AmpliSeq Exome Panel, TP53 Panel, and CFTR Panel. Design details of those panels are also presented. *For Research Use Only. Not for use in diagnostic procedures.* .

1622T

Normalization for correcting systematic variation in microarray genotyping. *J. Liu, B. Eynon, J. Brodsky, J. Gollub, T. Webster.* Affymetrix, Inc. , 3420 Central Expressway, Santa Clara, CA 95051.

Genotyping on array platforms has very high accuracy, exceeding 99. 5% across a large number of markers and samples. Occasional, systematic, artifactual variation in genotypes of a small set of markers can arise from non-genetic sources, including differences in sample collection and handling; laboratory, instrumentation, and reagents used in the genotyping assay; time period of processing; or other environmental influences. This can result in biased allele measurements and lead to incorrect genotype calling. Advanced normalization techniques to mitigate these systematic variations can significantly improve the quality of results.

We have developed a normalization approach to remove systematic effects arising from identified sources from genotyping data. Log-linear regression is used to quantify and subtract identified effects from the measured intensities for each allele. A critical assumption of this approach is that the identifiable effect being mitigated is not confounded with genetically relevant differences. For example, if a systematic effect is introduced by one or more collection centers, it is critical that genetically relevant factors such as population of origin and case-control status be randomized across centers so that meaningful information is not removed along with the irrelevant bias.

To evaluate our normalization approach, we examined a set of approximately 1,100 samples genotyped using Axiom® Genotyping Solution. Samples were arranged on Axiom® array plates: (1) with population of origin strongly confounded with the array plate, and (2) with samples randomized across the array plates, with no confounding. Genotyping was performed with no adjustment of signals and with our advanced normalization using plate of origin as a possible source of bias. Results from the experiment showed that normalization improved concordance to 1000 Genomes Project genotypes for a small set of probesets with performance below standard quality metrics in the non-normalized data. 1. 2% of probe sets were filtered out using Axiom® SNP quality control methods (SNPolisher) from Affymetrix. Normalization improved performance for 30% of the probesets below QC thresholds, and their concordance increased from 92. 4% to 98. 4%. However, when the randomization assumption was violated, this approach was not successful. We conclude that normalization to remove systematic biases can be a useful tool when coupled with appropriate experimental design.

1623F

APIGenome: A public library for big data genomic analysis tools. *A. Liu, H. M. Kang.* Department of Biostatistics, University of Michigan, Ann Arbor, Ann Arbor, MI.

Querying over large-scale genomic data requires many different computational tools for each particular questions of interest. In many cases, in-house software tools that were written to answer to a particular type of task can be recycled for answering different types of questions by others. Simple examples tasks include extracting particular subset of samples or markers from a VCF file, producing basic summary statistics from a VCF file, comparing genotypes between VCF files, plotting allele frequency spectrum from a VCF file. Here we present APIGenome, a public library of software tools for analyzing large-scale genomic data. APIGenome is a collection of reusable genomic analysis software tools hosted as a github repository and open to contribution from public. APIGenome provides generic interfaces that allow automated documentation of software (wGetOptions) and seamless parallelization or a large number of tasks (run-make). These features substantially reduce the burden of software maintenance and facilitate further extensions through collaborative contribution from the community with minimal effort. As of June 2015, APIGenome includes a rapid VCF analysis tool written in C++ (vcfast), genotype comparison and visualization tool between VCF and/or PLINK-formated files (bed-diff), visualization tools of allele frequency spectrum (draw-afs), likelihood-based variant filtering tool from VCFs of family-based samples or duplicate samples (MiFt), various utilities for VCF summarization or manipulation (vcf-summary, vcf-add-rsid, vcf-lookup-rsid, vcf-resolve-chrX-hets, vcf-issac-summary), and software summarizing QC metrics from a large-scale sequencing study (now-seq-scan). We expect that APIGenome will continually grow as a library of practical utilities for big data genomic studies.

1624W

WGSA: an annotation pipeline for human genome sequencing studies. *X. Liu[1,2], S. White[3], B. Peng[4], A. D. Johnson[5,6], J. A. Brody[7], A. H. Li[1], A. Carroll[8], Z. Huang[3], N. Rustagi[3], P. Wei[1,9], R. Gibbs[3], F. Yu[3], R. J. Klein[10], E. Boerwinkle[1,2,3].* 1) Human Genetics Center, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA; 2) Department of Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 4) Department of Bioinformatics and Computational Biology, University of Texas MD Anderson Cancer Center, Houston, TX, USA; 5) NHLBI Framingham Heart Study, Bethesda, MD, USA; 6) Population Sciences Branch, NHLBI Division of Intramural Research, Bethesda, MD, USA; 7) Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA; 8) DNAnexus, Mountain View, CA, USA; 9) Department of Biostatistics, School of Public Health, University of Texas Health Science Center at Houston, Houston, TX, USA; 10) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, Icahn Institute for Genomics and Multiscale Biology, New York, NY, USA.

Large-scale sequencing technologies have made unprecedented progress during the recent years. We are witnessing an exciting era when genetic variants—including single nucleotide variants (SNVs) and insertion-deletion variants (indels)—in both healthy human subjects and those with Mendelian or complex diseases can be comprehensively identified by re-sequencing studies. Unfortunately, there is a large gap in the ability to ascertain enormous amounts of variants and the ability to understand their functionality. This gap will grow even wider as we transition from whole exome capture sequencing to whole genome sequencing (WGS). There are currently few tools available to integrate coding and non-coding functional annotation resources and provide a convenient and efficient pipeline for annotating millions of variants discovered in a WGS study. As a first step to narrow this gap, we developed WGSA, an annotation pipeline for human genome sequencing studies, to facilitate the functional annotation step of WGS. Currently WGSA supports the annotation of SNVs and indels locally without remote database requests, so it can be scaled up for large studies. Compared to existing annotation tools, our study has the following unique aspects:1. We integrated annotations from three popular annotation tools (ANNOVAR, SnpEff and VEP) versus two popular gene model databases (Ensembl and RefSeq). A recent study has shown that the consequence prediction of a variant often varies depending on the underlying gene model and the annotation tool used. Therefore, it is suggested to obtain annotations from multiple tools across multiple databases for a more complete interpretation of the variants. Here we provide a one-stop resource to achieve that goal. 2. To further speed up the process, we precomputed the annotations for all potential human SNVs (a total of 8,584,031,106) based on human reference hg19 non-N bases and use it as a local database. 3. We integrated many annotation resources for both coding and non-coding variants, including functional prediction scores (CADD, FunSeq, etc), population allele frequencies (ExAC, UK10K, etc), conservation scores, mappability, and annotations from epigenomic projects (ENCODE, Roadmap, etc). 4. To provide convenience to a broader community, we have built an Amazon Machine Image for running WGSA on the cloud from Amazon Web Services, in addition to a downloadable version. More information can be found at https://sites. google. com/site/jpopgen/wgsa.

1625T

Slippage-Associated Repeat Identification and Analysis (SARIA) *in silico* at Sites of Chromosomal Structural Change. *S. Ma[1], N. Pannunzio[2], K. Wang[3], M. R. Lieber[4].* 1) Cancer Biology and Genomics Program, Keck School of Medicine, University of Southern California, Los Angeles, CA; 2) Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA; 3) Zilkha Neurogenetic Institute, Biostatistics Division, Keck School of Medicine, University of Southern California, Los Angeles, CA; 4) Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA. lieber@usc. edu.

DNA structural aberrations, such as translocations, insertions and deletions, often have a non-random distribution in the human genome, but are predisposed to occur in certain regions. Despite this, the relation between structural abnormalities and their surrounding sequences has not yet been fully studied. One major cause of these structural abnormalities is transient DNA slippage, which usually occurs during replication or transcription, when the DNA duplex melts and re-anneals. One common cause of DNA slippage is mis-alignment at short, closely spaced DNA repeats (direct or inverted, less than 15bps in length, distance between repeats less than 15bps) due to passage of a polymerase. To precisely illustrate the distribution of sequences prone to DNA slippage throughout the human genome, and to determine the differences in DNA repeat locations between healthy individuals and patients with various diseases, we have developed a python program: Slippage-Associated Repeats Identification and Analysis (SARIA). Because of the nature of DNA slippage, our program focuses on searching the surrounding sequences for repeats. Our program includes convenient user defined parameters, such as minimum/maximum distances between repeats, shortest/longest repeat length, and percentage of mismatches allowed. It provides follow-up analysis support, such as G/C content calculation, labeling of nearby histone modification sites, and statistical comparisons between different inputs. Together with its visualization functions, SARIA will provide users with a straight forward output displaying the repeat sequences and their genomic locations. Currently, there is no existing software serving to extract such information for neighboring repeat pairs in a large genome range. Using SARIA, one can easily locate potential DNA slippage sites in genome regions of interest, and examine whether or not slippage is a possible cause of genome structural changes. This program seeks to provide insights at the DNA level in diseases, such as cancer, which involve DNA abnormalities.

1626F

A pairwise genomic distance measure to evaluate the effect of donor/recipient genomic proximity on unrelated stem cell transplantation. *A. Madbouly, V. Paunic, M. Albrecht, M. Maiers.* Bioinformatic Research, National Marrow Donor Program, Minneapolis, MN.

Survival after hematopoietic cell transplantation (HCT) is dependent on donor/recipient (D/R) HLA matching. However, other factors such as D/R non MHC admixture can affect outcomes. To investigate these effects, we developed a robust pairwise D/R genomic distance that captures signal variability while minimizing signal-to-noise ratio. The study was designed in two phases. We initially genotyped 300 D/R pairs for 500 autosomal ancestry SNPs using SequenomiPLEX assay. In the 2nd phase 995 D/R pairs were genotyped on the HumanOmniExpress BeadChip array, which contained 307 of the 500 pilot SNPs. The remaining 193 SNPs were imputed to the 1000 Genomes reference panel using IMPUTE2. After running basic sample QC (SNP and sample call rate >98%, MAF > 1%), gender concordance was run using X-chromosome homozygosity. Twelve samples were found discordant with reported gender and were further evaluated for sample mix-up. All gender discordant samples were disregarded from further analysis. Pairwise D/R Euclidean distance was evaluated on PCA eigen vectors by varying the number of PCs in each vector and comparing correlation with distance from raw genotypes (ʃ). Selected distance vector size (N=88 PCs) was chosen such that the first derivative of the correlation ʃ is minimized, thus minimizing signal-to-noise ratio. Average PCA distance was 11. 67 (range [1. 83, 19. 5]). Average distance was higher for D/R pairs with mismatched versus matched reported race (N-mismatched=160, mean = 12. 2, [8. 09, 19. 2]), (N-matched= 1075, mean = 11. 6, range [1. 83, 19. 4]). Minimizing signal-to-noise ratio in the developed distance measure can potentially help reduce the effect of confounding factors on subsequent outcomes analyses. A multivariate analysis is currently being conducted to evaluate the association of D/R genomic distance and D/R admixtures with multiple HCT outcomes.

1627W

SPRITE: A Fast Parallel SNP Detection Pipeline. *K. Madduri[1], V. Rengasamy[1], P. Medvedev[1,2].* 1) Computer Science and Engineering, The Pennsylvania State University, University Park, PA; 2) Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA.

Variant calling analysis of whole genome sequencing data is a computationally-intensive process, often taking several hours to even multiple days. We develop SPRITE, an open-source toolkit for detecting Single Nucleotide Polymorphisms (SNPs). By combining fast new algorithms and parallel computing technologies, SPRITE significantly accelerates the computational tasks of alignment, intermediate file processing, and SNP detection. For a benchmark human genome with 30X coverage, SPRITE takes just 30 minutes (from FASTQ data ingestion to VCF file creation) on a compute cluster with 16 servers. We also evaluate SPRITE on several Illumina Platinum Genome data sets. PARSNIP, the SNP detection tool in SPRITE, is orders-of-magnitude faster than the SNAP callers in GATK, samtools, and Freebayes. In our preliminary analysis, we find that that the quality of results obtained (PARSNIP precision and recall using high-confidence variant calls as ground truth) is comparable to state-of-the-art SNP-calling software. A prototype implementation of SPRITE is available at sprite-psu. sourceforge. net.

1628T

Alignment, phasing and structural variant detection with linked reads. *P. Marks, M. Schnall-Levin, S. Kyriazopoulou-Panagiotopoulou, H. Heaton, D. Stafford, G. X. Y. Zheng, S. Saxonov.* 10X Genomics, Pleasanton, CA.

We have developed algorithms for the GemCode platform from 10X Genomics to recover long-range information from short-read sequencers. Our approach exploits the throughput and accuracy of short-read sequencing while enabling haplotype phasing and vastly improved structural variant detection and short-readalignment. In the GemCode system, long fragments from 1ng of DNA are randomly partitioned into hundreds of thousands of individually-barcoded microfluidic droplets (~3Mbp/droplet) from which libraries are made that are compatible with standard exome capture. Many linked reads cover each input fragment, providing a powerful new data for novel algorithms that enable phasing and structural variant detection from WGS and WES samples. Our barcode-aware alignment algorithm exploits prior knowledge that the long molecules within one partition cover a very small fraction of the genome (0. 1%). A group of reads mapping to distant duplicated regions have a 99. 9% prior probability of having originated from a single copy of the locus. Unique alignments from flanking regions and single unique nucleotides within the copies combine to inform the correct mapping location and mapping quality score for each group of linked reads. Our phasing algorithm uses a maximum likelihood approach to optimize the conditional probability of read and barcode support for each allele given a phasing configuration. The algorithm first finds near-optimal local haplotype configurations using a beam-search approach to phase blocks of ~50 variants. It then joins neighboring blocks together, and makes local corrections until convergence. For SV detection, our algorithm first searches for all pairs of genomic loci with significant barcode intersection, encoding this search as an efficient sparse matrix-multiplication. Candidates from this first stage are then filtered utilizing a probabilistic model that incorporates read-pair, split-read, and barcode data. We tested our phasing algorithm on data from the CEPH cell line NA12878 and the GiaB Ashkenazi trio. For WGS, we achieve mega-base N50 phase blocks, >95% of SNPs phased and long switch error rates <0. 05%, without relying on statistical phasing. For WES, we phased 95% genes <100Kb into single phase blocks. We tested our SV-calling on NA12878 and NA20847, calling multiple large-scale structural variants and phasing them with respect to adjacent phase blocks, showing consistency of phasing with inheritance patterns in the nuclear trio.

1629F

Geisinger GenomeFIRST™ and Targeted Family History Collection. *S. A. Martin, M. A. Giovanni, A. H. Buchanan, M. L. Barr, J. B. Leader, C. J. Seiler, J. R. Hill, M. F. Murray.* Geisinger, Danville, PA.

BACKGROUND:The Geisinger Health System (GHS) MyCodeTM Biorepository-Regeneron Genetics Center collaboration (DiscovEHR) is assembling a cohort of 250,000 participants who will undergo whole exome sequencing (WES). Through the GenomeFIRST™ return-of-results program, participants found to have a pathogenic or likely pathogenic variant in a gene associated with one of 27 medically actionable conditions (inclusive of the ACMG secondary findings list) will have the result CLIA-confirmed and clinically returned. This program includes a web-based, patient-entered family history tool with EHR compatibility, the GenomeFIRST™ Family History Tool. METHODS:The GenomeFIRST™ Family History Tool was developed utilizing response-based logic to collect only targeted family history information relevant to the returned result. After results are reviewed with participants, they are directed to a secure URL where they enter a unique identification number and are presented with a condition-specific family history survey. Through brief prompts and questions, the survey collects family structure information (including first and second degree relatives) and condition-specific family health information for each relative. The patient-entered data is used to generate a detailed report to enable clinical decision support. RESULTS:The family history tool was beta tested by 35 individuals who responded for all first and second degree relatives. Testers were surveyed regarding usability and total time for completion. The average completion time was: 8 minutes (range 4. 5 to 14). A total of over 42,000 patients have undergone WES to-date. It is anticipated that at least 840 patients (2%) will receive a report of a pathogenic/likley pathogenic variant. The first participants will receive results in July 2015. The tool will be used by these participants and their relatives who receive cascade testing for familial mutations (> 1,000 patients). CONCLUSIONS:The use of sequencing to screen populations for clinically actionable genomic findings will lead to the need for easy to use tools for patient-reported data entry, provider decision support, and interface with EHRs. The “GenomeFIRST™ Family History Tool” will be used to satisfy these needs. The utility of this approach for identifying both at-risk family members and affected family members is under investigation. Detailed results on the first 50 cases will be presented.

1630W

Massively Parallel Demultiplexing of Raw Illumina Sequencing Data. *B. R. Myers[1], S. M. L. Griffith[1], K. N. Hetrick[1], B. D. Craig[1], D. Mohr[1], K. Roberts[1], M. Z. Mawhinney[1], A. Robinson[1], D. R. Leary[2], J. D. Newcomer[3], L. Watkins, Jr. [1], K. F. Doheny[1].* 1) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 2) The New York Genome Center, New York, NY; 3) The Johns Hopkins Applied Physics Lab, Laurel, MD.

The Center for Inherited Disease Research (CIDR) provides high quality genotyping and sequencing services and statistical genetics consultation to investigators working to discover genes that contribute to disease. Before Illumina sequencing data can be analyzed, raw base call files (BCL) produced by the various Illumina sequencing technologies (HiSeq, MiSeq, etc.) are demultiplexed and converted into FASTQ files. While the demultiplexing tools provided by Illumina are typically used for this conversion process, the software development group at CIDR has designed and implemented a distributed BCL demultiplexing tool. Written in Java, the CIDR BCL conversion tool is integrated with CIDRSeqSuite, a software application designed to implement and run sequencing analysis pipelines in a distributed fashion across a computational cluster. Because of this integration, the BCL conversion tool takes advantage of the high degree of parallelism made available by the CIDRSeqSuite toolset. The data associated with each of the tiles of an Illumina flowcell are not directly related to one another, meaning that each tile can be processed independently. Because the set of tile-conversion tasks is embarrassingly parallel, the BCL conversion and demultiplexing process scales well with available computational resources. Previously, FASTQ files were generated in two general steps. Illumina's Offline Base Caller (OLB) was used to generate QSEQ files, and CIDRSeqSuite demultiplexed and converted them into the FASTQ format. Whereas the combined aforementioned methods could take many hours, the current approach takes anywhere from five to 30 minutes, depending upon data size and the number of concurrent tasks. This increase in speed is particularly germane for high-throughput runs that are available on the current generation of HiSeq sequencers. The tool works for single- and paired-end runs; and supports both currently used location file formats (. clocs and . locs). In addition to the benefits mentioned above, the CIDR BCL conversion tool also writes detailed reports on the sample index distribution discovered while demultiplexing. This allows users to quickly determine incorrect or underrepresented indices present in a given run. The tool is also capable of removing tiles that have corrupt or truncated associated BCL files. Current work includes enabling dual-index demultiplexing.

1631T

CANNOTATE: A Genomic Sequencing and Annotation Database. *J. D. Newcomer[1], S. M. L. Griffith[2], E. W. Pugh[2], D. R. Leary[3], R. T. Hider, Jr. [1], P. Zhang[2], D. Mohr[2], L. Watkins, Jr. [2], C. Bradburne[1], K. F. Doheny[2].* 1) Applied Physics Lab (APL), The Johns Hopkins University Applied Physics Laboratory, Laurel, MD; 2) Center For Inherited Disease Research (CIDR), The Johns Hopkins University School Of Medicine, Baltimore, MD; 3) The New York Genome Center, 101 Avenue of the Americas, New York, NY 10013.

The Center for Inherited Disease Research (CIDR) provides high quality genotyping and sequencing services and statistical genetics consultation to investigators working to discover genes that contribute to disease. The Johns Hopkins University Applied Physics Lab (APL) supports independent research and development programs that pioneer and explore emerging technologies and concepts to address future national priorities. CIDR and APL staff have collaborated to produce CANNOTATE, a software tool designed to store and analyze large-scale genomic data. CANNOTATE was developed because CIDR staff encountered the following two problems:1. Genomics data is typically stored in flat VCF files, which leads to slow data access and analysis2. Some current genomics annotation tools are extremely resource (RAM) and time intensive when running many samples over large data sets. CANNOTATE is designed to solve these problems by organizing variant (from VCF) and annotation data in a NoSQL database using Accumulo, a NoSQL technology built on top of the Hadoop File System (HDFS). Because Accumulo is based on HDFS, the data are spread across a distributed computing cluster (Hadoop HDFS), which leads to faster data access and storage. This software stack organizes, compresses, and indexes the variant and annotation data for fast query and retrieval. CANNOTATE also includes software tools that perform fast and efficient sample annotation. CANNOTATE is a work in progress; our prototype has yielded very promising results. Data is ingested two to three orders of magnitude faster than a previous RDBMS implementation allowed. Annotations complete 26 times faster than comparable tools: a 51-sample annotation went from 4. 5 hours down to 10 minutes. Current work includes a GUI toolset and additional analysis capabilities. Eventually, CANNOTATE will be released to the genomics community as an open-source tool.

1632F

Integrated Genome Mapping in Nanochannel Arrays and Sequencing for Better Human Genome Assembly and Structural Variation Detection. *A. W. C. Pang, A. Hastie, P. Sheth, T. Anantharaman, Z. Džakula, H. Cao.* BioNano Genomics, Inc. , San Diego , CA.

De novo genome assemblies using purely short sequence reads are generally fragmented due to complexities such as repeats found in most genomes. These characteristics can hinder short-read assemblies and alignments, and that can limit our ability to study genomes. The BioNano Genomics Irys® System linearizes long DNA molecules, thus yielding single-molecules containing long-range information. These hundreds of kilobases molecules can capture structural information that may be missed by other sequencing platforms. The assembled genome maps from these molecules can scaffold sequencing assemblies to validate the accuracy of the sequences, and to anchor the adjacent sequences into the proper order and orientation. The long-range hybrid scaffolds are able to identify novel chromosomal rearrangements undetectable by short-read alignment or reference-guided assembly approaches. We present a comprehensive analysis of a human genome by combining single molecule genome mapping with one of the most annotated sequence assemblies, the HuRef (Human Reference Genome – J. Craig Venter Institute) assembly. Overall, we found that assemblies from the sequencing and genome mapping technologies correspond well, and the resulting hybrid scaffolds are highly contiguous, with a N50 exceeding 35 Mb, a value typically unachievable by short-read sequencing technologies alone. In addition, we compared the structural variation with calls previously detected in the HuRef assembly, and found multiple novel variants spanning over hundreds of kilobases in size. Some of these variants reside in areas where the sequence assembly was poorly covered or was highly fragmented; yet these variants encompass numerous genes, and can be of functional importance. Finally, we identified genome maps that span over the remaining reference gaps, and maps that resolve and measure long tandem repeats.

1633W

An accurate and sensitive approach to detecting source of variance in DNA methylation Beta values. *Y. Park[1,2], L. B. Chibnik[2,3,4,5], M. Kellis[1,2], P. L. De Jager[2,3,4].* 1) Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, Massachusetts; 2) Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts; 3) Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Departments of Neurology and Psychiatry, Brigham and Women's Hospital, Boston, Massachusetts; 4) Department of Neurology, Harvard Medical School, Boston, Massachusetts; 5) Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts.

From DNA methylation we keep track of cumulative epigenetic changes in cells, and we can utilize DNA methylation as a good indicator of diagnosis and prognosis. In case-control studies, differential DNA methylation levels can occur due to abnormal acceleration of enrichment or depletion. However, methylation is result of disease status and other factors such as aging and gender. Unless we factor out unwanted effects from DNA methylation, downstream analysis is affected by spurious factors will dilute valuable information of subtle, yet statistically significant, disease associations. This is true of the Beta values provided by the Illumina Humanmethylation450 beadset. Nonetheless common practice of modeling heavily rely on a simple linear regression, which assumes Gaussian noise distribution. We propose a sparse regression model that builds on the Beta regression. We extend the original regression model with spike-slab prior (Mitchell and Beauchamp, 1988), and computed posterior inclusion probability of known and latent variables. The spike-slab prior has been adopted to variable selection problems and showed success in modeling Gaussian variables (e. g. , Carbonetto and Stephens, 2013). However the Beta probability distribution is non-conjugate to the spike-slab prior, we performed hybrid algorithm using elliptical slice sampling (Murray et al. , 2010) and expectation propagation (Hernandez-Lobato et al. 2014). We regressed DNA methylation separately in the Brain and CD4+ cells on Alzheimer's Disease (AD) phenotypes (beta amyloid and tau tangles), cognitive decline score, and other confounding factors such as age, gender, and experimental batches. First we evaluated the utility of cognitive decline score as a non-invasive proxy to AD pathology. We found strong agreement between the cognitive score association and AD pathology. We then tested a clinical value of CD4+ methylation searching for a convergence point of the Brain and CD4+ cells. We confirmed convergence points at pathway level. Examples includes WNT signaling, notch signaling, gap juction, hypoxia, and SMAD2 pathways. Many of them are supported by current knowledge of metabolism and neurodegenerative disease mechanism. Moreover we extended our regression method to hierarchical models that take into account genetic bias and downstream gene expression regulation. Finally we evaluated statistical power of our methods in extensive simulations. We will release source codes in a public repository.

1634T

Discovering Combinatorial Patterns of histon modification regulating gene expression. *S. Park.* Department of Bioinformatics and Life SCience, Soongsil University, Seoul, South Korea.

To dissect histone code hypothesis, it is challenged that computational methods can be capable to discover de novo combinatorial patterns of chromatin modifications across cell types. We report a new computational approach based on association rule mining for de novo combinatorial pattern discovery of chromatin modifications. We identify the patterns correlated to gene expressions and enriched in functional categories. The method, an unsupervised approach to discover combinatorial chromatin modification patterns, can identify globally occurring chromatin modifications to locally rare modifications. It has advantages over existing methods since the patterns as rules are a quantitative description of properties and their interpretation is straightforward and simple to biologists.

1635F

GenomeWhiff and DigitalKindred: Ultra-Fast Similarity Search, Clustering and Classification of Very Large Number of NGS Samples Using Bitwise Operations. *J. Patel, A. Abhyankar.* New York Genome Center, New York, NY.

GenomeWhiff is a marker-based super-fast sample concordance and similarity search. DigitalKindred extends the ideas of bitwise representation and fast similarity search to biological applications like GWAS/case-control, clustering/classification of samples and single cell genomics. **GenomeWhiff**GenomeWhiff was developed as the next step of concordance check, wherein a sample is cheaply genotyped using a chip to be compared with the sequencing data for that sample to check for sample swaps and contamination. GenomeWhiff maintains a compact index of all the previously processed samples and checks the new sample against the index efficiently. A sample can be checked against 26,000 samples in 200 milliseconds. This is a similarity search that exposes inherent relationships like parent-child and siblings as well as indicates sample contamination. GenomeWhiff uses a set of LD-pruned, biallelic markers that have high minor allele frequency. Each marker is represented as a pair of bits encoding 2 alleles. Specific combination of allele bit values creates a sample signature. Using series of bitwise operations, a ranked list of samples similar to the query sample is created. Each sample signature is accompanied by a bitmask that indicates valid markers for the sample. The bitmask also allows partial searches involving a subset of total marker set. In case of contamination, a candidate list for the contaminating sample is created by looking at discordant calls and by searching for samples that would explain the allele change at the discordant calls. **DigitalKindred**For biological applications, the above ideas are generalized by separating the encoding scheme from the fast bitwise similarity search mechanism. Each application has its own encoder and decoder that describe how the data is represented as bitsets. The markers need not be positions on the genome. For a DNA sample set, the set of all the variants seen in the sample set is considered and a sample signature encodes whether each of those variants is present in the samples. For an RNA sample set, individual exonic expression levels are encoded after binning. Irrespective of encoding, systematic and exhaustive enumeration over marker combinations reveals the combinations that would cluster or classify the samples in line with observed phenotypes. Due to compact and efficient bitwise representation, this approach can scale to millions of samples and millions of markers.

1636W

Software for the visualisation of genetic variation and its implication on protein sequences and 3D structures. *A. Prlic[1], R. Bhattacharya[2], S. K. Burley[3,1], H. Tilgner[4], P. W. Rose[1].* 1) RCSB Protein Data Bank, San Diego Supercomputer Center, University of California San Diego, San Diego, CA; 2) Bioinformatics and Medical Informatics, San Diego State University, San Diego, CA; 3) RCSB Protein Data Bank, Center for Integrative Proteomics Research and Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, New Brunswick, NJ; 4) Stanford University, Stanford, CA.

Statement of purpose: Knowledge of an atomic-level resolution structure of a protein, provides deeper insight into the molecular mechanisms that underpin the function of a gene/protein, as well as their role in disease. To enable a deeper analysis of the possible changes related to genetic variation, we have developed novel tools that facilitate mapping of any genetic location onto corresponding protein sequence isoforms and 3D protein structure available through the RCSB Protein Data Bank. Our tools permit annotation of large sets of genetic variation data by correlating genomic location and sequence variation data with over 60 external resources and protein 3D structural data housed within the data-warehouse that supports the RCSB PDB website. We are apply these tools in several collaborations, wherein we annotate genetic variation data obtained through Next-generation sequencing. We have also developed novel visualisation tools that provide a graphical depiction of the impact of genetic variation on protein sequence and 3D structure. **Methods used**:We developed a mapping for every nucleotide in a known human gene (as available from the Human Gene Nomenclature Committee) at a specific position in the reference genome to mRNA and codon position together with the corresponding amino acid sequence position and atomic coordinates. A mapping pipeline is available for manual analysis of single genomic positions via a user interface. It can also be scripted for batch-analysis of large number of genetic locations (e. g. ,for SNP analysis). It forms the core of new user interfaces at the RCSB PDB web site, including Gene View (e. g. http://www. rcsb. org/pdb/gene/BRCA1) and Protein Feature View (e. g http://www. rcsb. org/pdb/protein/P38398). **Summary of Results:**Our database contains annotations for about 19,000 human genes. Using our analysis tools, at http://www. rcsb. org/pdb/chromosome. do, any genomic location can be manually mapped to UniProt and the PDB archive of 3D structures (if a mapping is possible). At present, ~5100 distinct human genes can be mapped onto ~7900 PDB entries. The same pipeline is also being used to analyze the impact of experimentally verified (synthetic long-read mRNA sequencing) alternative pre-mRNA splicing events on 3D protein structures, where we can associate alternatively spliced exons with protein features.

1637T

Improved Transcription Factor Binding Site Prediction using DNase-seq Footprinting in a Supervised Discriminatory Machine Learning Framework. *B. C. Quach[1,3], T. S. Furey[1,2].* 1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Curriculum in Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Transcription factors (TFs) bind to sequence motifs and regulate gene expression through interactions with other components of the transcriptional machinery. Identifying TF binding sites (TFBSs) remains a long-standing challenge due to TFs tolerating motif degeneracy, and only a small fraction of motif sites across the genome exhibit binding activity in a given cell-type. DNase-seq assays accessible chromatin corresponding to active regulatory regions, represented as reads enriched for nucleosome-depleted open chromatin. TFs bind at smaller sites within the broader open chromatin and protect the TF-bound DNA from DNaseI digestion. This creates a local depletion of reads within open chromatin known as footprints, which can be utilized to locate active TFBSs. Existing footprint detection methods differentiate between TF bound and unbound sites using only a single generative model for each binding state. With DNaseI digestion data from ENCODE DNase-seq experiments, we show that this assumption fails to fully capture footprint signal heterogeneity. To address this weakness, we developed a support vector machine (SVM) classification framework called DeFCoM (Detecting Footprints Containing Motifs) to better model DNase-seq footprint diversity. DeFCoM starts with a training set of ChIP-seq defined bound and unbound sites that contain a motif for a factor of interest. DNaseI cut patterns for these sites are clustered within both the bound and unbound groups, where clusters denote subclasses or profile patterns. DeFCoM models each of these diverse digestion profiles based on features of the digestion pattern, then uses these feature vectors to train an SVM to distinguish profile patterns characteristic of active and inactive TFBSs. To assess the performance of DeFCoM, we compared it to 5 existing footprinting methods. The comparative analysis included 18 TFs and DNase-seq data from 6 ENCODE cell-lines. At low false positive rates (<0. 05), we find DeFCoM is consistently better or at least comparable in accuracy toexisting methods across these diverse factors, demonstrating its ability to correctly identify true binding sites. Our results suggest that modeling DNaseI cut profiles using DeFCoM enhances active TFBSs prediction accuracy relative to current footprint detection methods.

1638F

Authorial: An author list management system for scientific papers. *N. W. Rayner[1,2,3], N. R. Robertson[1,2], M. I. McCarthy[1,2,4].* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom; 2) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, United Kingdom; 3) Wellcome Sanger Institute, Hinxton, Cambridge, United Kingdom; 4) Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, United Kingdom.

The continued decrease in the cost of genotyping has led to an increase in the size and number of data sets genotyped. This abundance of cohorts with, usually, multiple phenotypes has led to a proliferation of data sharing via disease specific consortia. These large collaborative efforts have resulted in a greatly increased number of authors on publications and managing the long lists of authors, author contributions and institutions by hand is error prone and extremely time consuming. To address this issue we have developed a web application, Authorial, designed to collect authorship details for a publication and create an ordered, annotated author list in Microsoft Word format. Authorial is a server-side application written in Java and runs in all major internet browsers including Firefox, Chrome, Safari, Internet Explorer and Opera. The system operates in a hierarchical manner and there are a number of roles defined in the application, these are; publication organiser, author group leader and author. The publication organiser assumes overall responsibility for the publication and defines a series of author groupings and an author group leader for each. The author group leaders are emailed directly by the application and are asked to confirm their own details and to populate their assigned author groups with participating authors. Once the membership of all author groups has been completed, individual authors are directly emailed a unique access link URL which allows them to check their details and the details of their affiliated institutions. To avoid issues with multiple institutional spellings the system caches institutions entered to date so that these can be selected via autocomplete and reused by subsequent authors. Within Authorial there are various reports to allow the organiser to review and sift out duplications and inconsistences in the collected data. Authorial tracks the progress of each author so the organiser is able to re-email and chase up non-responders. Once the information is complete a Word Document is created containing the ordered author list. Should the author order change or an author or institution be added or removed, unlike traditional methods, it is quick to create the new Word format author list. To date the application has been used successfully by a number of different publication leaders representing multiple consortia, such as GoT2D which consists of over 250 authors and 170 institutions.

1639W

Best practices for deploying a high throughput pipeline for variant calling in an extremely large cohort of whole genome sequenced samples. *N. Rustagi[1], Z. Huang[1], N. Veeraraghavan[1], A. Carroll[2], M. Gorentla Venkata[3], E. Boerwinkle[1,4], R. Gibbs[1], F. Yu[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) DNAnexus, Palo Alto, CA; 3) Oak Ridge National Laboratory, Oak Ridge, TN; 4) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX.

Large cohort studies have the potential to inform gene discoveries and directly influence personalized medicine. The surge of large cohort studies aided by decreasing costs of genome sequencing are driving a need for variant calling pipelines with the ability to handle ultra-large-scale computes and a short turn around time. Even with low coverage(~6x-10x) whole genome sequencing (WGS), the amount of genomic data processing for cohorts with >5000 samples is beyond the capacity of commodity compute clusters, typical of most research environments. Joint calling on cohorts with over 200 Tb of aligned read data imposes constraints on the length of the atomic genomic compute window. Minimizing data and memory footprint, transfer latency and costs, IO bottlenecks, and file system inconsistencies also becomes challenging at this scale. In this work, we present a high throughput framework for SNP and INDEL calling, which leverages heterogeneous computing infrastructures and is easily scalable to accomplish cohorts with over 10k samples. As a proof of principle, we present results of analysis on Cohorts for Heart And Aging Research in Genomic Epidemiology(CHARGE) WGS freeze 3 dataset in which joint calling, imputation and phasing on over 5200 whole genome samples was produced under six weeks using four state of the art callers, namely SNPTools, GATK-HC, GATK-UG and GotCloud. We used Amazon AWS provided by DNAnexus, a 4000-core in-house cluster at HGSC, IBM power PC Blue Biou at Rice and Rhea at Oakridge National Labs. We had approximately 180 TB of raw aligned read data, which was split into a genomic window size of 1MB across all the BAMs for the variant calling phase. The total number of compute core hours used were 5. 2 million. The SNP and INDEL site calling was accomplished on Amazon AWS architecture, with 1000 concurrent runs on average, with an average schedule time of approximately 60hrs. To optimize transfer costs and file system stress, four callers were executed on the 1MB genomic region at one go. Approximately 75% of the imputation was accomplished on Rhea, and the rest at Blue Biou, with 64,000 and 1000 concurrent runs respectively. Average time to completion was 30 hrs and 13 hrs for each window on Rhea and Blue Biou respectively. Our data footprint never exceeded three times the data through the whole operation. We hope that our experience with this compute can serve as a roadmap for future large scale variant calling efforts. .

1640T

VarAFT: A variant annotation and filtration system for exome sequencing data. *D. Salgado[1], J. P. DESVIGNES[1], M. BARTOLI[1], V. DELAGUE[1], M. KRAHN[1,2], M. MILTGEN[1], C. BEROUD[1,2].* 1) Aix-Marseille Université, Inserm, GMGF UMR_S 910, 13385, Marseille, France; 2) APHM, Hôpital TIMONE Enfants , Laboratoire de Génétique Moléculaire, 13385, Marseille, France.

High-throughput DNA sequencing technologies provide an unprecedented opportunity to identify disease-causing mutations for Mendelian disorders. However even if primary analysis of WES raw data such as alignment, variant calling and annotationare almost standardized and usually performed by bioinformaticians, the filtration step aiming at reducing the number of variationsto a handfulof candidate disease-causing mutationsremains a challenge. To facilitate this selection process an ideal system should allow: the combination of data from multiple individuals in order to evaluate autosomal recessive, dominant, X-linked and *de novo* hypothesis; the collection of extensive annotations for each variant; and the selection of the most relevant mutations based on multiple selection criteria. Additional features such as the coverage at the gene/exon/nucleotide levels are critical for clinical applications to ensure proper gene exclusion. As such tool is not currently publically available, we designed VarAFT (**Var**iant **A**nnotation and **F**iltration **T**ool - http://varaft.eu) as a freely available systemrunning on MacOS, Windows and Linux operating systems. It has been tailored with the recommendations of researchers, geneticists and clinicians from the Department of molecular diagnosis of the La Timone Children Hospital and the INSERM UMR_S910 research unit, taking into account international recommendations from ACMG and ESHG. VarAFT offers three application modules:1) The “CoverageAnalysis tool”is particularly useful for clinical applications as it allows evaluation of the coverage and depth for any nucleotide, exon or transcript;2)The “Annotation Module” can be used to perform variant annotation from any VCF file. It is a mandatory step to further apply filtration criteria to highlight potential disease-causing mutations. VarAFT makes use of the ANNOVAR annotation system in combination to theUMD-Predictor for pathogenicity prediction of exonic SNVs and the Human Splicing Finder system to evaluate the impact of any exonic or intronic mutation on splicing signals;3)The “Analysis and Filtration module”provides various filtration options to accommodate the international recommendations such as:localization, impact on the protein, pathogenicity, quality, zygosity, and frequency. . . VarAFT is currently designed for WES analysis. It can also handle gene panels. Optimizations are ongoing for WGS analysis.

1641F

cnvScan: a CNV screening and annotation tool to improve the clinical utility of computational CNV prediction. *P. S. Samarakoon[1,2], H. S. Sorte[1,2], A. Stray-Pedersen[3,4,5], O. K. Rødningen[2], T. Rognes[6], R. Lyle[1,2].* 1) Department of Medical Genetics, University of Oslo, Oslo, Norway; 2) Department of Medical Genetics, Oslo University Hospital, Oslo, Norway; 3) The National Neonatal Screening Unit, Oslo University Hospital, Oslo, Norway; 4) Center for Human Immunobiology/Section of Immunology, Allergy, and Rheumatology, Texas Children’s Hospital, Houston, TX, USA; 5) Baylor-Hopkins Center for Mendelian Genomics of the Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 6) Department of Informatics, Faculty of Mathematics and Natural Sciences, University of Oslo, Oslo, Norway.

With advances in next generation sequencing technology and analysis methods, single nucleotide variants (SNVs) and indels can be detected with high sensitivity and specificity in exome sequencing data. While recent studies have demonstrated the ability to detect disease-causing copy number variants (CNVs), exonic CNV prediction programs have shown high false positive (FP) CNV counts. The high false discovery rate is the major limiting factor for the applicability of these programs in clinical studies. Thus, we developed a tool (cnvScan) to improve the clinical utility of computational CNV prediction by reducing the false discovery rate. cnvScan consists of two stages: CNV screening and CNV annotation. The screening stage utilizes CNV quality scores reported from prediction programs and further evaluates the quality of CNV calls by using an in-house CNV database. The annotation stage is designed to annotate predicted CNVs with functionally and clinically significant information using different source datasets. In addition, known CNVs within predicted variants were annotated using publicly available CNV datasets. CNV quality scores and annotated information reported from cnvScan are then used to filter high-quality rare CNVs from computational CNV predictions. To assess the utility of our method, we tested cnvScan using 64 exomes from Primary Immunodeficiency (PID) patients. Here we demonstrate that cnvScan can be used to generate a significantly small set of high-quality rare CNVs from a relatively large set of computationally predicted CNVs (4 high-quality rare CNVs in 4 patients from 17147 CNVs predicted from ExCopyDepth). Three CNVs out of these four high-quality rare CNVs were confirmed as PID-causing CNVs. We further compared CNV predictions from five programs (ExomeCopy, ExCopyDepth, ExomeDepth, CoNIFER and XHMM) and demonstrated the ability to derive high-quality rare CNVs using cnvScan. Finally, we assessed the efficiency of cnvScan using different filtration criteria and showed the ability to reduce false discovery rate by filtering-out FP CNVs while retaining rare true positive variants. In summary, cnvScan improves the clinical utility of computational CNV prediction by reducing the false discovery rate and providing annotation.

1642W

Benchmarking Strand NGS RNA aligner against TopHat2 and other common splice aligners. *A. Sathyanarayanan, R. Gupta.* Strand Life Sciences, Bangalore, India.

Background and objective: RNA alignment involves mapping of short reads produced from next generation sequencing of cDNA obtained from RNA, to a transcript annotation or reference genome. Alignment of millions of short reads with high accuracy is one of the biggest challenges of RNA Seq analysis. It is pivotal for discovering splicing events, differential gene expression, gene fusion and RNA editing. In this study, we compared the performance of the RNA mapping algorithm implemented in Strand NGS with TopHat v2 and several other algorithms. Strand NGS RNA aligner is a split read aligner with three options: 1) Alignment against transcriptome with no novel splices detection, 2) Alignment against transcriptome with novel splices detection and 3) Alignment against transcriptome and genome with novel splices detection. **Results:** We used four samples for the study, two samples each of mouse and human. Human samples comprise of RNA-Seq libraries from K562 cell line and the mouse data comprises of RNA-seq data obtained from brain tissue from adult mice of C57BL/6NJ strain. Alignment was performed in both, Strand NGS and TopHat v2, using similar parameters. Assessment of the alignment results was done using several benchmarks such as alignment yield, mismatch detection, indel placement and splice detection. While similar results were obtained for alignment yield and mismatch frequency, splicing events were detected more effectively in Strand NGS than TopHat v2. The number of known and novel junctions was plotted for increasing supporting reads cut-off to obtain receiver operator curves (ROC). For all the datasets, the ROC curve showed that Strand NGS outperforms TopHat v2 in identifying similar number of annotated junctions for much fewer false/novel junctions. Based on metrics of alignment yield and mismatch frequency, Strand NGS results are comparable to GSNAP, GSTRUCT, MapSplice and STAR, and better than GEM, ReadsMap and PALMapper. **Conclusions:** Several algorithms have been developed to overcome the challenges posed in RNA Seq analysis. High alignment accuracy, right placement of indels and gaps, and efficient detection of splice junctions are some desired features of RNA spliced read aligner. Detailed assessment shows that Strand NGS compares well with other state-of-the-art aligners for RNA alignment and also offers comprehensive approaches to data visualization, filtering and downstream analysis such as quantification and differential analysis.

1643T

Accurate somatic mutation detection at low allelic frequency and low copy number applied to liquid biopsy samples. *M. Schwartz[1], C. Ionescu-Zanetti[1], N. Kamps-Hughes[1], R. Brobey[2], K. Rosenblatt[3], M. Dehghani[3], R. Amato[2].* 1) Research and Development, Fluxion Biosciences, South San Francisco, Ca; 2) University of Texas Health Science Center at Houston, Houston, TX; 3) Research and Development, CompanionDx, Houston, TX.

Purpose: Tumor genotyping experiments are increasingly moving from the characterization of solid primary tumor biopsies to blood samples to track dynamic mutational changes over time. These sample types require tumor DNA enrichment steps and modifications of the molecular analysis protocols be able to detect rare variants. For circulating tumor cell (CTC) samples, the processing must also accommodate low copy input as sometime only a few cells (<5) are recovered from a blood draw. Here, we present validation of a method to detect somatic mutations from a blood draw, where typical CTC enrichment above 10% of total cell numbers enables the use of standard amplicon libraries. **Methods:** Analytical samples were created by spiking tumor cell lines into whole blood from healthy donors, followed by CTC enrichment using immunomagnetic beads in a microfluidic chamber. Further reference samples were created by spiking tumor cells into known numbers of donor-derived white blood cells (WBCs). CTCs were lysed and DNA was amplified by whole genome amplification (WGA). Targeted libraries were created using different amplicon sets, then sequenced using either the PGM (ThermoFisher) or MiSeq (Illumina) instruments. Data was analyzed using a customized variant pipeline based on standard alignment tools, variant filtering, and functional interpretation. Clinical samples were processed using the same protocols. Matched clinical samples were enumerated to determine the CTC load (CK+, CD45-, nucleated cells). All data was analyzed in a blinded manner. **Results:**Multisite analytical validation data demonstrated a detection limit down to 10 cells from a blood draw with a very low false positive rate. This translates into a robust detection of variants down to 0. 5% mutant allele frequency, without the use of nonstandard NGS techniques like single molecule sequencing. Results are shown for a number of analytical validation experiments, detecting variants down to 2 mutated copies in a background of wild type cells. Exemplary clinical data is presented from 10 patient sample sets, including biological replicates and analysis replicates using different amplicon sets and sequencing technologies to validate the results. **Conclusions:** A sensitive assay for the detection of somatic variants from a blood draw using standard amplicon panels developed for somatic mutation detection from solid cancers has been validated using analytical and clinical samples.

1644F

COSMOS: accurate detection of complex somatic structural variations through asymmetric comparison between tumor and normal samples. *J. Sese[1], K. Yamagata[1], A. Yamanishi[2], C. Kokubu[2], J. Takeda[2].* 1) BRD, AIST, Tokyo, Japan; 2) Grad. School of Medicine, Osaka U. , Osaka, Japan.

An important challenge in cancer genomics is precise detection of structural variations (SVs) by high-throughput short-read sequencing, which is hampered by the high false discovery rates of existing analysis tools. Here we propose an accurate SV detection method named COSMOS, which compares the statistics of the mapped read pairs in tumor samples with isogenic normal control samples in a distinct asymmetric manner. COSMOS also prioritizes the candidate SVs using strand-specific read-depth information. Performance tests on modeled tumor genomes revealed that COSMOS outperformed existing methods in terms of F-measure. We also applied COSMOS to an experimental mouse cell-based model, in which SVs were induced by genome engineering and gamma-ray irradiation, followed by polymerase chain reaction-based confirmation. The precision of COSMOS was 84. 5 %, while the next best existing method was 70. 4%. Moreover, the sensitivity of COSMOS was the highest level, indicating that COSMOS has great potential for cancer genome analysis. COSMOS is available from http://seselab. org/cosmos/.

1645W

A tool for developing and validating supervised learning risk prediction models. *S. Shankaracharya, Y. Ye, H. Hu, X. Wu, CD. Huff.* Department of Epidemiology, University of Texas MD Anderson Cancer Center, Houston, TX.

Machine learning (ML) methods have the potential to substantially improve the accuracy of disease risk prediction, but the technical complexities of these methods have been a barrier to widespread adoption. Here we present the PREdiction by SUpervised Learning Toolkit (PRESULT), a toolkit designed to simplify the development, validation, and dissemination of supervised machine learning risk prediction models. PRESULT supports four supervised ML methods: Mixture of Experts, Support Vector Machine, Random Forest (RF), and Regression Tree. PRESULT generates receiver-operating characteristics (ROC) curves and calculates the area under the curve (AUC) for training and validation datasets, comparing each ML model with a traditional logistic regression model. PRESULT produces a compact, distributable, and machine-readable description of each training model, ensuring reproducibility and enabling external validation. The toolkit also performs k-fold cross validation and provides robust support for parameter tuning and optimization for each ML method. PRESULT calculates 5-year and 10-year absolute risk by incorporating overall incidence and mortality rates with relative risk, which we estimate by fitting a logistic regression model to the continuous prediction score for each ML method. We evaluate the performance of this toolkit in two common, complex disease datasets: a case-control study of type 2 diabetes in Pima Indians with epidemiological data and a case-control study of lung cancer in European Americans with epidemiological and GWAS data. We found that RF consistently outperformed logistic regression and other ML methods in these particular datasets, with an average 10-fold cross validation AUC of 75. 6 in lung cancer and 84. 3 in type 2 diabetes compared to 69. 2 and 72. 5, respectively, for logistic regression models. PRESULT mitigates many of the major challenges of ML risk prediction modeling, providing new opportunities for more accurate assessment of personalized disease risk.

1646T

QuicK-mer: A rapid paralog sensitive CNV detection pipeline. *F. Shen, J. Kidd.* Human Genetics, University of Michigan, Ann Arbor, MI.

Copy number variation (CNV) is an important class of variation that contributes to genome evolution and disease. CNVs that become fixed in a species give rise to segmental duplications; whereas already duplicated sequence is prone to subsequent gain and loss leading to additional copy-number variation. Multiple methods exist for defining CNV based on high-throughput sequencing data, including analysis of mapped read-depth. However, accurately assessing CNV can be computationally costly and multi-mapping based approaches may not specifically distinguish among paralogs or gene families. We describe a rapid, mapping free, paralog-specific pipeline for CNV detection from second generation sequencing datasets by counting the occurrence of predefined 30-mers. The pipeline achieves paralog specificity while maintaining an efficiency of less than 7 hours for analysis of a typical 10x mammalian genome. We have utilized the QuicK-mer pipeline to explore variation among duplicated sequences in human, chimpanzee, and mouse sequencing data. Preliminary results demonstrate concordance between CNV prediction and experimentally validated copy number across mouse strains C57BL/6J, 129S5/SvEvBrd and BALB/cJ for KRAB-ZFP genes associated with sex-specific phenotypes in the *Rsl1/2* locus. Data further reveals significant overlaps between CNV hot spots and the existing segmental duplications in human and mouse genomes.

1647F

GENALICE MAP: efficient and accurate population genomics. *B. Tolhuis, L. Baarspul, H. Karten.* GENALICE BV, Harderwijk, GLD, Netherlands.

Current DNA sequencing platforms can analyze tens of thousand individual genomes per year, enabling large scale population genetics. Indeed, several nation-wide genomics initiatives are emerging, which aim at deciphering hundreds of thousands human genomes. Detection of sequence variants is ideally performed in the context of the entire cohort, begging that thousands of samples be analyzed in parallel. This is a computational and bandwidth challenge due to the huge amount of data that needs to be processed.

Here, we present a novel population scale variant caller as an extension of our proprietary GENALICE MAP NGS secondary data analysis suite. Earlier, we showed that GENALICE MAP aligns sequence reads for a single human genome (37x coverage) in 25 minutes and subsequent variant detection requires 5 minutes. The population caller investigates mapped reads from multiple genomes in parallel while each individual genome benefits from the population context. In a benchmark test, analyzing 17 individuals from Illumina's Platinum Genomes, we observed that variant detection in population context takes 7 minutes per genome using a single compute node with a Dual Intel Xeon E5 2620 V2 CPU architecture. This is a time reduction of approximately two orders of magnitude compared to GATK's joint variant calling method run on the same data and hardware. Moreover, our population caller has a single step workflow, employed directly on aligned reads without any additional and pre- and post-processing. Finally, compared to single sample variant calling the population caller yields substantially more and accurate genotype information. For example, the number of false negative SNPs in a 'truth' set comparison is reduced with 30%.

GENALICE MAP's population caller significantly reduces computational and workflow requirements of multi-genome variant detection. It allows individual genomes to take advantage of the population context resulting in more genetic information with increased accuracy per genome. Currently, the GENALICE MAP population caller works on a single compute node and its runtimes scale linearly. In the future, it will run in parallel on multiple compute nodes, enabling the analysis of thousands of genomes using compute grids or cloud environments. Given its low computational requirements per genome, the GENALICE MAP population caller is a highly time and cost effective manner to detect sequence variants in large cohorts.

1648W

Applying compressed sensing to genome-wide analysis of qualitative traits. *S. Vattikuti[1], J. J. Lee[2], C. C. Chang[3], S. D. H. Hsu[4], C. C. Chow[1].* 1) MBS, LBM, NIDDK, NIH, BETHESDA, MD; 2) Department of Psychology, University of Minnesota, Minneapolis, MN; 3) Complete Genomics, Inc. , Mountain View, CA; 4) Research and Graduate Studies, Michigan State University, East Lansing, MI.

Genome-wide analysis is an underdetermined problem due to many fewer samples than measures. This poses a challenge for discovering genetic predictors and causal variants. Compressed sensing (CS) is the theory for selection that explains the properties of this problem for well-conditioned sensing matrices. As we have shown before for quantitative traits, the genome is such a matrix and there is a phase transition for complete selection of true associated variants [1]. For narrow-sense heritability of 0. 5 this is expected to occur around 30 times the number of trait-associated variants. Here we extend our prior work to qualitative traits and ask whether the same predictions can be made for these discrete traits. We explore factors such as case prevalence, sampling bias, and regression model (e. g. , linear, logistic). 1. Vattikuti S, Lee JJ, Chang CC, Hsu SD, Chow CC. Applying compressed sensing togenome-wide association studies. Gigascience. 2014 Jun 16;3:10. doi:10. 1186/2047-217X-3-10. eCollection 2014. PubMed PMID: 25002967; PubMed Central,PMCID: PMC4078394.

1649T

Computational Framework for Heterogeneity Assessment and Characterization in Single Cell Sequencing Data. *K. Volyanskyy[1], Y. Mao[1], T. Baslan[2], N. Dimitrova[1].* 1) Philips, Briarcliff Manor, NY; 2) Memorial Sloan Kettering Cancer Center, New York, NY.

Background: Heterogeneity is a common feature in both normal and diseased tissues. This phenomenon manifests itself in multiple modalities of genetic and epigenetic biological cell data such as gene expression, copy number variations, methylation, peptide profiling and others. For various purposes it is important to accurately identify and characterize distinct sub-populations of cells constituting heterogeneity profile of a sample. This becomes especially relevant with the adoption and rapid development of single cell sequencing technologies allowing previously unseen resolution of genomic data. In our work we developed an efficient computational framework to accurately identify heterogeneity profile from single cell sequencing data and analyze multiple modalities of genomic data. *Method*: We developed a computational framework based on optimal partition selection characterizing distinct sub-populations in a cohort of samples from single cells sequencing. The framework identifies optimal minimum variance partitions of the input in various modalities such as, RNASeq or copy number variation, or combined data set consisting of data from several modalities at once. In addition, we developed phylogeny reconstruction algorithm allowing identification of sub-populations in iterative procedure of minimum distance search. We tested our algorithm on a publicly available single cell transcriptome data set from [1] consisting of 82 single-cell transcriptomes from mouse lung epithelium obtained via microfluidic single-cell RNA sequencing. *Results*: We successfully identified distinct subpopulations in considered sets of samples. Compared to the ground truth published in [1] where transcriptomes were analyzed using unbiased genome-wide approach, we independently generated analogous partitions of the transcriptome into four main distinct sub-populations. *Summary:* Developed algorithms allow splitting heterogeneous data into a set of distinct sub-populations with guarantees of minimum within sub-population variability. The algorithm flow is contextually interpretable and the performance has been successfully tested on several data sets. The developed framework can find immediate applications in genomic studies that map out tumor heterogeneity at a deeper sequencing level. References[1] Treutlein B, Brownfield DG, Wu AR, Neff NF et al. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature 2014 May 15;509(7500):371-5.

1650F

Deconvolution for Mixed Cancer Transcriptomes from Heterogeneous Tumor Samples with Immunal Infiltration. *Z. Wang[1], C. Holmes[4], W. Lu[5], X. Tang[5], I. Wistuba[5], J. Morris[3], W. Wang[2].* 1) Statistics, Rice University, Houston, TX; 2) Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX; 3) Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX; 4) Statistics, University of Oxford, Oxford, United Kingdom; 5) Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX.

Tumor tissue samples are comprised of a mixture of cancerous and surrounding stromal cells. Understanding tumor heterogeneity is crucial to analyzing gene signatures associated with cancer prognosis and treatment decisions. Compared with the experimental approach of laser micro-dissection to isolate different tissue components, in silico dissection of mixed cell samples is faster and cheaper. A computational approach “DeMix” was previously developed to address this challenging problem when the observed signals are assumed to come from a mixture of two cell compartments, the cancerous tissue and its surrounding stroma. We have significantly extended DeMix to develop a three-component deconvolution model, DeMixT, that can account for a third component such as the immune cell compartment explicitly and is more realistic than a two-component model. Straightforward generalizations of the two-component DeMix model to the three-component setting are not computationally feasible because of the need to compute high-dimensional integrals. DeMixT therefore involves a novel two-stage filtering method yielding accurate estimates of cell purities and compartment-specific expression profiles. Simulations and real data analyses have demonstrated the good performance of our method. Compared to DeMix, DeMixT can be applied more widely and provides deeper insight in cancer biomarker studies. It allows for a further understanding of immunal infiltration in cancer and assists in the development of novel prognostic markers and therapeutic strategies.

1651W

Correcting nucleotide-specific biases in high-throughput sequencing data. *J. R. Wang[1], T. S. Furey[2].* 1) Department of Genetics, University of North Carolina, Chapel Hill, NC; 2) Departments of Genetics and Biology, Carolina Center for Genome Sciences, Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC.

High-throughput sequencing (HTS) assays use short (50-250bp) reads to determine the genome, transcriptome, and functional activity of biological samples. HTS-based methods include whole-genome sequencing (DNA-seq), transcription profiling (RNA-seq),chromatin accessibility (DNase-seq, FAIRE-seq, ATAC-seq), and transcription factor binding (ChIP-seq). These methods exhibit nucleotide-specific biases at specific read positions as a result of assay and/or HTS preparation steps. For example, DNase-seq data shows large biases at the 5' end as a result of the DNase I hexamer binding/cutting preferences. Common HTS protocol steps include sonication and adapter-mediated PCR, which may also introduce subtle biases in the final reads. We have developed a method to adjust for observed nucleotide-specific biases in HTS data, regardless of the bias source. Briefly, we first build a model of both the observed and expected nucleotide content within and surrounding each read. We then calculate read weights based on combined ratios of expected component k-mer frequencies relative to observed frequencies. We applied our method to DNA-seq, DNase-seq, FAIRE-seq, ATAC-seq, and ChIP-seq data from GM12878, K562, and LNCaP human cell lines produced at different labs, times, and under various experimental conditions. Using principal component analysis, we observed that the bias signature of these data cluster primarily by assay and lab, as expected. After bias correction, the average nucleotide frequency variance across reads is reduced from 0. 012 to 0. 0052. DNase, FAIRE, and ATAC commonly measure chromatin accessibility where transcription factors bind. Therefore, we compared correlations between these assays before and after bias correction. We used ChIP-seq data from CTCF, a widespread and abundant regulator. After bias correction, we observed an average increase in covariance from 0. 532 to 0. 544 between features predicted by these assays from the same cell line. The convergence of these corrected data implies they more accurately represent true epigenetic features. Additionally, corrected DNA-seq exhibited a shift in the ratio of alleles represented at heterozygous loci toward 1:1. The reduced level of nucleotide-specific biases at these loci can help identify true heterozygotes and differentiate these from misalignments, sequencing error, and homologous sequences.

1652T

HGSC Variant warehouse, a scalable, varianted-centered database running in hadoop. *S. J. White[1], S. Ambreth[1], T. Chiang[1], M. Dahdouli[1], A. English[1], A. Hawes[1], Z. Kahn[1], W. Salerno[1], N. Veeraraghavan[1], E. Boerwinkle[1,2], DM. Muzny[1], RA. Gibbs[1].* 1) HGSC, Baylor College of Medicine, Houston, TX; 2) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX.

Personalized medicine is being ushered in tandem with large cohort studies, leveraging DNA sequencing technologies with higher throughput, lower cost per sample, and larger sample sizes. Consequently there are demands in terms of scale, heterogeneity of data and richness of annotation. We present a Hadoop variant data warehouse, which is flexible and scalable and provides rapid access to massive data sets. We use Hbase to store variant data in a NOSQL schema where rows represent variants and columns represent samples. This allows for rapid retrieval of data for a given variant / loci, typically with millisecond response times across thousands of samples. Sources of annotation such as DBSNP, HGMD, COSMIC are easily imported into the same schema. By holding all these data together we can move away from sequentially annotating and prioritizing variants for each sample to a paradigm of continual annotation across all samples, using allele frequencies generated on the fly. Using this technology in a CAP-CLIA environment will enable us to automatically flag unsolved cases for re-analysis as soon as new disease associations are published. In addition to variants we also encode read depth and base quality. By allowing us to distinguish between reference homozygous calls and regions with no coverage, these data provide us with a variant-caller agnostic solution to the "N+1" population level VCF (pVCF) problem. This significantly reduces compute time and simplifies the process of comparing large cohort studies consisting of heterogeneous data from different pipelines and sequencing centers. Storing coverage and quality increases the data footprint significantly; by using a novel, highly compressed, run-length encoded format we store at base level resolution, all the required data for a 30x whole genome in around 5Gb. This represents around 2% of the BAM size and includes data triplication within Hbase, making practical the persistent storage of 10's of thousands of whole genomes, such as the anticipated Illumina X-Ten production for the CHARGE cohort. Hadoop provides us with a flexible ecosystem of open-source tools, allowing us to easily integrate other sources of meta data such as LIMS, phenotypic data and electronic medical records. We will describe applications of the Variant Warehouse to more than 500 whole genomes, including variant annotation, summary statistics and QC, data footprint and pVCF generation.

1653F

NGS-SWIFT: A Cloud-Based Variant Analysis Solution Using Control-Accessed Sequencing Data. *C. Xiao, E. Yaschenko, S. Sherry.* NIH, 45 Center Drive, Bethesda, MD.

Variation analysis plays an important role in elucidating the causes of various human diseases. The drastically reduced costs of genome sequencing driven by next generation sequence technologies now make it possible to analyze genetic variations with hundreds or thousands of samples simultaneously, but currently with the cost of ever increasing local storage requirements. The tera- and peta-byte scale footprint for sequence data imposes significant technical challenges for data management and analysis, including the tasks of collection, storage, transfer, sharing, and privacy protection. Currently, each analysis group facing these analysis tasks must download all the relevant sequence data into a local file system before variation analysis is initiated. This heavy-weight transaction not only slows down the pace of the analysis, but also creates financial burdens for researchers due to the cost of hardware and time required to transfer the data over typical academic internet connections. To overcome such limitations and explore the feasibility of analyzing control-accessed sequencing data in cloud environment while maintaining data privacy and security, here we introduce a cloud-based analysis framework that facilitates variation analysis using direct access to the NCBI Sequence Read Archive through NCBI sratoolkit, which allows the users to programmatically access data housed within SRA with encryption and decryption capabilities and converts it from the SRA format to the desired format for data analysis. A customized machine image (ngs-swift) with preconfigured tools (including NCBI sratoolkit) and resources essential for variant analysis has been created for instantiating an EC2 instance or instance cluster on Amazon cloud. Performance of this framework has been evaluated using dbGaP study phs000710. v1. p1 (1000Genome Dataset in dbGaP, http://www. ncbi. nlm. nih. gov/projects/gap/cgi-bin/study. cgi?study_id=phs000710. v1. p1), and compared with that from traditional analysis pipeline, and security handling in cloud environment when dealing with control-accessed sequence data has been addressed. We demonstrate that it is cost effective to make variant calls using control-accessed SRA sequence data without first transferring the entire set of aligned sequence data into a local storage environment, thereby accelerating variation discovery using control-accessed sequencing data.

1654W

A hidden Markov random field based Bayesian method for the de-tection of long-range chromosomal interactions in Hi-C Data. *Z. Xu[1,2,3], G. Zhang[3,4], F. Jin[5], M. Chen[1,2], T. S. Furey[2], P. F. Sullivan[2,6], Z. Qin[7], M. Hu[8], Y. Li[1,2,3].* 1) Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Curriculum in Bioinformatics and Computational Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 5) Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, CA 92093, USA; 6) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; 7) Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, 1518 Clifton Road Atlanta, GA 30322, USA; 8) Division of Biostatistics, Department of Population Health, New York University School of Medicine, 650 First Avenue, New York, NY 10016, USA.

Advances in chromosome conformation capture and next-generation sequencing technologies are enabling genome-wide investigation of dynamic chromatin interactions. For example, Hi-C experiments generate genome-wide contact frequencies between pairs of loci by sequencing DNA segments ligated from loci in close spatial proximity. One essential task in such studies is peak calling, that is, the identification of non-random interactions between loci from the two-dimensional contact frequency matrix. Successful fulfillment of this task has many important implications including identifying long-range interactions that assist in interpreting a sizable fraction of the results from genome-wide association studies (GWAS). The task – distinguishing biologically meaningful chromatin interactions from massive numbers of random interactions – poses great challenges both statistically and computationally. Model-based methods to address this challenge are still lacking. In particular, no statistical model exists that takes the underlying dependency structure into consideration. We propose a hidden Markov random field (HMRF) based Bayesian method to rigorously model interaction probabilities in the two-dimensional space based on the contact frequency matrix. By borrowing information from neighboring loci pairs, our method demonstrates superior reproducibility and statistical power in both simulations and real data.

1655T

Pathway analysis of genomic loci uisng the *seq2pathway* Bioconductor package. *X. H. Yang, B. Wang, J. M. Cunningham.* Section of Hematology/Oncology, Department of Pediatrics, the University of Chicago, Chicago, IL.

IntroductionPathway analysis improves biological interpretability of genomic features with respect to prior knowledge and has led to important biological discoveries (Khatri, et al. , 2012). However, non-coding regulatory loci pose a critical challenge, as nearly 99% of the nucleotides in the human genome do not code for proteins but instead are enriched for critical functional genetic regulators (Kellis, et al. , 2014). The first sequence-derived pathway analysis method is GREAT (the Genomic Regions Enrichment of Annotations Tool) (McLean, et al. , 2011). GREAT quantifies a pathway score by determining if the total number of loci within the regulatory domain of genes in a gene-set is greater than expected. The ChIP-Enrich method then empirically adjusts the length of the gene body and its surrounding non-coding sequence (Welch, et al. , 2014). However, both methods treat all loci equally, despite the fact that a threshold for significance is always experimentally context-dependent and arbitrary. We propose a new sequence-derived pathway analysis approach named “seq2pathway”, considering quantitative sequencing measurements (Wang, et al. , 2015). Seq2pathway provides the community a Bioconductor package that **1)** quantifies the functional impact of both coding genes and non-coding loci (Yang, et al. , 2015), **2)**generalizes four pathway analysis methods, including FAIME (Yang, et al. , 2012), to variable next-generation sequencing (NGS) data, and **3)** wraps gene locus definitions for human and mouse genomes from GENCODE into R objects. **Main features**Seq2pathway is designed for knowledge discovery using a variety of NGS data (eg, ChIP-seq, RNA-seq, SNPs, etc) by taking the contribution of distal non-coding loci and their experimental significance scores into consideration. Users can apply the “seq2pathway” function jointly, or in a two-step algorithm consisting of “seq2gene” and “gene2pathway” components separately. The seq2gene step collapses significant scores per gene when a gene was mapped to multiple neighboring genomic loci, removing the bias of linkage disequilibrium for the downstream pathway analysis. The gene2pathway step condenses gene-by-sample measurements to pathway-by-sample measurements, which are gene-coverage-difference tolerable and free of pathway size preference. With a search radius of 100 kbps, seq2pathwaydefined a novel cis-regulatory element from both ChIP-seq and transcriptomic data (Hoffmann, et al. , 2014). **Availability**www. bioconductor. org/.

1656F

Development of an effective clinical interpretation solution for genomic data. *P. Yu, M. Wei, J. Li, G. Hu.* Admera Health LLC, South Plainfield, NJ.

With the rapid cost decline of high-throughput sequencing data generation, one major hurdle in making precision medicine a reality is to turn genomic data into clinical information that can be easily used in medical practice. Even though it has been pursued by many institutions and companies, a truly effective solution is still elusive, which at minimal requires: 1) a fully automated data analysis and interpretation workflow from raw data to clinically actionable report; 2) a scalable and reliable IT infrastructure; 3) a comprehensive and high-quality database; 4) a clinically validated process that meets regulatory requirements; 5) accurate, fast and cost-effective; and 6) being demonstrated in real medical practice with great satisfaction. To address this unmet need, we have developed Actionable Genomic Interpretation System (AGIS) that has all above features. Currently it contains three major modules that can be used separately or in combination: Pharmacogenomics Module for guiding personalized disease treatment; Disease Risk Module for disease risk assessment; Genetic Disease Module for diagnosing hereditary diseases. In this presentation, besides introducing the overall design and functionalities of AGIS, we will also report a real clinical application case that AGIS has successfully enabled the delivery of pharmacogenomic testing results for thousands of patients. The prototype of the AGIS system is available at: https://agis. admerahealth. com/.

1657W

Introducing Awsomics: a cloud-based bioinformatics toolbox. *Z. Zhang.* Department of Biomedical and Health Informatics, The Children’s Hospital of Philadelphia, Philadelphia, PA.

The processing and management of evergrowing genomics data is dominating the attention of bioinformatics field. However, the long-term bottleneck of using genomic data in biomedical research is the information-to-knowledge process, which is usually carried out by biomedical researchers who have been studying specific biomedical questions through their careers. Proper and efficient usage of genomic data and information are often challenging to these researchers, and we developed Awsomics toolbox to assist them through this process. The frontend of this toolbox is a series of independently run APPs that were designed to address research questions commonly arisen from genomic data sets. The APPs are supported by a backend collection of curated, high quality collection of genomic data, named Rchive. The contents of Rchive range from standard annotation databases, such as dbSNP and NCBI Gene, and curated experimental data, such as GWAS results and transcriptome data sets, that can be directly applied to integrative analysis without further processing. The backend Rchive and frontend APPs are connected through a number of data mining algorithms that integrate multiple data sets or different types of data. Two APPs have been implemented in Awsomics. GWAS Hub is a tool that searches and summarizes GWAS results from 7,000+ analyses reported by databases such as dbGaP and GWAS Catalog. Users can use it to look up their favorite SNPs, genes, and genomic regions in previous GWAS results, and link them to chemicals, diseases, and other genes through the annotation of these results. The other APP, GeEx (Gene expression Explorer), is a tool for exploring collections of transcriptome data sets. Data sets in each collection could be generated from different experiments, experimental conditions, technical platforms, or even species, but all share a common theme. For example, the ToMD (Transcriptomes of Mitochondrial Dysfunctions) collection currently includes 119 sample groups in 35 data sets, each including samples with impaired mitochondrial functions or respective control samples. All data sets were properly normalized and formatted, so users can easily utilize GeEx functions to query, compare, and visualize them together. In conclusion, Awsomics toolbox provides platform for faster access of genomic data and knowledge discovery. The number of its APPs will keep growing to cover a variety of research applications. Awsomics is available at: http://awsomics. org.

1658T

Sequence Analyzer (SeqA): a computationally efficient large-scale sequence data quality control pipeline and analysis framework. *D. Zhang, G. T. Gao, S. M. Leal.* Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX.

The ability to harvest massive genomic data through cutting-edge technologies has opened spectacular opportunities for discovering moderate effect variants in complex disease etiology via sequencing based association studies involving hundreds of thousands samples. Current analytical tools for association analysis are becoming obsolete, because of their inability to scale with "big data", which is a major impediment to successful gene-mapping practices. Here we report on Sequence Analyzer (SeqA), a new analysis platform for extremely rapid quality control and association testing of large-scale sequence data. Powered by modern distributive computing system, i. e. , Spark and Hbase, SeqA performs quality control for association studies and provide reports on various data attributes including but not limited to genotype call quality, variant/sample level missing rates evaluation, global ancestry and cryptic relatedness analysis, genetic sex check, and batch-effect analysis. Data query and processing using SeqA is drastically faster than PSEQ and VAT and can also handle much larger data sets due to the distributed database infrastructure optimized for management of large quantities human genome data which are sparse in nature. SeqA also intensively utilizes Spark's statistical libraries to perform ultra-fast rare variant association analysis via automatically paralleled computational procedures, and in addition to providing most widely used statistical methods for common and rare variant association testing, an Application Program Interface (API) is available for incorporation of user-specific analyses in the SeqA platform. Although SeqA is capable of handling up to a million whole genome sequences on a moderate local cluster environment, it can also be deployed to commercially available cloud computing platforms such as Amazon EC2. To our knowledge, SeqA is the first tool to implement the current best practice of sequence based association analysis workflow in a scalable and simple fashion readily available for use in modern genetic epidemiological research projects, maximizing the potential to uncovering novel variants in massive amount of data formidable to traditional analysis toolkits.

1659F

Speciman: A new low-frequency variant caller with significantly increased specificity. *J. Zhang[1,2], H. Hikbakht[1,2], J. Majewski[1,2].* 1) McGill University, Montreal, Quebec, Canada; 2) Genome Quebec Innovation Centre.

Introduction The identification of point, tumor-only mutation remains challenging, mainly because the ratio of DNA harboring the mutation in the sample is often low. However, precisely identifying low frequency variants (LVFs) in tumor samples is crucial to understand cancer heterogeneity, tumor evolution and treatment response. Many variant callers have been developed to increase the sensitivity of detecting LVFs, but lead to a large number of false positives. Accurately estimating site-specific error rates can significantly increase specificity. Here, we introduce a new tool that calling variants from exome sequencing using information from large set of tumor or normal samples. **Methods**To distinguish true LVFs from errors, we prepared a control cohort using 500 blood samples and for each sites we first filtered reads based on base quality and mapping quality. We also filtered reads if the same read name or read position were seen in the same sample. We then counted the reference and altered reads for all samples. For variant calling from only one pair of tumor and matched normal, we calculated error rates across all sites with potential variants based on the read count of control samples with mutant allele ratio less than 10%. We then performed binomial tests followed by FDR corrections in both tumor and matched normal. For variant calling from multiple tumor samples, we calculated the mutant allele ratio in each tumors and controls, and identified LVFs using *k*-means clustering. **Results and Conclusions** To evaluate the sensitivity of our algorithm, we merged one tumor sample benchmarked by Sanger validation with matched blood sample to lower the mutant allele ratio. To evaluate the specificity, we used one sample underwent exome sequencing twice, and therefore any variants should be false positives. Compared to MuTect, our method showed similar sensitivity, but significantly increased specificity. Moreover, we performed somatic variant calling in exome sequencing studies of five spatially distinct biopsies from a primary glioblastoma multiforme (pGBM) tumor and the matched blood sample, followed by targeted re-sequencing validation. We found low-frequency mutation in PIK3CA contributed to tumor profession, and provided the potential applications of our approach to similar studies.

1660W

Zodiac: A Comprehensive Information System of GeneticInteractions in Cancer Based on Statistical Models of TCGA Data. *Y. Ji[1,2], R. Mitra[6], Y. Xu[7], S. Yang[1], L. L. Pesce[3], P. Mueller[4], W. Guo[5], I. Foster[3], S. Sengupta[1], K. W. White[8,9,10], Y. Zhu[1].* 1) Program of Computational Genomics & Medicine, NorthShore University HealthSystem, Evanston, IL; 2) Department of Public Health Sciences, The University of Chicago, Chicago, IL; 3) Computation Institute, The University of Chicago, Chicago, IL; 4) Department of Mathematics, University of Texas at Austin; 5) School of Public Health, Fudan University, Shanghai, China; 6) Department of Bioinformatics & Biostatistics, University of Louisville, Louisville, KY; 7) Department of Applied Math and Statistics, Johns Hopkins University, Baltimore, MD; 8) Institute for Genomics and Systems Biology, The University of Chicago, Chicago, IL; 9) Department of Human Genetics, The University of Chicago, Chicago, IL; 10) Department of Ecology & Evolution, The University of Chicago, Chicago, IL.

Genetic interactions play a critical role in cancer development. Existing knowledge about cancer genetic interactions is incomplete, especially lacking evidences derived from large-scale cancer genomics data. The Cancer Genome Atlas (TCGA) produces multi-layer measurements across genomics features of thousands of tumors. We introduce Zodiac, a computational resource to integrate existing knowledge about cancer genetic interactions with new information contained in TCGA data. It is an evolution of existing knowledge by treating known genetic interactions and pathwaysas a prior graph, integrating it with a likelihood model derived by Bayesian graphical model based on TCGA data, and producing a posterior graph as updated and data-enhanced knowledgebase. In short, Zodiac realizes "*Prior interaction map + TCGA data –Posterior interaction map.* "**Zodiac provides molecular interactions for about 200 million pairs of genes.** All the results are generated from a big-data analysis and organized into a comprehensive database allowing customized online search. In addition, Zodiac provides data processing and analysis tools that allow users to customize the prior networks and update the genetic pathways of their interest. Zodiac recapitulates and extends existing knowledge of molecular interactions in cancer. It can be used to explore novel gene-gene interactions, transcriptional regulation, and other types of molecular interplays in cancer. Zodiac has independently confirmed recent case studies reported in cancer literature, such as novel gene-gene interactions related to BRCA1/2 and TP53. In addition, potentially novel discoveries are being reported in our online website**www. compgenome. org/ZODIAC**. Some discoveries appear to be useful for drug discovery in oncology. We will report some results in this presentation.

1661T

Genomic analysis of multiple data types in cancer patient cohorts in GenePool platform allows fast and efficient multi-dimensional view of biological mechanisms. *A. Vladimirova, R. D. Goold, S. Sanga, T. Klingler.* Station X , Inc. , San Francisco, CA.

The ability to analyze genomic data generated by using different and complementary technologies allows us the opportunity to generate multi-dimensional views of the biological mechanisms involved in disease, and to formulate integrative hypotheses based on all-encompassing results from orthogonal molecular tests that reveal the full spectrum of molecular alterations observed in the patient samples. GenePool™, a commercial software application for biomarker discovery and validation in large patient cohorts (Station X, San Francisco, CA), is a cloud-based genomics platform for analysis, management, visualization and sharing of genomics data and results, where we have integrated, curated and organized TGGA multi-dimensional reference data from 33 cancer patient cohorts spanning copy number, methylation, protein arrays, miRNA, RNA-seq and exome information, and including primary tumor, blood, and in many cases, metastasis and adjacent normal tissue samples per patient. We used this multi-dimensional TCGA data in GenePool platform to integrate diverse data types, and probe entire cohorts of hundreds of patients to uncover biological mechanisms that span multiple data types. We demonstrate examples where GenePool was able to quickly and efficiently integrate a variety of genomic data, analyze and visualize it in the cloud to explore entire patient cohorts and generate testable hypotheses in real time. We present results of integrative analysis of multiple molecular assays in a patient cancer cohort. In GenePool platform reference data can be further combined and extended with proprietary data sets to illuminate new mechanisms or validate existing hypotheses to further our understanding of the disease and highlight approaches for tumor treatment.

1662F

Decoding the code of chromosomal translocation: a computational approach. *R. M. Rawal.* Cancer Biology, The Gujarat Cancer & Research Institute, Ahmedabad, Gujarat, India.

Recurrent chromosomal rearrangements including deletion, duplication and non-random chromosomal translocations are hallmark characteristics of leukemogenesis however molecular mechanisms underlying these rearrangements are still not completely understood. The fundamental question is, why and how chromosomes break and reunite so precisely in the genome. Meticulous understanding of mechanism leading to chromosomal rearrangement can be achieved by characterizing breakpoints. To address this hypothesis, a novel multiparametric computational approach for characterization of most frequent leukemic translocations within and around breakpoint region was performed. Breakpoint islands were also analyzed for other complex genomic architecture and physical properties. Our study distinctly emphasizes on the probable role of SDs, quadruplex and various genomic features in breakage and reunion of chromosomes leading to translocation. Various physico-chemical parameters were also taken into consideration for characterization of chromosomal breakpoint island. To best of our knowledge, this is an unique approach to decipher the basis of chromosomal translocation in leukemias.

1663W

Design and Implementation of an Cloud-Based Graphical Interface for Variant Exploration and Classification. *X. Xu, J. Cohen, D. Caplan, D. Gross.* SolveBio, New York, NY.

High throughput sequencing has transformed the field of genetic diagnostics. Current practices for curation and clinical significance classification of sequenced genetic variants stem from Sanger-based sequencing of single genes or less and cannot handle the avalanche of variants identified from next-generation sequencing technologies. Geneticists currently take a median of 54 minutes per genetic variant to classify the variant's clinical significance (Dewey, 2014). The purpose of this project is to build a cloud-based graphical interface, designed specifically for the clinical geneticist or other variant curation expert, to improve speed and accuracy in clinical decision-making regarding sequenced genetic variants.

Variant analysis requires accurate and up-to-date genetic reference annotation data. We have built a Data Library, that, as of June 2015, has indexed 25 publicly available reference data sources with 154 individual datasets, ~1. 2 billion independent records, and ~28 billion data points. All of these data has been integrated into one platform, versioned, and made programmatically accessible via an Application Programming Interface (API). With this underlying data infrastructure, we have built the Variant Explorer (VE): a cloud-based graphical interface that displays variant-specific data. The VE consists of modular components: variant identification, genomic visualization/localization, clinical evidence from publicly available sources, and variant effect/consequence prediction. Each displayed variant is the result of querying and combining a dozen disparate reference datasets.

We have also designed a cloud-based graphical interface for variant curation and classification built on top of the Variant Explorer. This interface guides an expert curator through a series of steps that systematically and computationally codify each evidence criteria in the American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) standards and guidelines for the interpretation of sequence variants (Richards, 2015).

NGS-based tests has the potential to dramatically improve molecular diagnostics and understanding of the genetic etiology of diseases. Geneticist-guided variant classification will ameliorate the analysis bottleneck in the variant curation and classification step.

1664T

Run Smarter, Not Harder: Scalable Joint Variant Discovery with GATK. *S. Chandran[1], G. Van der Auwera[1], R. Poplin[1], V. Ruano-Rubio[1], M. DePristo[1], T. Fennell[1], M. Carneiro[1], D. Kling[1], L. Gauthier[1], A. Levy-Moonshine[1], D. Roazen[1], B. Haas[1], K. Shakir[1], J. Thibault[1], D. Benjamin[1], C. Whelan[1], S. Ashrafzadeh[2], M. Lek[2], S. Gabriel[1], D. Altshuler[1], M. Daly[1,2], D. MacArthur[1,2], E. Banks[1].* 1) Data Sciences and Data Engineering, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA 02114, USA, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, and Department of Medicine, Harvard Medical School, Boston, Ma.

Studying a single individual's genome or exome in isolation is almost never useful: practically every genomic study design involves comparing one or several individuals to another group of individuals. However, traditional multisample analysis methods in which all samples are processed simultaneously through variant calling and genotyping scale very poorly. This is especially the case of the newer assembly-based methods, which yield superior indel calls but impose a heavy computational cost. In addition, adding any new samples to existing analyses entails re-running compute-intensive algorithms on the entire cohort, which is costly and slow -- the so-called "N+1 problem".

Here we present a scalable approach that overcomes these obstacles without sacrificing accuracy. Firstwe determine genotype likelihoods individually per-sample for all possible alleles using the assembly-based HaplotypeCaller. Thenwe perform joint genotyping across all samples within a project withGenotypeGVCFs. This yields the same great accuracy as traditional multisample calling and the superior indels calls of an assembly-based method, while scaling efficiently to ridiculously large cohort sizes. It also solves the N+1 problem since samples can be added incrementally; only the joint genotyping step needs to be re-run, which is computationally trivial. On top of that, the Reference Confidence Model (RCM) implemented within HaplotypeCaller produces a fully squared-off matrix of genotypes across all samples at every genomic position being investigated.

The HaplotypeCaller + GenotypeGVCFs workflow is the most scalable and incremental workflow to yield highly accurate variant calls, with abundant applications for population genetics and clinical studies.

1665F

Risk Allele Distribution Visualizer (RADViz-QT): A novel computational tool to estimate continent-specific differences in Cumulative Genetic Risk (CGR) for quantitative traits and diseases. *A. Choudhury[1], M. Ramsay[1,2].* 1) SBIMB, University of the Witwatersrand, Johannesburg, Gauteng, South Africa; 2) Division of Human Genetics, National Health Laboratory Service and School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa.

Background : Population genetic variation is shaped by history (population size, migration, bottlenecks and admixture), environment (selection) and chance (new mutations and genetic drift). These forces also influence genetic susceptibility to diseases and may lead to population-level differences in predisposition. Previous studies based on cumulative frequency distribution comparison of risk alleles of SNP sets identified in genome wide association studies (GWAS) have shown the genetic risk for some diseases to vary significantly across continents. These studies have not addressed quantitative traits (QTs) nor have they investigated intra-continental differences in risk allele distributions. We examine the spread of risk alleles from ~90 QT GWA studies, across and within continents. **Methods:** The risk-associated alleles for SNP sets identified in GWA studies were retrieved from the GWAS catalog and examined using the 1000 Genomes Project (1KGP) population genetic data. A simple additive model was used to integrate the risk status for all the SNPs identified in a GWA study into a cumulative genetic risk (CGR) statistic. The CGR distributions were compared across populations and continents. **Results:** Risk allele sets from about 10% of the QT studies (including eGFR levels related to chronic kidney disease (CKD), bone mineral density (BMD), and menopause) exhibit significant high differentiation between one or more pairs of continental populations. SNPs sets associated to 15% of the QTs (including breast size, waist to hip ratio and systolic blood pressure) were found to show high intra-continental CGR differentiation. Both inter and intra-continental CGR differentiations were observed for risk alleles from some of the QTs. Moreover, alignment between the geographical distribution of CGR and the associated phenotypes was observed for SNP sets related to QTs like CKD and BMD. **Conclusions**: The levels of inter and intra-continental differentiation of genetic risk for complex traits and diseases can be used to formulate hypotheses on geographical variation in traits and diseases, and to inform country-specific public health intervention policies. As GWAS data accumulates for more populations, this approach will become both more accurate and useful. A tool- Risk Allele Distribution Visualizer (RADViz-QT) has been developed to assess and compare CGR scores between continents and populations (www. bioinf. wits. ac. za/software/radvizqt).

1666W

Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *J. Fan[1], N. Salathia[2], R. Liu[3], G. Kaeser[4], Y. Yung[4], J. Herman[1], F. Kaper[2], J. B. Fan[3], K. Zhang[3], J. Chun[4], P. V. Kharchenko[1].* 1) Center for Biomedical Informatics, Harvard University, Boston, MA; 2) Illumina Inc. , San Diego, CA; 3) Department of Bioengineering, University of California, San Diego, CA; 4) Department of Molecular and Cellular Neuroscience, Dorris Neuroscience Center, The Scripps Research Institute, La Jolla, CA.

Single cell RNA-sequencing provides the means to dissect the transcriptional composition of complex tissues and organs, including the mammalian brain. The transcriptional state of each cell, however, reflects a variety of biological factors, including persistent cell-type specific regulatory configurations, transient processes such as cell cycle, local metabolic demands, and extracellular signals. All such aspects of transcriptional heterogeneity may be of potential interest, but detecting and resolving them from inherently noisy single-cell data presents analytical challenges. We developed PAGODA to recover multiple, potentially overlapping, aspects of transcriptional heterogeneity by identifying known pathways and novel gene sets that exhibit a significant excess of coordinated variability among the measured cells. We demonstrate that our method effectively recovers known cellular subtypes and their corresponding functional characteristics in a variety of single-cell samples. Furthermore, we apply PAGODA to characterize the transcriptional diversity of neuronal progenitor cells in the developing mouse brain.

1667T

BDQC: a general-purpose analytics validation tool for Big Data Discovery Science. *G. Glusman[1], R. Kramer[1], E. W. Deutsch[1], I. Foster[2], C. Kesselman[3], R. Madduri[2], K. Chard[2], B. D. Heavner[1], I. D. Dinov[4], J. Ames[5], J. Van Horn[5], N. D. Price[1], L. Hood[1], A. W. Toga[5].* 1) Institute for Systems Biology, Seattle, WA; 2) Computation Institute, University of Chicago and Argonne National Laboratory, Chicago, IL; 3) Information Sciences Institute, University of Southern California, Los Angeles, CA; 4) Statistics Online Computational Resource (SOCR), UMSN, University of Michigan, Ann Arbor, MI; 5) Laboratory of Neuro Imaging, USC Stevens Neuroimaging and Informatics Institute, University of Southern California, Los Angeles, CA.

Biomedical data acquisition is generating exponentially more data: thousands of whole-genome sequences (WGS) are now available; brain data is doubling every two years. Analyses of Big Data, genomic or otherwise, presents qualitatively new challenges as well as opportunities. Among the challenges is a proliferation in ways analyses can fail, due largely to the increasing length and complexity of processing pipelines. Anomalies in input data, runtime resource exhaustion or unavailability of nodes in a distributed computation can all cause pipeline hiccups that are not necessarily obvious in the output. Flaws that can taint results may persist undetected in complex pipelines, a danger amplified by the fact that research is often concurrent with the development of the software on which it depends. On the positive side, the huge sample sizes increase statistical power, which in turn can motivate entirely new analyses. We have developed a framework for Big Data Quality Control (BDQC): an extensible set of analyses, heuristic and statistical, that identify deviations in data without regard to its meaning (domain-blind analyses). BDQC takes advantage of large sample sizes to classify the samples, estimate distributions and identify outliers. Such outliers may be symptoms of technology failure (e. g. , truncated output of one step of a pipeline for a single genome) or may reveal unsuspected “signal” in the data (e. g. , evidence of aneuploidy in a genome). We have applied the framework to validate our WGS analysis pipelines. BDQC successfully identified data outliers representing genome analyses missing a whole chromosome or part thereof, hidden among thousands of intermediary output files. These failures could then be resolved by reanalyzing the affected samples. BDQC both identified hidden flaws (in some cases, in software deemed “too simple to fail”) as well as yielded unlooked-for insights into the data itself. BDQC is meant to complement quality software development practices. There are multiple benefits from application of BDQC at all pipeline stages. By controlling input, it can help avoid expensive computations on flawed data. Analysis of intermediary and final results facilitates recovery from aberrant termination of processes. All these computationally inexpensive verifications reduce cryptic analytical artifacts that could seriously preclude clinical-grade genome interpretation. This Big Data for Discovery Science work is supported by NIH 1U54EB020406.

1668F

A novel approach to genome-wide association analysis: *S*NP *P*airs *I*n *N*eighboring *D*istance and *L*inkage *E*quilibrium (SPINDLE). *A. P. Khawaja[1], L. R. Pasquale[2,3], J. H. Kang[3], J. N. Cooke-Bailey[4], M. A. Hauser[5,6], J. L. Haines[4], J. L. Wiggs[2], NEIGHBORHOOD Consortium.* 1) NIHR Biomedical Research Centre, Moorfields Eye Hospital NHS Foundation Trust and UCL Institute of Ophthalmology, London, UK; 2) Department of Ophthalmology, Harvard Medical School, Massachusetts Eye and Ear Infirmary, Boston, MA; 3) Channing Division of Network Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 4) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 5) Department of Ophthalmology Duke University Medical Center, Durham, NC; 6) Department of Medicine, Duke University Medical Center, Durham, NC.

Standard analysis of genome-wide association studies (GWAS) is limited by examining SNPs one at a time and ignoring potentially important nearby SNPs. Gene-based approaches are limited by the fact that all SNPs in a region are considered in aggregate, even though only a small proportion of independent SNPs within a gene may be relevant for disease. SPINDLE (**S**NP **P**airs **i**n **N**eighboring **D**istance and **L**inkage **E**quilibrium) is a new approach that only examines **pairs** of independent SNPs, thereby avoiding association signals being lost due to inclusion of nearby SNPs that are disease irrelevant. SPINDLE uses summary GWAS results as input and uses HapMap data to identify pairs of SNPs that are neighboring (<20Kb) and in linkage equilibrium (r2<0. 2). A combined *P*-value (com*P*) for each SNP pair is calculated using Fisher's method, and a Bonferroni corrected genome-wide significance threshold is calculated based on the number of SNP pairs SPINDLE identified. To test SPINDLE, we used data from the GLAUGEN and NEIGHBOR consortia (3430 cases, 3108 controls) which examined associations with primary open-angle glaucoma (POAG [MIM 603383]). We deleted genome-wide significant SNPs from the standard GWAS results and tested whether SPINDLE would identify significant SNP pairs in the same regions despite no individually significant SNPs. We further analysed normal tension glaucoma (NTG) and high tension glaucoma (HTG) subgroups, as these have been shown to have distinct genetic associations. SPINDLE identified the originally significant loci for POAG (in *CDKN2A* [com*P* = 3. 7x10-11] and near *PPM1A* [com*P* = 2. 5 x 10-8]) and NTG (in *CDKN2A* [com*P* = 1. 2x10-9]), despite individually significant SNPs not being present in the input files. Additionally, SPINDLE identified new loci that were not significant in the standard GWAS; in *GAS7* for POAG [com*P* = 1. 4x10-8], *SLC27A1* for NTG [com*P* = 3. 7x10-9], and in *AFAP1* for HTG [com*P* = 2. 0x10-8]. *GAS7* and *AFAP1* have previously been associated with POAG in independent studies. We sought replication of the *SLC27A1* locus in the Women's Genome Health Study (158 cases, 22508 controls); one SNP in the pair was nominally associated with POAG with the same direction of association (rs11672333, *P*=0. 016). In summary, SPINDLE is a new method for analysing GWAS data that has advantages over current gene-based approaches. Testing SPINDLE in a large GWAS supported its validity and identified a new locus associated with NTG in *SLC27A1*.

1669W

Genvisis: Visualization and batch correction of copy number variation at the population level leads to higher accuracy and the ability to analyze data on sex chromosomes. *J. Lane, B. Cole, Z. Xu, N. Pankratz.* University of Minnesota, Minneapolis, MN 55455.

Genvisis is an open-source, platform-independent Java software package that we designed to efficiently store, normalize, analyze and visualize genetic data. The underlying databases are fully scalable to hundreds of thousands of samples with genome-wide data that would comprise terabytes of data even when using our unique compression algorithms. Even so, we structure the data so that any individual or any marker can be loaded into memory in milliseconds for visualization or analysis. This has therefore become a platform for the efficient review of sample quality, locus quality, sex aneuploidy, chromosomal arm mosaicism, DNA contamination, batch effect correction, population stratification, joint calling, and more. It is also the foundation for the processing and analysis of copy number variants (CNVs), including data from the X and Y chromosomes. We have also developed novel methodology to improve blind duplicate concordance from 68% using standard methodology to as high as 91% for heterozygous copy number changes and 99% for homozygous deletions. This is accomplished by performing a principal component analysis (PCA) of the normalized intensity data at the allelic level (as opposed to at the marker level) in order to detect and correct for batch effects and DNA quality. Originally designed for Affymetrix and Illumina array data, the software has recently been extended to take advantage of next generation sequencing data either in processed (. VCF files) or raw alignments (. BAM files). One of the unique aspects of our visualization tools is the ability to interactively and dynamically link data points from one plot to another. For example, the genome browser can be used to visualize all of the CNVs on a particular arm of a chromosome, color-coding them by study, copy number, and calling algorithm. Clicking on a particular CNV will open up the raw intensity data for that individual for that region. Clicking on that intensity data for a particular marker will then open up a plot of intensity data for all individuals for that marker. The marker can then be manually re-clustered, which automatically re-computes the Log R ratio intensities, which can be fed back through the calling algorithms and updated for all samples. The uniform application of this software to multiple studies will allow for the meta-analysis of many traits and should identify new targets for functional research and future therapies. Manual and tutorials are available at http://genvisis. org.

1670T

KeBABS: an R/Bioconductor package for kernel-based analysis of biological sequences. *J. Palme, S. Hochreiter, U. Bodenhofer.* Institute of Bioinformatics, Johannes Kepler University, Linz, Austria.

The computational analysis of biological sequences is a fundamental task. On the one hand, sequence analysis methods have supplied highly valuable insights into how patterns/motifs in amino acid sequences govern protein structure. On the other hand, a large proportion of our current knowledge about how DNA sequence patterns control transcription factor binding, nucleosome positioning and remodeling, alternative splicing, etc. , is the result of computational sequence analysis. In genetics, discriminative sequence analysis is becoming increasingly important to predict potential effects of single-nucleotide variations in the context of surrounding sequences.

In the last two decades, kernel methods have been established as an important class of sequence analysis methods. For the classification of sequences, in particular, support vector machines (SVMs) have emerged as a sort of best practice. To apply SVMs for sequence analysis, it is necessary to either use a vectorial representation of the sequence data or to use kernels, that is, positive semi-definite similarity measures for sequences. The use of sequence kernels, however, is not limited to sequence classification. For example, they can also be used for regression tasks and similarity-based clustering.

This contribution is devoted to introducing KeBABS, a powerful, flexible, and easy-to-use framework for kernel-based analysis of biological sequences based on the widely used scientific computing platform R. KeBABS is publicly and freely available via the Bioconductor project (for more information, see http://www. bioinf. jku. at/software/kebabs). It includes efficient implementations of the most important sequence kernels, also including variants that allow for taking sequence annotations and positional information into account. KeBABS seamlessly integrates three common support vector machine (SVM) implementations with a unified interface. It allows for hyperparameter selection by cross validation, nested cross validation, and also features grouped cross validation. The biological interpretation of SVM models is supported by (1) the computation of weights of sequence patterns and (2) prediction profiles that highlight the contributions of individual sequence positions or sections. The features of the package will be described in detail along with illustrative biological examples.

1671F

HPMV: Human Protein Mutation Viewer. *W. A. Sherman[1], D. B. Kuchibhatla[1], V. Limviphuvadh[1], S. Maurer-Stroh[1,2], F. Eisenhaber[1,3,4], B. Eisenhaber[1].* 1) A*STAR Bioinformatics Institute (BII), Singapore, Singapore; 2) School of Biological Sciences (SBS), Nanyang Technological University (NTU), Singapore, Singapore; 3) Department of Biological Sciences (DBS), National University of Singapore (NUS), Singapore, Singapore; 4) School of Computer Engineering (SCE), Nanyang Technological University (NTU), Singapore, Singapore.

Next-generation sequencing advances are rapidly increasing thefraction of genetic disorders for which causative variants can beidentified. Our Human Protein Mutation Viewer (HPMV) can helpidentify causative variants in genetic disorders that are caused by asingle small change in a protein sequence (e. g. a non-synonymous pointmutation). HPMV fills a niche between initial variant filtering anddetailed analysis of individual mutations. It allows a researcher toquickly assess dozens of potential causative variants by presenting aninteractive cartoon that shows where each mutation of interest fallsalong the corresponding protein sequence - in relation to proteinfeatures that can help interpret the mutation. These protein featuresinclude post-translational modification sites, targeting signals,transmembrane helices, charge clusters, known conserved (Pfam)domains, and relevant 3D (PDB) structures, among others. As input, HPMV accepts protein mutations - as UniProt accessions withmutations (e. g. HGVS nomenclature), genome coordinates, or FASTAsequences. The main output of HPMV is its interactivecartoon. Clicking a sequence feature in the cartoon expands a treeview of additional information including multiple sequence alignmentsof conserved domains and a simple 3D viewer mapping the mutation torelevant PDB structures. A multiple sequence alignment of similarsequences from other organisms is provided directly below the cartoonand the cartoon itself includes a band showing the conservation ateach sequence position. In cases where a mutation is likely to havea straightforward interpretation (e. g. a point mutation in a wellunderstood targeting signal), this interpretation is suggested. The interactive cartoon is implemented as a web page using standardHTML, CSS, and JavaScript. It is also available as a standalone viewerimplemented in Java that can be downloaded in jar format, withembedded data, to be saved and viewed later with only a standard javaruntime environment. The HPMV website is: http://hpmv. bii. a-star. edu. sg/.

1672W

The ENCODE Analysis Pipelines: Tools for Repeatable, Standards-Based Analysis and Quality Control of Chromatin, Expression, and Methylation Experiments. *J. S. Strattan, T. Dreszer, B. C. Hitz, E. L. Hong, J. M. Cherry, ENCODE Data Analysis Center, ENCODE Data Coordinating Center.* Stanford University School of Medicine, Department of Genetics, Stanford, CA.

From Ammon's horn to zone of skin, members of the ENCODE Consortium have measured RNA quantity, RNA-protein interactions, DNA-protein interactions, DNA methylation, replication timing, chromatin structure, and histone modifications in over 4,000 experiments on more than 400 cell or tissue types. The ENCODE Data Analysis Center (DAC) have specified uniform processing pipelines for four ENCODE datatypes: ChIP-seq, RNA-seq, DNase-seq, and whole-genome bisulfite sequencing. The ENCODE Data Coordinating Center (DCC) have implemented these pipelines and deployed them to a cloud-based platform. Importantly, the pipelines can be run on the cloud via a web interface with no technical prerequisites other than input data. In this way researchers can perform the same analyses as ENCODE on their own data and repeat ENCODE analyses on ENCODE data. For ENCODE, the results of these analyses and metadata describing them are distributed through the ENCODE Portal, and illustrate general methods of accessing and interpreting ENCODE data. The ENCODE Portal is https://www. encodeproject. org/. The DCC codebase is freely available at https://github. com/ENCODE-DCC/.

1673T

PhASTESt, a user friendly tool for power calculations in pharmacogenomic studies with "time to event" outcomes. *H. Syed, A. Jorgensen, A. P. Morris.* Biostatistics, University of Liverpool, Liverpool, United Kingdom.

With the increased scale of genome-wide pharmacogenomic association studies, and the complexity of clinical outcomes they consider, there is a need for software to perform power calculations over a range of trial designs. There are already genetic power calculators available for binary phenotypes and quantitative traits, but the key outcome of interest in pharmacogenomics studies is often "time to event" (survival) data, which cannot adequately be modelled by existing software. To address this issue, we have developed the user friendly software tool PhASTESt (Pharmacogenomic Time to Event Association Study) to perform power calculations and generate sample pharmacogenomic data with time to event outcomes over a range of study designs. PhASTESt was built using C#, utilising pre designed methods packages Math. NET and Accord. NET for the generation of pharmacogenomic data and statistical analyses, respectively. PhASTESt is able to calculate the power to detect association of a SNP with a time to event outcome (at a pre-specified significance threshold) over a range of trial design scenarios. The software requires specification of genetic parameters, such as the magnitude of the SNP effect on the outcome and the effect allele frequency. The varied collection of design scenarios includes adding a recruitment period, multiple treatments with differential effects on outcome, SNP-treatment interactions, and/or different censoring options (for example, withdrawal due to adverse treatment event). Time to event and censoring data are then simulated on the basis of this model by making use of a Weibull distribution, so as to allow for the possibility of a deviation from a proportional hazards assumption. The data are then analysed by means of a Cox proportional hazards model. To allow for flexibility of analysis using methods not supported by the power calculator, individual simulated data sets can also be output from the software. In conclusion, this application offers a much needed user-friendly and flexible software tool for power calculations for time to event outcomes in pharmacogenomic association study designs.

1674F

Homozygous and hemizygous deletion CNV detection from exome sequence in a Mendelian disease cohort. *T. Gambin[1,2], Z. C. Akdemir[1], C. Shaw[1], S. Jhangiani[1,3], P. M. Boone[1], M. K. Eldomery[1], D. Muzny[1,3], E. Boerwinkle[3,4], R. A. Gibbs[1,3], A. L. Beaudet[1,3], J. R. Lupski[1,3,5,6], The Baylor-Hopkins Center for Mendelian Genomics.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Institute of Computer Science, Warsaw University of Technology, Warsaw, Mazowieckie, Poland; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 6) Texas Children's Hospital, Houston, TX.

We developed an informatics algorithm, HMZDelFinder, to discover rare, small (potentially single exon) homozygous and hemizygous deletions with a high sensitivity (84%) and specificity (53. 3%). HMZDelFinder is a joint CNV calling algorithm that uses read count information at each exon (RPKM) provided by BAM files together with independent B-allele frequency information calculated from VCFs. The identified homozygous and hemizygous intragenic deletions may play a role in the complete loss-of-function of the indicated gene in that personal genome. Our algorithm detected 13,868 deletions in total, consisting of homozygous (9,482) or hemizygous (4,386) deletion CNV calls from whole exome sequencing (WES) data for 4,124 samples in the Baylor-Hopkins Center for Mendelian Genomics (BHCMG) cohort. From 4,124 samples, 3,033 samples were found to have at least 1 homozygous or hemizygous deletion with the median number of 3 calls per personal genome. Interestingly, 97% of these deletions (13,449 out of 13,868) encompass merely a single exon. Investigation of homozygous and hemizygous deletions in the BHCMG cohort revealed 74 deletions of known disease genes in patients with phenotypes matching known OMIM (http://www. omim. org/) diseases associated with these genes. A subset of these known disease gene deletions was subjected to array comparative genomic hybridization (aCGH) and breakpoint junction specific PCR for orthogonal technology experimental verification of the bioinformatically identified CNV. The presence and zygosity of 17/17 known disease gene deletions were confirmed by at least one orthogonal experimental validation. Only 5 out of those 17 deletions (0 out of 6 single exon deletions) were detected by other state-of-the-art CNV calling methods (CoNIFER, CoNVeX, exonDel). Utilization of WESdata as a single comprehensive assay for both the detection of point mutations, InDels and CNVs is urgently needed. HMZDelFinder outperforms currently available CNV calling methods in terms of detection of small homozygous and hemizygous deletions. We demonstrated that this algorithm has a potential to increase the molecular diagnostic 'solution rate' and facilitate the process of novel disease genes discovery.

1675W

Detection, validation and visualization of copy number variations in targeted panels without matched normal samples using next generation sequencing data. *R. Gupta, A. Sathyanarayanan, M. Nagarajan, S. Katragadda, V. Veeramachaneni.* Strand Life Sciences, Bangalore, Karnataka, India.

Background and Objectives: Copy number variations (CNVs) are a specific form of structural variations (SVs) that are defined as large-scale amplifications or deletions (typically of size more than 1kb). Several algorithms are developed in the past to detect CNVs in whole-genome sequencing data, however recently detecting CNVs from targeted panel data is gaining attention of researchers and clinicians. Further to reduce the cost, data from matched normal samples is often not available. Therefore, the objective of this study is three-fold: 1) to accurately detect CNVs in targeted sequencing data without matched normal samples; 2) to independently validate the detected CNVs using supporting read % of the heterozygous SNPs in the CNV region; and 3) to visualize CNVs with estimated copy number values and confidence scores such that patterns either within or across multiple samples can be elucidated. **Methods:** Detection of CNVs in targeted sequencing data is challenging due to several technical and biological issues including noise in the data, alignment artifacts, absence of a reference from matched normal sample, tumor aneuploidy, normal/stromal cell contamination, heterogeneity in tumor cells etc. We have implemented a robust approach in Strand NGS software to detect, validate and visualize CNVs while addressing most of the above concerns. To alleviate the need for a matched normal sample, a coverage profile is first generated which stores mean and standard deviation of coverage across a group of samples for each target region. The group of samples can be a carefully selected pool of normal samples, which can be used as a reference to detect CNVs in other samples. Robust ways have been developed to ignore outlier values and corresponding target regions or samples from coverage profiles. **Results:** We demonstrated the applicability of the proposed approach using data from several clinical research samples available at Strand Life Sciences. We were able to detect CNVs in samples and narrow them down to few high confidence CNV calls using a z-score cut-off of 3. Z-score is computed for each target region by comparing sample mean to the profile mean coverage. In addition, these CNV calls are validated by measuring the deviation of supporting reads % from 50% of the heterogyzous SNPs called in the CNV regions. Copy number values and z-scores are also visualized across single sample as well as across multiple samples to quickly identify significant patterns.

1676T

CNValloc : a Bayesian approach for identifying copy unit alleles of CNVs from population-scale high-throughput sequencing data. *T. Mimori, N. Naoki, K. Kojima, Y. Sato, Y. Yosuke, Y. Yamaguchi-Kabata, M. Nagasaki.* Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan.

Recent development of high-throughput sequencing (HTS) technologies enable number of studies to reveal genetic variation among individuals, such asSNPs and short indels at nucleotide resolution. However, prediction of exact sequences of CNV is still challenging due to limited read length of HTS data, and CNV analysis has been mainly confined in predicting its existing regions and total copy numbers. We propose an approach that is named CNValloc to identify allele sequences at CNV locus from population-scale HTS data. CNValloc infers the copy unit alleles existing in population and their copy numbers for each individual simultaneously. The approach employs variational Bayesian inference for a graphical model inspired by latent Dirichlet allocation, which is a well known topic model in natural language processing field. We conducted simulation and real data studies to evaluate utility of CNValloc. In the simulation studies, degree of concordance between inferred and true alleles were evaluated for dataset with lower, middle, and higher copy number populations. We devised precision- and recall-like metrics for the evaluation and confirmed that the metrics were ≥ 0.9 for data with mean coverage $\geq 10\times$ per copy unit, which is a typical depth of coverage of HTS data. In real data study, we applied CNValloc to 1123 samples at highly variable salivary amylase gene locus and a pseudogene locus, and confirmed consistency of the estimated alleles within samples belonging to a trio of CEPH/Utah pedigree 1463 with 11 offspring. We also demonstrate that possible diplotypes of amylase alleles in the trio could be reconstructed from our results.

1677F

BAM Consensus: Copy Number Data Analysis for Targeted Panel NGS. *A. O'Hara, Z. Che, A. Venier, J. Fierro, S. Verma, S. Shams.* BioDiscovery Inc. , Hawthorne, CA.

Introduction: Targeted panel Next-Generation Sequencing (NGS) provides cost-effective single base pair resolution, with results focused on targeted genes and regions that have a clear diagnosis, are actionable with prescription drugs or compounds in clinical trials, or have known impacts on prognosis or outcome. Until recently, DNA microarray technology has been the gold standard for detection and analysis of copy number events. However, many labs with this technology must simultaneously test each sample on targeted NGS to detect sequence mutations and on DNA microarray to obtain the coordinating copy number information, which requires extra cost, labor, and time. Thus, using targeted panel NGS results for both copy number and sequence mutations is extremely valuable on many levels. **Methods**: We have developed a method called BAM Consensus which only requires the loading of BAM files from affected samples and targeted panel design files to generate copy number results. Germline normal reference samples are not required for analysis. In combination with the VCF files that contain the sequence mutations, an integrated analysis of both events can be easily carried out. **Results**: BAM consensus was applied to targeted sequencing panels in constitutional sample sets. Differences in overall read-depth resulted in variable sample quality across the cohorts, however most sample quality was adequate for copy number estimation and a quality threshold was assessed. Results indicate that relative copy number can be estimated for targeted regions comparable to the results achieved with microarray for the same targeted regions. Additionally, some larger arm length changes can also be estimated in genomic areas with wider targeted coverage. Integration with VCF files identified several samples with co-occurring copy number and sequence variant changes in targeted regions of interest. **Conclusions**: Targeted panel NGS is useful in constitutional samples as it provides a second tier of analysis for genes of interest, especially when no clear sequence mutation is detected. We have tested BAM Consensus in a variety of constitutional sample cohorts and it proves a very easy and straightforward way for targeted panel analysis.

1678W

A Multiple Feature Approach for Robust and Accurate Structural Variation Discovery for Whole Genome Sequencing Data of Varied Designs. *L. C. Xia[1,2,3], H. P. Ji[1,2], N. R. Zhang[3].* 1) Division of Oncology, Department of Medicine, Stanford University School of Medicine, Stanford, CA; 2) Stanford Genome Technology Center, Stanford University, Palo Alto, CA; 3) Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA.

We present SWAN, a statistical framework with open source software package for detection of genomic structural variants in next-generation sequencing (NGS) data. SWAN accommodates mixed library sequencing designs and adjusts for variation in fragment length distribution across multiple libraries. SWAN is shown to have improved sensitivity for small deletions, for insertions of all ranges, and for variants that may be present at low proportions in the sample. The scan statistics used by SWAN collectively quantify evidence available from read-pair, read-depth and hang-read through statistical likelihoods based on Poisson field models. At the core of SWAN is a novel marked Poisson Process model for paired-end sequencing, giving rise to likelihood ratio scan statistics that collectively quantify evidence from read-pair, hang-reads, and coverage. SWAN also remaps soft-clipped reads, thus allowing base-pair resolution for breakpoint junctions whenever possible. Using a spike-in dataset, we compared SWAN against mainstream structural variant callers in recovering simulated homozygous and heterogeneous variants of various sizes and types. SWAN maintained high accuracy, high precision and low false positives across a broad spectrum of SV sizes and purities. In addition, we benchmarked SWAN calls on NA12878 using long-read sequence data, PCR validation, and new targeted sequencing data. SWAN also performed the best among existing callers, attaining 30% higher sensitivity in detecting validated deletions than the runner-up caller at 50x coverage. At 5x coverage, SWAN was able to recover 40% of the validated deletions while the runner-up recovered 20% at the same false discover rate. SWAN is quite efficient in memory utilization and speed, and can be used for large-scale genome-wide structural variant genotyping and discovery based on NGS data. As a demonstration, we analyzed the whole genome sequencing dataset of a sample sequenced under 3 different library designs, and illustrate how the effective aggregation across libraries improves the sensitivity of calls. SWAN is freely available for academic use at http://bitbucket. org/charade/swan.

1679T

molQTL, An integrated NCBI resource for molecular QTL results. *N. Sharopova, A. Borkar, A. A. Schaffer, S. Sherry, M. Feolo.* NCBI/NLM/NIH, Bethesda, MD.

Molecular QTL (Quantitative Trait Loci) data are results of the analysis of association between genotypes and quantitative molecular phenotypes often generated by high throughput analysis pipelines processing data from technologies such as genotyping/expression/methylation arrays as well as DNA and RNA sequencing. The increasing volume and complex nature of molecular QTL data has motivated our development of a data archive and a web-based browsing tool that integrates genotype-phenotype association analysis data with a multitude of meta-information about genotypes, phenotypes, and analysis to facilitate data sharing, mining and interpretation. Here, we present a newly developed NCBI resource, molQTL, which archives and distributes molecular QTL data integrated with a variety of NCBI resources including dbSNP, BioSystems, Gene, Clinvar, dbGap, and GEO. Currently, molQTL database contains data from several published or submitted molecular QTL datasets with more than 17 million eQTL, methQTL, and mirQTL associations between 7,022,835 SNP and indel genomic markers and 38,195 molecular phenotypes derived using mRNA expression, methylation level, and microRNA expression assays. The integration of molQTL data is achieved by re-annotation of the datasets using the same versions/builds of annotation resources (SNP, GEO, Gene, BioSystems, Clinvar, dbGap, and MirBase). Annotation-based data integration allows investigators to conduct cross-dataset and cross-resource searching, for example, mining through 9,570,794 molQTLs using 2,840 Biosystems' pathways/terms. Investigators can publically access the data archived at molQTL using molecular QTL Browser, which currently has two views (i) by 'Study' - https://preview. ncbi. nlm. nih. gov/gap/eqtl/studies/ and (ii) by 'Bioprocess/Molecular Function', and via FTP. The browser allows investigators to search and filter the data using genome location, parameters of association analysis, and functional annotations (GWAS/Clinvar phenotypic traits, genes, biological pathways). By integrating data from a variety of molecular QTL datasets, molQTL promotes data sharing and data validation which is important for datasets generated using high-throughput lab and analysis methods. The presentation will summarize the resource data content and annotation pipelines.

1680F

Random Walk on Ontology for Rare Disorder Diagnosis. *C. Wu[1], M. A. Deardorff[2,3], M. Sarmady[1].* 1) Division of Genomic Diagnostics, Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, The University of Pennsylvania Perelman School of Medicine, Philadelphia, PA; 3) Division of Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA.

Recognition of a rare diagnosis to best explain a patient's clinical features has always been a challenging task. Computational methods have been designed to assist this process involving searching of structured and unstructured electronic databases. Specifically, the development of the Human Phenotype Ontology (HPO) to relate diseases, phenotypes and genes has provided an invaluable dataset from which to develop algorithms to assist in diagnosis. Despite of these efforts, clinical diagnosis remains formidable when an enormous number of clinical features with varying levels of specificity are shared among numerous rare disorders. In this work, we present a network-based approach using a stochastic process to simulate a random walker on the HPO to rank candidate diseases. Given a set of precise standardized HPO terms describing patient phenotypes, our method traverses the ontological tree of HPO to retrieve all of the direct and indirect children terms of each individual phenotype respectively and thereby builds a sub-network of all these terms. Following this, a PageRank *with priors* model is implemented on the formed sub-tree to prioritize the nodes and assign a score to each individual term. Finally, candidate diseases are ranked accordingly by summing the scores of the phenotypes they are associated with. Since it is difficult to evaluate a diagnostic method by using real patient data, simulations were performed to validate our method. A collection of 32 monogenic Mendelian diseases from OMIM was used to generate synthetic patient profiles. 100 simulated patients were generated for each disorder taking into account the disease's penetrance information, and the target disease was being ranked against all annotated disorders from OMIM. Additionally, since the method relies on human input, we modeled human factors by introducing noise (i. e. irrelevant phenotypes) and imprecision (i. e. parent terms of precise phenotypes) as well. The target diseases were ranked on top in 95% of the optimal cases with an area under curve (AUC) over 0. 99. Meanwhile, our method demonstrated robustness with similar performance in noise and imprecision scenarios with AUCs of 0. 99 and 0. 96 respectively. In summary, our computational method shows great promise to improve clinical diagnosis by implementing a random walk model on HPO. We expect a statistical model such as Monte Carlo sampling will further enhance the method by adjusting scores of the original rankings.

1681W

Virtual progeny analytics providesa multidimensional view of phenotype expectationsfor couples prior to conception. *A. J. Silver, M. J. Silver, J. L. Larson, N. Delaney, R. Lim, L. M. Silver.* Genetics, GenePeeks, New York, NY.

Purpose

Genetic test results are typically limited to two possible outcomes: positive or negative. This binary choice may work well in clinical diagnostics, but reveals itself to be a rigid and uncompromising framework in reproductive risk assessment. Variants differ in their effect on gene function and expression, and certain manifestations can result in an altered phenotype in some allelic combinations but not others. By treating variant pathogenicity as an all-or-none phenomenon, the impact of allelic heterogeneity on disease severity, penetrance and expressivity is lost. To capture this biological reality, we developed a generalizable methodology for predicting the likelihood of heritable traits in a couple's potential offspring.

Methods

Virtual Progeny Analytics (VPA) simulates genetic recombination to produce "virtual gamete" sequences that are combined pairwise to create a pool of novel diploid genomes - "virtual progeny" (VP). A neural network-inspired Variant-specific Gene Dysfunction (VGD) score, combining all informative quantitative and qualitative metrics of variant impact on protein function, was computed for each variant in the haploid contribution to the VP. Important nodes include large population-based allele and genotype frequencies and measures of "clinical visibility. " By interpreting the VP at the level of a particular gene and genotype, rather than at the level of individual variants, we calibrate specific gene dysfunction to a gene's acceptance of general variation. We then sum together the functionality of the two gene copies to determine if the total gene product is sufficient to avoid a disease state.

Discussion

The distribution across the entire VP pool provides an estimate of the multidimensional probability space of progeny phenotype expectations for any two people defined solely by their genome sequences. VPA can operate easily on clinically identified genotype-phenotype correlations and is applicable to any combination of simple and complex genetic traits.

Conclusion

By focusing on gene dysfunction in the context of potential offspring genomes, the unrealistic imperative of a simplistic binary classification system is eliminated. The impact of this alternative perspective on reproductive risk rises as VGD scores retreat from the two extremes. Together, a large pool of virtual progeny provides an opportunity, preconception, for two prospective parents to obtain a nuanced view of their joint reproductive disease risk.

1682T

Geneious R9: CRISPR design tool. *C. Olsen[1], H. Shearman[2], K. Qadri[1], R. Moir[2], M. Kearse[2], S. Markowitz[2], J. Kuhn[2], S. Dunn[2], T. Paulin[2], A. Cooper[2].* 1) Biomatters Inc. , San Francisco, CA; 2) Biomatters Ltd. , Auckland, New Zealand.

The CRISPR/Cas9 system, a pioneering genome-editing technique, allows for the manipulation of genes and enables researchers to examine the consequence of sequence modifications in a precise and repeatable manner. The CRISPR tool in Geneious R9 will identify potential CRISPR target sites, including paired nickase sites, in a selected sequence and assess the target sites based on off-target interactions. Identified CRISPR sites are annotated on the target sequence using innovative heatmap-style annotations.

Each CRISPR site in the selected sequence is scored according to how many off-target sites it potentially will bind to and how similar the off-target sites are to the original sequence. Scores are calculated according to the method developed by the Zhang lab at the Massachusetts Institute of Technology1. CRISPR sites are displayed as an annotation track on your original sequence and the annotations are colored according to the corresponding score to enable you to see the best sites at a glance.

CRISPR sites need to be targeted to the gene of interest and not cut elsewhere in the genome; however, many other design tools provide a small number of choices for the target genomes. In Geneious, the researcher creates their own off-target database by simply creating a folder within Geneious and placing sequences within this folder. Any sequence may be placed inside this folder, including reference genome sequences downloaded from external databases and/or proprietary sequences created in-house. This allows researchers to use genomes from their organism of interest or from closely related organisms. The researcher then specifies the folder in the CRISPR options.

As this tool is contained within Geneious, researchers can leverage all of the other functionality in Geneious, including creating and sharing workflows for CRISPR design standard operating procedures. This poster aims to demonstrate the functionality of the CRISPR tool, including the benefits of providing the researcher the control to create their own off-target database.

References

[1] Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013 Nov;8(11):2281-308. doi: 10. 1038/nprot. 2013. 143. Epub 2013 Oct 24.

1683F

Apollo: A Production Tested, Vertically Integrated, Operations Enhanced, Science Aware Framework for Launching Large Cohorts of Genomics Pipelines. *R. Z. Castellanos[1], Y. Wang[1], Z. Giles[2], H. Shah[1], E. Schadt[1].* 1) Department of Genetics and Genomic Sciences, Icahn, Mount Sinai School of Medicine, New York, NY; 2) Scientific Computing, Icahn School of Medicine at Mount Sinai, New York, NY.

Apollo is a Platform As A Service (PAAS) that handles dependency resolution, workflow management, visualization, reproducibility, error reporting, error resolution and command-line integration while addressing the plumbing associated with heterogeneous genomic analyses workflows run in a High Performance Computing (HPC) environment.

Apollo integrates the running, tracking, managing and metadata of the Illumina Demultiplexing, GATK Whole Exome or Whole Genome or custom captures, RNA-Seq, and ChIP-Seqχ pipelines into a single framework. Apollo also integrates with SMRT Portal to help manage PacBio data flows. All pipelines are tracked in an online interface which can be configured to send notifications to clients if Service Level Agreements (SLAs) are violated.

Clients can take advantage of Apollo's Standard Development Kit (SDK) to add their own pipelines. This allows investigators to spend less time worrying about operational details and environmental overhead and more time on their actual research. A Service Desk provides a web based entry point for clients to submit data processing requests. "Recipes" provide a concise JSON specification for configuring analyses on cohorts of samples. Strict versioning and provenance facilitate experimental reproducibility. In addition to command line and online access, many of Apollo's services can be accessed via RESTful APIs.

Apollo's backend is tightly integrated with the JIRA project management suite. Each run has it's own JIRA issue. If a given task - human or machine - is not completed *correctly* a JIRA issue is automatically assigned to the appropriate party. Apollo and its many services are self-aware.

System-wide issues as well as local exceptions raise automatic alerts that notify engineers. These service alerts are tracked on Apollo's Services Page and on JIRA.

Apollo's software lifecycle employs industry standard best practices including strict versioning, source control via git, and continuous integration. Nightly regression tests are performed on all pipelines. Changes to the codebase are not permitted until all unit tests have passed.

The result is more correct results, faster turnaround, fewer man-hours, and better science. As of June 11, 2015 the framework has 1. 2 petabytes under management and has been used to process 26,618 samples.

1684W

A virtuous cycle of large cohort research, personal genome analysis, and clinical deployment. *N. Veeraraghavan[1], A. English[1], S. White[1], T. Chiang[1], Z. Khan[1], A. Hawes[1], S. Ambreth[1], J. Farek[1], A. Carroll[2], S. Ma[2], W. Salerno[1], K. Andrews[1], D. Muzny[1], E. Boerwinkle[1,3], R. Gibbs[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) DNAnexus, Mountain View, CA, USA; 3) Human Genetic Center, University of Texas Health Science Center, Houston, TX 77030, USA.

Large common disease studies such as the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium and Mendelian disease gene discovery and diagnostic efforts like the Baylor College of Medicine's Whole Genome Laboratory (BCM-WGL) and Center for Mendelian Genetics (CMG) are transitioning from targeted or whole-exome sequencing to whole-genome sequencing. Whole genome sequence data is very large in size and complex because of a paucity of non-genic annotation. We have developed, deployed and validated an integrated data processing and analysis pipeline, Mercury, for whole genome sequence data. The features of the pipeline include:(a) incorporating multiple protocols for whole exome and genome sequencing with particular optimization for the Illumina X Ten (b) expanding annotation capabilities to include regulatory information from Encode, GTEx, promoter and enhancer sites from FANTOM5 and deleteriousness scores from CADD, RegulomeDB, and Funseq2; (c) tackling discovery of structural variants (SV) under the Parliament framework, a multi-source consensus SV discovery and evaluation tool that automates assembly-based force calling and can incorporate long-read data via PBSuite; (d) developing a NoSQL-based data warehouse to integrate annotation, phenotype, and variant information, with capabilities to provide up-to-date annotations and an "N+1" pVCF solution for large cohorts studies; (e) porting analysis tools and protocols to the DNAnexus cloud platform, with emphasis on security, accessibility, trackability, data integrity, validation and reproducibility, towards compliance under CAP/CLIA guidelines. Efforts have been invested in developing a high-quality personal genome, HS-1011, for calibrating new features as they arise. This reference genome has been characterized across multiple sequencing platforms, including aCGH, Illumina, Pacific Biosciences, Hi-C, and BioNano Irys. These data have been subject to deep variant calling, trio analysis, whole genome assembly, long-read structural variant analysis, and Mendelian prioritization, generating high-quality variants against which novel tools, methods and protocols can be automatically and objectively evaluated, and then transplanted into a diagnostic environment. Our model of development, validation, and deployment delivers comprehensive genomic data to signout teams in a scalable and secure manner, and provides a mechanism for clinical findings to re-inform research protocols.

1685T

Multiple Testing Correction in Linear Mixed Models. *J. W Joo[1], F. Hormozdiari[2], B. Han[3,4], E. Eskin[5].* 1) Bioinformatics IDP, UCLA, Los Angeles, CA; 2) Computer Science Department,, UCLA, Los Angeles, CA; 3) Asan Institute for Life Sciences, Asan Medical Center, Seoul 138-736, Republic of Korea; 4) Department of Medicine, University of Ulsan College of Medicine, Seoul 138-736, Republic of Korea; 5) Department of Human Genetics, University of California, Los Angeles, CA, USA.

Multiple hypothesis testing is a major issue in genome-wide association studies (GWAS) that often analyze millions of markers. The permutation test is considered as the gold standard in multiple testing correction, which accurately takes into account the correlation structure of the genome. Recently, linear mixed model (LMM) has become the standard practice in GWAS, effectively addressing two important issues in GWAS: population structure and low power. However, the current approaches for multiple testing, such as the permutation test as well as many alternative approaches that aim to increase efficiency ofpermutation test, are not applicable to LMM. This is because all these approaches are based on the assumption that uniformly permuting phenotypes will generate the null distribution,which is not the case in LMM. By simulations, we show that permutation leads to inaccurate multiple testing correction when heritability is non-zero. In this paper, we first set up the gold standard approach for multiple testing correction for LMM, called parametric bootstrapping,which is the equivalent of the permutation test for LMM. This approach is accurate, but is computationally expensive because of the large number of resampling. We then propose an efficient and accurate multiple testing correction method for LMM, MultiTrans. The key idea behind MultiTrans is to take advantage that the null LMM statistics asymptotically follow a multivariate normal distribution,whose covariance matrix is obtained by transforming the genotype data based on the kinship matrix. Utilizing this approach, we were able to estimate per-marker thresholds as accurately as the gold standard approach in real and simulated datasets, while reducingrequired time from monthes to hours. We applied our approach to mouse, yeast, and human datasets to demonstrate the accuracy and efficiency of our approach.

1686F

Fast genome-wide assessment of the likelihood for variants being causal using only summary statistics. *D. Lee, T. Bigdeli, B. Riley, A. Fanous, S. Bacanu.* Department of Psychiatry, Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, Virginia, USA.

Genome-wide association studies (GWAS) have successfully identified thousands of genomic loci associated with human traits/diseases. However, these GWAS-detected loci typically possess hundreds of significant single nucleotide polymorphism (SNP) signals in strong linkage disequilibrium (LD). This makes it challenging to separate the true causal variants from significant but non-causal "LD buddies". Recently, two fine mapping methods (CAVIAR and CAVIARBF), which use summary statistics as input, were shown to outperform previous tools. These methods, while useful, have the limitations of: i) high computational burden due to the greedy-algorithm-based SNP selection procedure, ii), in their current form, the possibility of introducing false positive results when applied to multi-ethnic cohorts and iii) potential power loss due to the assumption of all causal variants being genotyped. To overcome these limitations, we propose a quasi-causality test (QCAT) for assessing the likelihood of measured/unmeasured genetic variants being causal. It uses as input only summary association statistics of measured markers and LD structures obtained from large reference population, e. g. 1000 Genomes data. QCAT is based on a simple but accurate regression test of (suitably transformed) association statistics for measured SNPs on (transformed) genotype correlations between the SNPs and a putative causal SNP. Unlike existing methods, QCAT allows i) fast and computationally efficient testing for causality of all SNPs in a particular genomic region or across the whole genome, ii) direct application to mixed ethnicity cohorts while controlling false positives and iii) capability of testing for causality SNPs not measured in the study but measured in the reference population. Using simulated null data sets with different ethnic composition, we demonstrate that, unlike the two greedy-algorithm-based methods, the proposed method controls the false positive rates for multi-ethnic cohorts. We also use alternative data sets simulated under different causality scenarios to show that QCAT provides comparable performance to the two methods for a much smaller computational burden. For more empirical assessment, we apply QCAT to summary data from the Psychiatric Genomic Consortium Schizophrenia Phase 2 and the Tobacco and Genetics Consortium for smoking quantity and present some interesting findings obtained from the results.

1687W

ZGP: a novel method for detecting candidate genes and pathways by integrating genome-wide association studies with pathway analysis. *H. Zhao[1,2], Y. Yang[3].* 1) Institute of Health and Biomedical Innovation, Queensland University of Technology, Queensland, Australia; 2) Queensland Institute of Medical Research, Brisbane, Queensland, Australia; 3) Institute for Glycomics, Griffith University.

ZGP: a novel method for detecting candidate genes and pathways by integrating genome-wide association studies with pathway analysisBackground The successes of GWA studies have provided the opportunities for detecting genetic signals beyond single markers through exploring combined effects of markers on complex diseases and traits. Recently, introduction of integrating the GWA tests with pre-knowledge on genes, such as pathways, have shown promising results to identify candidate genes associated with complex diseases. **Methods**Here we presented a new method, ZGP, to identify candidate genes for complex diseases according to gene-level GWAS results and pathway analysis results. ZGP exhaustingly searches the pathways enriched with genes having low P-values, and then detects the candidate genes. ZGP only requires GWAS summary statistics as input. **Results**We first tested the reliability of ZGP in assessing pathways by type I error rate, and compared ZGP with other available pathway analysis methods. The results show that the type I error rates of ZGP are lower than other methods across different databases. ZGP were compared with GSA-SNP and dmGWAS in detecting candidate genes when applying them to schizophrenia (SCZ) and Height GWAS databases by employing the reported genes as positive dataset. ZGP was demonstrated as an effective approach in detecting candidate genes through comparing the ROC curves with other pathway analysis approaches. **Conclusion**This study demonstrated ZGP as a pathway analysis approach with low type I error rate. This study indicated that ZGP has increased the potential for detecting candidate genes.

1688T

OpenCB: a scalable and high-performance platform for big data analysis and visualization in genomics. *J. Coll[2], CY. Gonzalez[3], J. Tarraga[4], M. Haimel[7], FJ. Lopez[3], JM. Mut[3], A. Rueda[2], R. Sultana[2], F. Salavert[4,5], M. Bleda[7], L. Garcia-Alonso[3], J. Dopazo[4,5,6], S. Grät[7,8], J. Paschall[3], P. Calleja[1], A. Rendon[2], I. Medina[1].* 1) High Performance Computing Service, University of Cambridge, Cambridge, United Kingdom; 2) Genomics England, Queen Mary University of London, London, United Kingdom; 3) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge, United Kingdom; 4) Computational Genomics Department, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain; 5) Bioinformatics of Rare Diseases (BIER), CIBER de Enfermedades Raras (CIBERER), Valencia, Spain; 6) Functional Genomics Node, (INB) at CIPF, Valencia, Spain; 7) Department of Medicine, University of Cambridge, Cambridge, United Kingdom; 8) Department of Haematology, University of Cambridge, Cambridge, United Kingdom.

Current technologies in Genomics are producing data at an unprecedented scale, with projects producing tens of terabytes per day. Existing bioinformatics solutions have not been designed to work at this scale, hindering the ability to store, analyze, or visualize data. OpenCB is an open-source, collaborative platform that hosts different projects designed to provide a scalable and high-performance solution for storing, processing, analyzing, sharing and visualizing big data in genomics. To achieve this, OpenCB uses modern computing technologies such as Hadoop and Spark for data processing and analysis; NoSQL databases (MongoDB, HBase) for data indexing; or HTML5 for data visualization. The knowledge-base database for all OpenCB projects is CellBase, a NoSQL database that integrates the most relevant biological information about genomic features and proteins, gene expression, regulation, functional annotation, genomic variation and systems biology information. Content relies on repositories such as ENSEMBL, Uniprot, and ClinVar, also supporting local knowledge. CellBase has variant annotation capabilities compatible with Ensembl VEP. All data is available through a command line or by a RESTful web service. Cellbase can annotate thousands of variants per second using a remote server. OpenCGA provides a scalable and high-performance solution for big data analysis. OpenCGA implements different components: 1) a *storage engine* framework to store and index alignments and genomic variants into different NoSQL such as MongoDB or Hadoop HBase; 2) a Metadata Catalog which keeps track of users, projects, files, samples, jobs and cohorts, also providing authentication and authorization capabilities; 3) an analysis engine to execute genomic analysis in a traditional HPC cluster or in Hadoop. OpenCGA implements a command line and RESTful web services to manage and query all data. Genome Maps provides a high-performance HTML5 SVG-based genome browser to interactively display data held by CellBase and OpenCGA. Users can also extend Genome Maps to display their own data and formats. OpenCB projects are compliant with the new GA4GH data models. OpenCB is used by projects such as ICGC, NIHR Bioresource-Rare Diseases (BRIDGE), EMBL-EBI EVA and Genomics England. OpenCB is actively developed by researchers and data analysts at University of Cambridge, CIPF, BRIDGE, EMBL-EBI and Genomics England. Source code is open and it is freely available at https://github. com/opencb. .

1689F

Global expression patterns and key regulators in epithelial to mesenchymal transition. *P. Parsana[1], S. Amend[2], K. Pienta[2], A. Battle[1].* 1) Department of Computer Science, Johns Hopkins University, Baltimore, MD; 2) Department of Urology, The James Buchanan Brady Urological Institute, Johns Hopkins University, Baltimore, MD, USA.

Epithelial to mesenchymal transition (EMT) is the process where epithelial cells transdifferentiate to mesenchymal cells with increased motility. EMT is integral in developmental biology, wound healing and stem cell behavior. However, in cancer it leads to acquisition of migratory and invasive properties. While accounting for batch effects and noise inherent in gene expression data, we integrated multiple EMT studies from diverse tissue types and used probabilistic graphical models (PGM) to identify the key regulatory mechanisms and interactions that are altered during EMT. We used data from 15 published gene expression microarray studies obtained from Gene Expression Omnibus, representing cell lines from 6 different tissues across 95 samples with different modes of induction of EMT. Applying normalization methods to account for technical variability between datasets, we performed differential expression analysis to identify the individual genes and pathways that vary significantly between epithelial and mesenchymal states. Among the differentially expressed genes were CDH1, ZEB1, FXYD3, AP1M2 and numerous others, known epithelial or mesenchymal marker genes. Enrichment results show increased representation of pathways implicated in wound healing, blood coagulation and cell-cell signalling. We compared the cross-tissue patterns with differential expression results from data restricted to particular tissues in order to identify both shared and tissue-specific genes. Further, we integrated additional sources of data using probabilistic graphical models to predict which of the differentially expressed genes are most likely to play key regulatory roles, with causal influencinge on other genes. Here, we incorporated large expression quantitative trait (eQTL) studies including the Depression Gene and Networks study of 922 individuals, identifying gene pairs where a cis-eQTL in one gene is also significantly associated with the other, distant gene (a trans-eQTL), strongly suggesting a causal, directed relationship between the two genes. From these results, we identified genes such as MRC2 affect the expression of multiple other genes differentially expressed in EMT. We are now extending this to incorporate epigenetic data, further refining our network model. Together, our approach will help build an integrative model that will account for genetic, transcriptomic and epigenetic interactions to identify causal genes and regulatory interactions in EMT.

1690W

***In silico* validation of allelic imbalance by assessments of SNP arrays, whole-exome sequencing and haplotype matching.** *S. Sivakumar[1,2], F. A. San Lucas[3], P. Scheet[2,1].* 1) Biostatistics, Bioinformatics and Systems Biology Program, UT Graduate School of Biomedical Sciences, Houston, TX; 2) Department of Epidemiology, UT MD Anderson Cancer Center, Houston, TX; 3) Department of Translational Molecular Pathology, UT MD Anderson Cancer Center, Houston, TX.

Copy number alterations (CNAs) play a critical role in the initiation and progression of cancer and are typically detected by array-based technologies. Stromal contamination or tumor heterogeneity confound detection of these events, rendering it desirable to observe these from a complementary technology. Although whole-exome sequencing (WES) is conducted mainly to detect single-nucleotide variants, it provides an opportunity for validating CNAs. Since subtle forms of allelic imbalance due to CNAs are particularly difficult to resolve with WES, we have implemented a sensitive, haplotype-based algorithm (hapLOHseq) for its detection. Here, we present a procedure for validating CNAs by assessing (1) the positional overlap of the calls, and (2) the directional concordance of the alleles at heterozygous markers (ie. whether a haplotype shows putative loss or gain). This procedure is straightforward when the markers overlap, eg. between whole-genome NGS and SNP array based techniques; we compute the haplotype consistency between the reference allele frequency (RAF) and the "B allele" frequency (BAF) of each overlapping marker. However, the utility of this direct approach is limited by the set of overlapping markers, which may be small (or empty) in the case of WES and SNP arrays, especially if the CNA covers a modest number of captured exomes. We address this limitation by implementing a linkage disequilibrium (LD) based technique to first generate a haplotype map from all the heterozygous sites in the 1000 Genomes Project, followed by a greedy approach of 1:1 mapping to identify maximal surrogate marker pairs between the platforms. Although the surrogate markers are not in complete LD, they dramatically improve marker availability for cross-platform comparisons. As a proof of concept, we tested our method on a pancreatic adenocarcinoma patient in the The Cancer Genome Atlas (TCGA) showing well distributed putative AI events across the Affymetrix 6. 0 SNP array and WES platforms. We obtained 1452 overlapping CNA markers, 81. 6% of which showed directional (haplotype) concordance. We additionally identified a total of 6197 surrogate pairs (cf. 1452), 76. 8% of which showed directional concordance, not far from the trend observed from the overlapping markers. These encouraging results support an intuitive, novel framework of *in silico* statistical validation for within-sample between-platform variation.

1691T

Identification of possible pleiotropic genetic variants using ClinVar and hospital discharge data. *A. Donahue, J. Weissert, P. Tonellato.* University of Wisconsin Milwaukee, Milwaukee, WI.

This work describes a novel method linking publically available genomic data with hospital discharge data to uncover pathological pleiotropic variants. When a variant arises in a pleiotropic gene, distinct diseases with different symptoms and clinical therapies can co-occur in a patient, obscuring the diseases' shared genetic origin. Identifying pathological pleiotropic genetic variation has significant clinical implications for directing research towards targeted therapy and impacting clinical guidelines for screening and monitoring. In contrast to other techniques for pleiotropic discovery that require extensive study design and resources, we hypothesized that pathological pleiotropic genetic variants can be identified using existing genomic and phenotypic information in the National Center of Bioinformation Technology (NCBI) database ClinVar. We mined the ClinVar database for specific diseases, colorectal cancer and asthma, both of which have at least one known genetic component and are the source of significant morbidity and mortality in the US population. We found 1,120 variants linked to colorectal cancer risk; 53 variants remained after excluding those provided by a single submitter and those linked only to colorectal cancer. We identified specific variants linking colorectal cancer to other diseases with already established relationships such as Lynch Syndrome and breast cancer. We also uncovered a novel relationship between colorectal cancer and age-related macular degeneration (ARMD), linked by independent studies of the variant *rs4986790* of the *TLR4* gene. Similarly, we found 24 variants linked to asthma, and 1 variant, *rs1042713* of the *ADRB2* gene, was independently associated with both asthma and metabolic syndrome. These results support our hypothesis that ClinVar can yield valuable information on possible pathological pleiotropic variants. Further, we matched ICD-9 codes with diseases found to be associated in ClinVar. We then proceeded to test for epidemiological relationships between the associated ICD-9 codes by mining nationwide hospital discharge data. Preliminary results confirm an increase in relative risk for diseases genetically linked in ClinVar across a broad population and suggests a method for validating the pleiotropic relationships found in ClinVar. Further work is being implemented to systematically search ClinVar for pleiotropic variants and cross-validate them with ICD-9 codes in hospital discharge data.

1692F

GRIPT: A novel method to identify disease causing genes in a cohort of patients using next generation sequencing data. *L. Zhao[1], X. Wang[2], Z. Ge[3,4], Y. Chen[5], F. Wang[5], R. Chen[1,3,4].* 1) Structural and Computational Biology and Molecular Biophysics Program, Baylor College of Medicine, Houston, TX; 2) Medical Genetics Laboratories, Baylor College of Medicine; 3) Department of Molecular and Human Genetics, Baylor College of Medicine; 4) Human Genome Sequencing Center, Baylor College of Medicine; 5) Department of Computer Science, Fudan University, Shanghai, China.

It has been extensively reported that disease-causing genes can be successfully identified in large families. However, finding genetic causes underlying sporadic cases is still challenging. With the rapid development in the next generation sequencing (NGS) technology, many large disease consortiums have been established. Therefore an algorithm to systematically prioritize and identify disease-genes from large numbers of sporadic cases is highly demanding. In this study, we developed a novel method called**GRIPT** (GeneRanking,Identification andPredictionTool) to evaluate the load of pathogenic mutations in each gene and also test the mutation load differences of each gene between cases and controls with NGS data, from which causative genes are prioritized. To evaluate the performance of our method, simulation was applied and our methods showed both high sensitivity and specificity comparing to other methods. Moreover, we applied our method on a cohort of patients with retinal degenerative disease with control individuals whose phenotypes are retinal unrelated. Both the patients and the control samples were genotyped by whole-exome sequencing (WES), and mutations were called and annotated in each individual using in-house bioinformatics pipeline. A list of genes ranked by p-value was provided. Known retinal disease genes have high ranks in this list and several novel candidate genes were also identified by this method for future research. Overall, our method can effectively and robustly prioritize and identify disease causing genes for sporadic cases using case-control NGS data.

1693W

SNP-Chip Sibling Ranker: Using SNPChip Data to Inform Next Gen Sequencing Choices. *B. N. Pusey, W. A. Gahl, D. R. Adams.* Undiagnosed Diseases Program, National Human Genome Research Institute, Bethesda, MD.

In 2008, the National Institutes of Health (NIH) began the NIH Undiagnosed Diseases Program (UDP), whose purpose is to find diagnoses and conduct research for study participants who remain undiagnosed despite an extensive medical workup. Clinical exome sequencing has rapidly become incorporated into the standard clinical armamentarium for evaluation complex patients. In an effort to address the needs of patients who remain undiagnosed, even after exome sequencing has been performed, the UDP continually seeks out new techniques to maximize diagnostic return. One staple strategy is to increase the number of family members included in the analysis. Adding additional, informative family members allows segregation filtering to remove a greater number of segregation-inconsistent variants. Sequencing the parents of a proband is particularly effective for the analysis of apparent de novo variants. Recessive variants can be filtered with the addition of affected and unaffected siblings. In practice, a balance must be struck between the cost of collecting, sequencing and analyzing family DNA samples compared with the final benefit to the exome analysis. Rational decision-making around these issues should include a quantitative assessment of the potential benefits of adding the additional family members. SNP chip genotyping offers one approach to estimating the benefit of sequencing additional family members. The SNP chip data can be used to map recombination events the occurred during recombination and segregation of chromosome into the prospective sibling. We present a framework for using this type of recombination mapping to estimate the potential benefit of any given sibling to a complete exome analysis. Our technique has the potential to avoid the addition of low-contribution siblings to an exome or genome sequencing project while maintaining maximum Mendelian filtration potential. This research is supported by the Intramural Research Program and Common Fund of the National Institutes of Health.

1694T

miRge – A rational, ultrafast, multiplexed method of processing small RNA-seq data for microRNAs. *A. S. Baras[1], C. J. Mitchell[2], T. C. Cornish[1], A. Pandey[1,2], M. K. Halushka[1].* 1) Pathology, Johns Hopkins University SOM, Baltimore, MD; 2) Institute of Genetic Medicine, Johns Hopkins University SOM, Baltimore, MD.

Small RNA RNA-seq for microRNAs (miRNAs) is a rapidly developing field in which there are still opportunities to create faster, better bioinformatics tools to parse these large datasets. We built miRge, to be the first rational, ultrafast, small RNA-seq solution to process samples in a highly multiplexed fashion. It utilizes a unique hierarchical multistep processing of small RNA-seq data. From a table of unique sequences and counts, a sequential alignment is made against customized mature miRNA, hairpin miRNA, noncoding RNA and mRNA sequence libraries. Customizations are performed to include SNPs and reduce alignment error. Output is given for miRNAs as raw reads and reads per million (RPM) in addition to isomiR data and reads for all other RNA species (tRNA, rRNA, snoRNA, mRNA). In head-to-head comparisons for single sample analyses with miRExpress 2. 0, sRNAbench, omiRAs and miRDeep2, miRge was 3 to 28-fold faster and always aligned more miRNA reads per sample. In a single multiplex experiment, we fully annotated all RNA species in 100 separate samples in 52 minutes. miRge accepts both fastq and fastq. gz files, can remove any adapter type, and can be used with any species. These features make miRge an ideal tool for high and low-throughput users alike. It is a significant improvement in processing speed and quality of alignment to known miRNAs. miRge is freely available at http://atlas. pathology. jhu. edu/baras/miRge. html.

1695F

Nonparametric tests for differential expression in paired tumor-normal microRNA data with imputation-induced dependence. *A. A Suyundikov[1,2], J. R. Stevens[2], C. Corcoran[2], J. Herrick[3], R. K. Wolff[3], M. L. Slattery[3].* 1) BioStat Solutions, Inc. , 5280 Corporate Drive, Suite C200; Frederick, MD 21703; 2) Department of Mathematics and Statistics, Utah State University, 3900 Old Main Hill, 84322-3900, Logan, UT; 3) Division of Epidemiology, Department of Internal Medicine, University of Utah School of Medicine, 383 Colorow Road, Salt Lake City, UT 84108, U. S. A.

Background: Imputed microRNA (miRNA) data are often correlated and may have imputation-induced dependence among subjects. The imputed miRNA expressions may also be non-normally distributed. In this respect, differential expression tests that do not assume a specific distribution but account for the dependence structure of the data are preferable. Currently applied nonparametric methods either do not consider the correlation structure of data or are not computationally feasible for high-dimensional miRNA data with imputation-induced dependence. *Methods:* To identify differentially expressed miRNAs in paired tumor-normal colorectal cancer data with imputation-induced dependence, we proposed three nonparametric methods: the computationally efficient nonparametric permutation t-test accounting for dependence (Permutation dependent), the signed-rank test accounting for dependence (SRT dependent), and the Wilcoxon signed-rank test after whitening transformation of data (Whitening). *Results:* The performances of these proposed methods over other nonparametric methods were evaluated using simulated and real data sets. The differential expression tests show that the Permutation dependent and the SRT dependent methods have moderate power and generally control the False Discovery Rate (FDR), while Whitening has better power but fails to control the FDR. The Permutation independent and the SRT independent methods, which ignore the dependence structure of data, showed higher power than the other nonparametric methods, but do not maintain control of the FDR. *Conclusion:* When miRNA data are not distributed normally and have imputation-induced dependence among subjects, nonparametric methods accounting for dependence should be used over conventional nonparametric methods that ignore the imputation-induced dependence.

1696W

SPARS - Sequencing-based pipeline for annotating novel small non-coding RNAs. *P. P. Kuksa[1,2], Y. Y. Leung[1,2], A. Amlie-Wolf[3], B. D. Gregory[3,4], L. -S. Wang[1,2,3].* 1) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Penn Institute for Biomedical Informatics, University of Pennsylvania Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Biology, University of Pennsylvania, Philadelphia, PA.

The human genome contains thousands of small non-coding RNAs (sncRNAs, <200nts), but most studies have focused on miRNAs or classes involved in splicing. Many new sncRNA classes have recently been uncovered, and they perform versatile roles across different cell states and types. Inferring their regulatory roles remains challenging due to a lack of comprehensive annotation and understanding of sncRNA transcription. To facilitate the simultaneous characterization and functional annotation of multiple types of sncRNAs, we integrate naming and sequence information from various independent annotation sources: miRBase v19 for miRNAs, UCSC for ribosomal RNAs, small cytoplasmic, nuclear and nucleolar RNAs, tRNA-scan for tRNAs and NCBI for piwi-RNAs. We also included tRNA fragments, the flanking 50bp of tRNAs. As sncRNAs are processed specifically, reflected by their stereotypical lengths, we combined different biologically meaningful features with a novel discovery-oriented smRNA-seq data processing pipeline. This allows us to quantify known and predict novel sncRNAs in human across many tissues and cell types. To systematically process smRNA-seq data generated by Illumina, we first performed sample level quality control after trimming the correct adapter. As a size selection step was performed during library preparation, we next examined the distribution of the read lengths, as this reflects whether the gel was cut correctly and allows our pipeline to capture sncRNAs of different lengths. After this, we performed local read alignment using STAR and allowed for an unlimited number of multi-mapping sites, enabling us to capture sncRNAs with high sequence similarity. Incorporating multi-reads increased the sensitivity of detecting different sncRNAs: compared with uniquely-mapped data, we are able to quantify 1. 5-2 times more human sncRNAs in eight different classes using smRNA-seq data obtained from 42 tissues across 30 different resources. Next, we computed confidence levels to identify novel sncRNA annotations based on the patterns of mapped reads and the specificity of 5' and 3' sncRNA processing. Using this novel annotation approach, we are able to recapitulate the majority of known microRNAs (76%) and small nucleolar RNAs (87%) and discover over 2,900 novel, previously unannotated mature small RNA products within promoter (187), exonic (599), intronic (1,091) and intergenic (621) regions. This flexible pipeline can be used for smRNA-seq data in any organisms.

1697T

Contrasting Association Results between Existing PheWAS Phenotype Definition Methods and Five Validated Electronic Phenotypes. *J. B. Leader[1], S. A. Pendergrass[2], A. Verma[5], D. J. Carey[3], D. N. Hartzel[1], D. H. Ledbetter[4], M. D. Ritchie[2,5], H. L. Kirchner[1].* 1) Center for Health Research, Geisinger Health System, Danville, PA; 2) Biomedical and Translational Informatics, Geisinger Health System, Danville, PA; 3) Weis Center for Research, Geisinger Health System, Danville, PA; 4) Geisinger Health System, Danville, PA; 5) Center for Systems Genomics, The Pennsylvania State University, University Park, PA.

Phenome-wide association studies (PheWAS) generally use various counts of International Classification of Diseases, Ninth Revision (ICD9) codes to identify case/control status for diseases that are used as comprehensive phenotypic variables. Defining high quality algorithms for phenotype extraction (gold standard) is a laborious task, thus is not a viable solution for PheWAS. However, differences between using gold standard phenotypes and various counts of ICD9 codes for defining phenotypes have not been thoroughly examined. The most common methods of defining case/control status using ICD9 codes include grouping the codes by three digits (e. g. 250. 02 rolls up to 250), five digits (e. g. 250. 02 stays 250. 02), and "PheWAS Codes" (e. g. 250. 02 grouped to 250. 2). Using data from the Geisinger Health System electronic health record (EHR) and MyCodeTM Biorepository, we identified ICD9 code based case/control status using these methods and compared the resultant cases and controls to the results of gold standard EHR phenotypes for type 2 diabetes (T2DM), acute coronary syndrome (ACS), non-obstructive coronary artery disease (CAD), obstructive CAD, and obesity. Finally, using ~5,000,000 genetic variants with minor allele frequency > 0. 01, we performed a genome wide association (GWAS) for each gold standard phenotype and contrasted highly significant results to those from PheWAS using various ICD9 code based case/control definitions. Comparing these methods of defining phenotypes has shown thus far that for certain diseases using ICD9 count based methods agrees with gold standard algorithms while other phenotypes cannot use codes alone and require additional data from the EHR. For example, good agreement was observed for T2DM, whereas ACS exhibited poor agreement due to the lack of specificity of PheWAS codes. Our future directions include comparing the positive and negative predictive value of the various ICD9 code based methods, potentially identifying additional methods that can broadly improve phenotype definitions. We will also compare additional gold standard phenotypes that can be deployed using EHR data to solely ICD9 based approaches to further identify disease areas that have a high degree of agreement between various approaches. As ICD9 codes will soon migrate to ICD10 codes, establishing improved principles and use of ICD or other ontological code systems (i. e. SNOMED) will be of critical importance for use in PheWAS for discovery now and in the future.

1698F

Text-mined phenotype annotation and vector-based similarity improve automated classification of Mendelian phenotypes. *J. R. Saklatvala, M. A. Simpson.* King's College London School of Medicine, London, United Kingdom.

The discovery of novel causative mutations in monogenic diseases typically requires analysis of variant profiles of multiple individuals with the same disorder. The falling cost and increased abundance of genomic data creates a bottleneck to efficiently identify patients with similar phenotypes. Systematic methods have recently been developed to calculate similarity between phenotypes – these involve the annotation of phenotypes with terms from clinical ontologies (such as HPO/SNOMED CT) and using ontology structure to infer similarity between phenotypes. Such phenotypic similarity is dependent on the process of annotating vast numbers of phenotypes in addition to the clinical ontologies being well defined and structured, which may not hold true throughout all areas in an ontology. It also requires an in-depth knowledge of the particular ontology used by the similarity measure program. Here we investigate annotation of free-text phenotypic descriptions using text mining, with a view to making phenotype annotation more efficient and automated. We calculate similarity between annotated phenotypes using vector space models, a flexible method that is well suited to text-mined annotations. We evaluate the performance of this approach through annotation and calculation of the similarity between OMIM and Orphanet phenotypes. Similarity methods were assessed based on their ability to group known similar phenotypes closely, using OMIM phenotypic series as the set of known similar phenotypes. Vector-based similarity greatly improved the identification of similar phenotypes when annotated through a text mining approach and provided equivalent performance when applied to manual annotations. Our study demonstrates that a combination of text mining and vector-based similarity is an effective combination to group related phenotypes within this dataset.

1699W

A Coalescent-Based Shotgun Sequence Simulator for Evolving Microbial and Tumour Cell Populations. *K. Liao[1], W. Hsieh[1], W. Hon[2], C. Tang[1,3].* 1) Department of Computer Science, National Tsing-Hua University, Hsinchu City, Taiwan; 2) Institute of Statistics, National Tsing-Hua University, Hsinchu City, Taiwan; 3) Department of Computer Science and Information Engineering, Province University, Taichung City Taiwan.

High-throughput sequencing technology has revolutionized the study of metagenomics and cancer evolution. For relatively simple environment, microbial community is dominated by a few species with sufficiently high coverage. By analyzing alignment of reads from these species, SNPs can be discovered and evolutionary history of the populations in the species can be reconstructed. Similar situation also arise in the study of cancer evolution in which the sets of reads represent random sample of individual DNA molecules from tumour tissues. Subclonal structure and relationship between subclones can be revealed by analyzing somatic mutations. The ever-increasing read length will allow more detailed analysis about evolutionary history of microbial or tumour cell population. A simulator of shotgun sequence from such populations will be helpful in the development or evaluation of analysis algorithms. In this study, we described an efficient algorithm, MetaSMC, which simulates reads from evolving microbial or tumour cell population. Based on coalescent theory, our simulator supports all evolutionary scenarios supported by other coalescent simulators. Our simulator also supports wide range of substitution models including Jukes-Cantor Model, HKY85 and Generalised Time-Reversible (GTR) model. Our simulator supports mutator phenotypes by allowing different mutation rate and subsitution model in different subpopulations. By ignoring unnecessary chromosomal segments, our algorithm is far faster than existing coalescent simulators including Hudson's ms and MaCS. The accuracy of our algorithm is evaluated by time to the most recent common ancestor (TMRCA) and likelihood curve derived from Monte-Carlo integration over large number of possible genealogies. By comparing our algorithm with Hudson's ms, we demonstrated that our algorithm is a good approximation to Standard Coalescent model. We also proved that the process behind our algorithm is Sequentially Markov Coalescent with incomplete sample.

1700T

Accurately inferring imbalanced allele expression using logistic regression models. *K. C. Olney[1], L. Skotte[2], R. Nielsen[3], M. A Wilson Sayres[1,4].* 1) School of Life Sciences, Arizona State University, Tempe, AZ; 2) Bioinformatics Center, Department of Biology, University of Copenhagen, Copenhagen; 3) Departments of Statistics and Integrative Biology, University of California, Berkeley, Berkeley, CA; 4) Center of Evolution and Medicine, The Biodesign Institute, Arizona State University, Tempe, AZ.

Biased allele expression, refers to the imbalanced expression of the two alleles in a diploid genome. Unequal transcription of alleles may occur due to cis-regulatory element variation or allele-specific epigenetic modifications. Allelic imbalance is associated in human diseases such as metabolic disorders and in ovary and breast cancers. However, allelic imbalance may be incorrectly inferred due to technical variation inherent in RNA-Seq data, including read depth, reference mapping bias, and the overdispersion of reads. To correct for technical variation we develop a logistic regression model with a mixed effects approach to combine information regarding biased allele expression from many individuals in a population, and across multiple genes. Simulations show that our method does not suffer from an excess of false-positives when inferring biased allele expression while standard ASE methods (a SNP-wise binomial test and a binomial- based logistic regression) test showed an excess of inflated p-values in the quantile-quantile plots. Further, we conducted additional simulations to predict the power of the method to detect the possible range of biased allele expression under assumptions of variable numbers of SNPs per gene and under variable depth of coverage. We then applied this method to inferring biased allele expression across the genome in 89 lymphoblastoid cell lines samples from a Central European Utah population, and are able to more accurately detect modest degrees of allelic imbalance.

1701F

Online human population genetics simulator: a tool for genetics/genomics education and research. *J. R. Shaffer[1], J. Rogan[2].* 1) Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Computer Science, University of Pittsburgh, Pittsburgh, PA.

The technology-driven advancement of ‘omics sciences and the lure of personalized medicine have generated increased interest in genetics and genomics from diverse health science professionals and society at large. Moreover, given its growing relevance to public health, policy, and medicine, there is an urgent need for greater “genomics literacy” of professionals across disciplines. Because many of the foundational concepts of ‘omics sciences are rooted in population genetics, which has historically remained a highly mathematical discipline, efforts to make these concepts accessible are needed. Toward this end we have developed an online population genetics simulator that enables users to manipulate and experience firsthand the effects of evolutionary forces and phenomena. The online simulator (popgensimulator. pitt. edu) focuses on the phenomena impacting human evolution and was designed as a tool to facilitate genetics education and research. The simulator can model various evolutionary forces (e. g. , mutation, selection, drift, migration, population bottleneck) and mating systems (assortative mating, inbreeding) individually or simultaneously. Graphical output depicts allele (or genotype) frequency changes across successive generations in finite or theoretically-infinite populations. Multiple simulations under the same or different parameters can be plotted together for comparison. Simulations can be run in batches with aggregate summary statistics (e. g. , average time to allele fixation) reported. Due to the extensive scope of the simulator, the user interface was designed to facilitate quick and clear input of desired simulation parameters without overloading new users. Input and output is intuitive. Interactive graphs of simulation results are automatically generated, which allow users to zoom in/out, pan left/right, trace individual simulation results, and save as various image file formats. Graphs can be toggled between two color and contrast themes that are optimized for screen viewing, and projecting/printing, respectively. Numerical simulation results can be outputted. URL links can be generated to pre-populate simulation parameters for fool-proof dissemination. The user interface has been optimized for both computers and mobile devices. Though extremely sophisticated, the user-friendly implementation allows the novice user to quickly master the simulator for exploring and experimenting with the forces and phenomena shaping the human genome.

1702W

Benchmarking of splice isoform quantification methods for RNA sequence data. *F. Aguet[1], D. S. DeLuca[1], T. Sullivan[1], G. Getz[1,2], K. Ardlie[1], The Genotype-Tissue Expression (GTEx) Project Consortium.* 1) Broad Institute, Cambridge, MA; 2) Department of Pathology, Massachusetts General Hospital, Boston, MA.

A majority of human genes, including ~80% of protein coding genes, undergo alternative splicing to produce distinct transcripts. How this process is regulated at a transcriptome-wide scale, and how genetic variants affect the expression of specific splice isoforms remains poorly understood due to challenges in accurately quantifying expression at the level of individual isoforms. Due to the short average length (~200-300bp) of sequenced fragments in a typical RNA-seq experiment, reads corresponding to a specific gene may align equally well to multiple of its isoforms, requiring a probabilistic approach to assign reads to their isoform of origin. Moreover, while RNA-seq can in principle capture the full range of expressed transcripts in a sample by proxy of sequenced fragment counts, several factors limit its accuracy in practice if left uncorrected. These factors include 5' or 3' biases (i. e. , due to a combination of RNA degradation and polyA selection during library preparation) and sequence-specific biases (i. e. , non-random sequence content) at the sites of fragment start/end positions and over the full fragment. Importantly, the accuracy of isoform assignment and bias correction depends on coverage, and consequently, on the read depth of the experiment. Several approaches for the deconvolution of isoform abundances from RNA-seq reads have been proposed (e. g. , Cufflinks, Flux Capacitor, eXpress, RSEM, Salmon, Kallisto), with bias correction models of varying complexity. Comparisons between different methods often produce diverging results, and objective evaluations are hindered by the absence of ground truth data. RNA-seq data from over 40 human tissues generated by the GTEx Project Consortium provide a unique opportunity to assess alternative splicing across the human transcriptome and to identify quantitative trait loci that modulate alternative splicing. To this end, high confidence in the estimates of transcript isoform abundance is critical. We present a rigorous framework for the evaluation of isoform quantification methods, based on simulations (which reproduce biases observed in GTEx data), replicate experiments, partial ground truth derived from a subset of long read (2x250bp) data from GTEx, and measures of isoform complexity, and apply this framework to assess the strengths and weaknesses of the aforementioned methods. We expect these results to serve as a resource for future improvements to isoform deconvolution methods for RNA-seq.

1703T

Accurate quantification of allele-specific expression from single cell RNA-Seq data. *K. Choi, N. Raghupathy, S. Munger, M. Vincent, G. Beane, A. Simons, G. Churchill.* Gary Churchill's lab, The Jackson Laboratory, Bar Harbor, ME.

In diploid cells, alleles can be genetically and even functionally different from each other, and their individual expression can vary significantly depending on genetic and cellular conditions. With the emergence of high-throughput sequencing (HTS) technologies, we can now resolve how alleles are preferentially expressed during a biological process of interest; indeed, recent HTS studies have shown imbalances in allelic expression to be common. This information not only reveals which allele is responsible for a phenotype, but enables us to decode how gene expression is regulated. This insight is fundamental for understanding the genetic architecture underlying normal phenotypic diversity and disease. The application of single-cell HTS methods to the study of allelic imbalance adds valuable information about the dynamics of allele expression. As each cell is unique in terms of its state and/or environment, individual cells may exhibit distinct allelic imbalances that are obscured at the level of the tissue. HTS technologies produce millions of short DNA or mRNA sequence reads, however it is often difficult to pinpoint the alleles from which they originate due to the complex, repetitive structure of genomes and noise in the reads themselves. Naïve approaches that discard obscure reads lack the power to accurately estimate allelic expression. A compounding challenge in single-cell genomics data is the increased variation stemming from transcriptional bursting and dropouts. Capturing and sorting out this variability is key to understanding the genetic regulation and dynamics of transcription. We developed an empirical Bayes model to unravel the true cellular heterogeneity from a seemingly homogeneous cell population and refine our understanding of the regulatory mechanisms underpinning cell function. We minimized data ambiguity and information loss by integrating individual-specific genetic information, e. g. , phased SNPs and indels. The new model also borrows information across a population of individual cells and model the transcriptional states of alleles. By referring to how other individual cells transcribe alleles, the method will enable us to accurately estimate allelic expression imbalance, identify subpopulations, and, thus, provide increased sensitivity to accurately estimate alleles with smaller sampling depths, as it combines information across all alleles in a population.

1704F

***derfinder*: Software for annotation-agnostic RNA-seq differential expression analysis.** *L. Collado-Torres[1,2,3], A. Frazee[1,3], M. Love[4], R. Irizarry[4], A. Jaffe[1,2,3,5], J. Leek[1,3,4].* 1) Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; 2) Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, USA; 3) Center for Computational Biology, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; 4) Department of Biostatistics, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, MA, USA; 5) Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

BackgroundDifferential expression analysis of RNA sequencing (RNA-seq) data typically relies on reconstructing transcripts or counting reads that overlap known features: genes, exons. We previously introduced an intermediate statistical approach called differentially expressed region (DER) finder that seeks to identify contiguous regions of the genome showing differential expression signal at single base resolution that does not rely on existing annotation or potentially inaccurate transcript discovery. ResultsHere we describe a complete suite of software built around the DER approach to RNA-seq analysis. We introduce the *derfinder* package that allows for: (1) a new, computationally efficient approach for base resolution annotation-agnostic analysis called expressed-region analysis, (2) genome-scale analyses in a large number of samples, (3) flexible statistical modeling, including multi-group and time course analyses, and (4) data visualization at base resolution in R. Our suite permits a comprehensive analysis of RNA-seq data at base resolution, from preprocessing, to modeling, to annotation and visualization. We perform a complete comparison to feature counts based methods and demonstrate that base-resolution analysis sacrifices a small amount of power to enable discovery. Finally we apply this approach to public RNA-seq data from the developing human brain and identify differentially expressed regions associated with human brain development. ConclusionsAnnotation-agnostic expressed-region RNA-sequencing analysis provides an alternative to feature-level analysis that allows discovery of differentially expressed regions outside known features. *derfinder* is software designed to identify, visualize, and interpret differentially expressed regions. The package is available from *Bioconductor* at www. bioconductor. org/packages/derfinderPre-printhttp://biorxiv. org/content/early/2015/02/19/015370.

1705W

SplineAdjust: Correcting length bias in differential expressionanalysis of RNA-Seq data. *L. Wang[1], X. Chen[1], C. Armoskus[2], J. Kim[2], J. Li[3], K. Wang[2].* 1) Department of Public Health Sciences, University of Miami Miller School of Medicine, Miami, FL; 2) Zilkha Neurogenetic Institute, Department of Psychiatry and Department of Preventive Medicine, University of Southern California, Los Angeles, CA 90089, USA; 3) Affymetrix Inc. , 3420 Central Expressway, Santa Clara, CA 95051, USA.

RNA-Seq has become a standard methodology for quantifying gene expression levels and detecting differentially expressed (DE) genes. However, technical artifacts may introduce biases into RNA-Seq. Recent studies have shown that DE analysis of RNA-Seq is often prone to gene length bias, that is, longer genes are more likely to be declared significant than shorter genes. When popular software tools such as CuffDiff, DESeq and edgeR are used to calculate DE p-values, substantial length bias often exist, even after normalizing counts by lengths. We proposed a novel statistical approach “SplineAdjust” that takes p-values from any DE analysis tool to generate adjusted p-values for unbiased assessment of DE. SplineAdjust models gene lengths nonlinearly using a penalized spline regression model, can accommodate a variety of study designs in addition to case-control studies, and is also useful for adjusting p-values of isoforms or exons. Using both simulated and real RNA-Seq datasets, we showed that SplineAdjust reduced bias more effectively than previously proposed methods, and the adjusted p-values had improved power and well-controlled type I error rate. With the ever-increasing application of RNA-Seq in gene expression studies, the proposed method will help improve our ability to detect genuinely DE genes in RNA-Seq studies.

1706T

Reveel2: an efficient and accurate method for large-scale population genotyping from low-coverage sequencing data. *L. Huang, S. Batzoglou.* Computer Science Department, Stanford University, Stanford, CA.

Identification of genomic variation in human DNA sequences is a key first step in associating alleles with human traits and diseases. Population low-coverage whole-genome sequencing is rapidly emerging as a prominent approach for discovering genomic variation and genotyping a cohort, because of its cost efficiency and its ability of whole-genome discovery. However, due to the sparseness and shortage of data in each individual, genotyping a cohort from low-coverage sequencing data is much more challenging than genotyping from full-coverage sequencing data. The problem becomes even more severe when attempting to genotype rare alleles at which the allele frequencies do not exceed sequencing error rates. A few existing computational methods can be applied to population genotyping. Despite their considerable success, existing methods are not ideally suited for application to large cohorts (1,000 – 1,000,000 individuals) because of their potentially massive computation time. Reveel, which is our first generation of population genotyper, has been demonstrated to be an effective method; compared to other state-of-the-arts, this method provides considerable improvement in the genotype calling of rare variants. Even though, Reveel could not fully capture the complex linkage disequilibrium (LD) patterns among high allele frequency sites. Thus, genotype phasing is used as complementary method to the genotype-calling algorithm. The phasing step reduced the efficiency of the overall algorithm by ~3x. To enable efficient and accurate population genotyping, we present Reveel2, a novel method for genotyping large cohorts that have been sequenced at low coverage. Based on a novel technique for leveraging the complex LD patterns in cohorts, Reveel2 incorporates significant enhancements to Reveel. The phasing step is not necessary in Reveel2, although we keep this option for the users place more importance on accuracy than efficiency. We evaluated Reveel2 on the empirical data from the 1000 Genomes Project, using the Complete Genomics data as a benchmark, and on extensive simulations using real data. We showed that Reveel2 outperforms its previous version Reveel and other state-of-the-art methods. Reveel2 achieved genotyping accuracy at 99. 7733% on a dataset with 100 samples; genotyping accuracy at 99. 9221% on 500 samples; genotyping accuracy at 99. 9504% on 1000 samples.

1707F

SplicER: A novel analytic scheme for the analysis of Splicing Efficiency in RNA-seq data. *L. Simon, T. Hsu, A. Renwick, N. Neill, T. Westbrook, C. A. Shaw.* Baylor College of Medicine, Houston, TX.

The advent of RNA-seq technology has created the opportunity to study transcriptomics at an unprecedented level. The nature of this technology allows researchers to study transcriptional mechanisms beyond gene expression. Global splicing efficiency is a crucial transcriptional mechanism. We define splicing efficiency as the ratio of the numbers of correctly and incorrectly spliced reads. Spliced alignments of RNA-seq data allow direct observation of splicing efficiency at exon boundaries. To better understand splicing efficiency, we have developed anew computational Python package to analyze this process (SplicER). Tovalidate the performance of the package, we used data from the Gene Expression Omnibus on known splicingmutants, and we were able to recapitulateknown differences. To extendtheanalyses with our package, we applied it to data from the Geuvadis project. Weobserved that there is great inter-individual variation insplicing efficiency that correlates with the expression of spliceosomal machinery.

1708W

Pleiotropic Variability Score: Quantifying Phenomic Associations of Genetic Variants. *M. A. Badgeley[1,2], K. Shameer[1,2], B. Kidd[1,2], B. Glicksberg[1], D. Ruderfer[3], M. S. Tomlinson[1,2], P. Wang[1], R. Chen[1], J. T. Dudley[1,2,4].* 1) Department of Genetics and Genomics, Icahn Insititute of Genomics and Multiscale Biology, Mount Sinai Health System, New York City, NY; 2) Harris Center for Precision Wellness, Mount Sinai Health System, New York City, NY; 3) Division of Psychiatric Genomics, Icahn Insititute of Genomics and Multiscale Biology, New York, NY; 4) Population Health Science and Policy, Mount Sinai Health System, New York, NY 10029, NY.

Genomic variant interpretation and assessment scores like GERP, PhyloP, SIFT, PolyPhen provides variant impact using evolutionary conservation and biophysical properties but do not account for variability induced by phenotypes. To incorporate phenomic variability into variant interpretation, we developed a method called “Pleiotropic Variability Sore”. Briefly, we mapped the phenotype information to two different phenotype ontologies (Human Phenotype Ontology and Disease Ontology) and estimated the pleiotropic similarities by combining the ontologies and semantic similarity measurements. We tested 226 unique semantic similarity assessment combinations and integrated products of six different semantic similarity metrics to obtain rankings of phenomic similarity associated with individual SNPs to define the final score. We have computed PVS for 10, 021 SNPs using 315 phenotypes mapped using GWAS catalog and PheWAS studies curated from the literature (PVSBMA=0. 777±0. 002; PVSAVERAGE=0. 525±0. 003). We used the pleiotropic SNPs defined by PVS and generated “pleiotropic hubs” using a set of genes in linkage disequilibrium with a SNP (r2<=0. 6) using SCANDB. A total of 10, 246 genes were annotated to pleiotropic hubs. PVS provide a single estimate to quantify the phenomic diversity of a given SNP. For example: 1) rs1939469 is an intronic SNP in *C11orf30* and associated with Alzheimer’s disease and vascular dementia; PVS=1 indicating low pleiotropy and the pleiotropic hub consists of *PRKRIR*, *C11orf30* and *LOC387790* 2) rs10751454 is an intronic SNP in *FAM41C* and pleiotropic hub consist of *RUNX3*, *LOC100128356*, *LOC100128356* and associated with cardiovascular system disease and cancer This variant was scored with 0. 317 indicating high pleiotropy. We examined various network properties of genes in pleiotropic hubs and found they hubs are encoded with genes mediating diverse set of drug targets, pathways and biological functions. Targeting pleiotropic hubs in the setting of complex disease could lead to better therapeutic efficacy in treating complex disease. We envisage that PVS will be a new genome interpretation metric to assess genetic variation to find SNPs with highly pleiotropic nature, and to study phenomic correlations and variants with lower PVS for explorative studies to understand specific roles of SNPs and pleiotropic hubs in mediating novel phenotypes.

1709T

A generalized non-parametric genotype caller using an EM-like Algorithm. *T. Benaglia[1], B. Carvalho[1,2], I. Lopes-Cendes[1,2].* 1) State University of Campinas, Campinas/SP Brazil; 2) Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas/SP - Brazil.

Current biomedical researches make constant use of high-throughput technologies for genotyping. Microarrays and, more currently, sequencing experiments provide information at millions of variant sites, which are later investigated for association with a phenotype of interest. Prior to this association assessment, genotyping algorithms quantify the allele-specific abundance and classify subjects into genotype clusters. The series of statistical procedures involved in the aforementioned quantification and classification steps is complex and includes strong hypotheses on the distribution of the data. Models for microarray data assume that intensity signals follow an Exponential distribution combined with Gaussian error, while count data from sequencing are seen as results of a Negative-Binomial processes. We propose an alternative genotype caller that does not rely on assumptions regarding the distribution of the data. Our method uses a non-parametric approach combined with an adapted Expectation-Maximization (EM-like) algorithm to determine the membership likelihoods of each subject at every *loci*. Following this strategy, we use the maximum posterior probability to accurately perform genotype calling.

1710F

AIPS: Ancestry Inference of Subpopulations using Scores from Principal Component Analysis in Genome-Wide Association Studies. *J. Byun[1], Y. Han[1], M. F. Seldin[2], C. I. Amos[1].* 1) Biomedical Data Science, Dartmouth College, Hanover, NH; 2) Rowe Program in Genetics, U. C. Davis, Davis, CA.

Genome-wide association studies (GWAS) in high-density single-nucleotide polymorphism (SNP) genotyping data have identified thousands of common variants associated to complex disease risks and traits. An accurate inference of genetic ancestry is extremely critical in many biomedical research areas, because the difference in genetic population structure between cases and controls can lead to false-positive results. A challenge in the larger genetic data to understand the data structure is whether the individuals are from a homogeneous population or any heterogeneous subpopulation. The Principal Component Analysis (PCA) is the most common approach in the population and admixture structures of genetic data. The R-package AIPS allows us to calculate the SNP weights and scores from PCA, predict the scores from SNP weights computed on the same pre-defined AIMs, and provide the ancestry definitions using pre-defined ancestry clustering information. To estimate ancestry definitions, we applied the Inverse Distance Weighted (IDW) interpolation method from spatial analysis. IDW interpolation assumes that points that are close to one another are more likely alike than those that are farther apart. To infer an ancestry proportion for an unknown sample, IDW can compute the distance metrics from each centroid of the known population. Those estimated values closest to the centroid of known population will have more local influence that diminishes with distance and be weighted greater than those farther away. To demonstrate, we used 4,376 samples with 22 pre-defined European substructures and 2,830 unknown ancestry samples on 25,732 ancestry informative markers. It takes about 15 minutes to calculate scores with the merged 7,206 samples or to predict unknown ancestry definitions after calculating SNP weights in 4,376 known ancestry samples. Also, it takes less than 10 minutes to calculate the distance metrics of substructures for ancestry inferences. The R-package AIPS will be publicly available.

1711W

Processing microarray data using alternative annotations through the Bioconductor oligo package. *B. Carvalho[1,2], I. Lopes-Cendes[1,2].* 1) State University of Campinas, Campinas/SP - Brazil; 2) Brazilian Institute of Neuroscience and Neurotechnology (BRAINN), Campinas/SP - Brazil.

The Bioconductor *oligo* package is a well-established microarray preprocessing software. Currently, it is in the Top 5% most downloaded Bioconductor packages rank, being downloaded 20,000 times in the period of July 2014 – June 2015. The package uses annotation files provided by manufacturers to estimate statistics, like expression abundance, that are required for downstream analyses. These annotation resources often require further curation to reflect the acquired knowledge in the recent past. Common examples of such tasks are removing probes known to deliver poor performance and subsetting preprocessing to genes uniquely mapped by the available sets of probes. We implemented an infrastructure that allows researchers to use alternative annotation resources under the *oligo* framework. Researchers can use our approach to incorporate curated annotations (BrainArray remapped designs, for example) to obtain improved preprocessing results. Our novel approach requires simplified input in the form of tables, and unlocks full access to the Bioconductor environment even for complex array designs, like Gene ST chips.

1712T

Gaussian processes for medical time-series extrapolation and event prediction. *L. Cheng[1], G. Darnell[2], K. Li[3], M. Draugelis[4], B. Engelhardt[3,5].* 1) Department of Electrical Engineering, Princeton University, Princeton, NJ, USA; 2) Lewis-Sigler Institute, Princeton University, Princeton, NJ, USA; 3) Department of Computer Science, Princeton University, Princeton, NJ, USA; 4) Hospital of the University of Pennsylvania, Philadelphia, PA, USA; 5) Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA.

In real-time monitoring of hospital patients, early detection of clinical events such as heart attacks or strokes are essential for successful medical interventions and improving patient outcomes. In this work, we develop a statistical framework for early prediction of sepsis that learns from clinically recorded physiological time-series data. Differences in the nature of sampling for clinical covariates make the data sparse and noisy, leading to challenges for real-time monitoring of a patient's status. For instance, vital signs are measured with fairly regular periodicity in the hospital setting, while other more informative measures, such as lab tests, are often invasive and expensive, and are less frequently and less regularly performed. In order to handle the incomplete data and improve the accuracy of sepsis prediction, we propose a statistical framework based on Gaussian processes (GPs), which defines a distribution over functions that map time to the set of clinical covariates for a single patient. Using GPs, we are able to characterize and predict the mean and variance (i. e. , uncertainty) forward in time for a single patient, given a large set of reference patients. In particular, we motivate our model using a collection of data from over 260,000 patent stays at the Hospital of the University of Pennsylvania. Since our data consist of multiple clinical covariates and is highly heterogeneous, we propose a model that shares strength across many clinical covariates with smoothness to reduce noise. In order to rapidly assess a new patient, we use our rich reference data to identify subgroups of similar patients by clustering patient demographic information, and we train one GP model for each group separately. Finally, we use Bayesian online changepoint detection, allowing online prediction of medically relevant events and updating the GP models on-the-fly as more information is gathered on each patient. Using the proposed framework, we show that the prediction error of covariate values is reduced compared with standard linear methods; furthermore, our model allow the explicit characterization of uncertainty in predictions, and the predictions may be traced back to the informative clinical covariates. In summary, the proposed GP framework enables tractable fine-grained prediction of temporal trends and disease onset in complex and heterogeneous time series medical data by exploiting observable dependencies among clinical covariates.

1713F

Realistic simulation of mutations for improving mutation assessment. *A. D. Ewing.* Mater Research Institute, University of Queensland, Woolloongabba, QLD, Australia.

Mutation detection is a fundamental step for many researchers studying human disease. The detection of differences between an individual genome and the reference genome, or between multiple human genomes is subject to false positive and false negative mutation calls. While false positives can be assessed through site-specific methods such as PCR or targeted resequencing, false negatives are more difficult to ascertain. This is a particularly important problem where somatic genomes are concerned, especially in the analysis of tumours with subclonal heterogeneity. Simulation of mutations provides a solution to the problem of accurate assessment of both false positive and false negative mutations, but the simulation needs to take into account normal human genomic variation and the error profile of the sequencing technology, which may vary substantially between individuals, applications, and sequencing platforms. To this end, I have developed BAMSurgeon, a software tool that allows the addition of mutations to pre-existing BAM files. BAMSurgeon is capable of simulating many mutation types including SNVs, INDELs, insertions, deletions, duplications, inversions, and translocations. It supports haplotype-aware spike-in, and many other customisable features to create realistic synthetic mutations in both germline and somatic contexts. BAMSurgeon has been used in a number of studies including the ICGC-TCGA DREAM-SMC Challenge, and the development of mutation calling methods. Through application in these studies and through user feedback, BAMSurgeon continues to improve in terms of usability and realism. In this presentation, I will cover the operation, use cases, and capabilities of BAMSurgeon, which can be obtained at https://github.com/adamewing/bamsurgeon.

1714W

Evolutionarily Derived Networks to Inform Disease Pathways. *B. E. Graham[1], C. Darabos[1,4], M. Huang[1], L. Muglia[3], J. H. Moore[4], S. M. Williams[1,2].* 1) Institute for Quantitative Biomedical Sciences, Dartmouth College, Hanover, NH; 2) Department of Genetics, Dartmouth College, Hanover, NH; 3) Cincinnati Children's Hospital, Cincinnati, OH; 4) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Identifying genes or pathways that associate with human phenotypes continues to be a useful means to understanding disease etiology. However, for complex diseases methods to do this are often inadequate to elucidate a majority of risk. We have previously developed human phenotype networks (HPN) as a means to identify connections between traits or diseases based on shared genetic backgrounds. This has proven useful in terms of demonstrating common processes among diverse phenotypes. In the present study, we expand HPN to include another novel method we have developed, Evolutionary Triangulation (ET), which is a statistical comparison of allele frequencies among three populations and their relationship to phenotype prevalences. ET can refine the selection of genes or pathways prior to implementing HPN. Briefly, with ET we select continental populations based on differences in disease frequencies, such that two populations have similar prevalence while a third is divergent. To build our network, we use ET to identify SNPs by triangulating for a target disease and we rely on prior association data to map the SNPs to their associated phenotypes. The resulting network is bipartite, containing two distinct sets of nodes: SNPs and traits. The ET derived HPN is obtained by projecting the bipartite graph into the disease space. For melanoma (our index disease in the present study), we selected YRI, GIH and CEU as sample populations based on disease prevalences. The ET analysis we used extracted 733 SNPs mapped to 25 phenotypes linked by 236 edges. Four of the phenotypes identified, including ALS and Hodgkin's lymphoma, follow the same distribution as melanoma in these populations. Although the index phenotype is missing from the network, we found tanning and vitamin D as traits, both of which are associated with melanoma. We also identified additionally associating traits, related to Type 2 diabetes (T2D) which also follows the same distribution. Specifically, we identified 18 related traits, including Glucose and HbA1c levels, both major markers for T2D. In conclusion, applying network analysis to ET datasets can extract subnetworks of traits related to multiple phenotypes that follow the same distributions across populations, demonstrating the utility of combining these two analytical approaches to further elucidate underlying etiological pathways of complex disease.

1715T

Rapid Generation of Illumina Infinium Genotyping Release Data. *S. M. L Griffith[1], H. Ling[1], K. N. Hetrick[1], J. Romm[1], E. W. Pugh[1], I. A. McMullen[1], D. R. Leary[2], B. R. Myers[1], M. Z. Mawhinney[1], M. E. Hurley[1], A. B. Robinson[1], L. Watkins, Jr. [1], K. F. Doheny[1].* 1) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 2) New York Genome Center, New York, NY.

The Center for Inherited Disease Research (CIDR) provides high quality genotyping and sequencing services and statistical genetics consultation to investigators working to discover genes that contribute to disease. Prior to release, raw data (in the form of image files, or . idats) generated by the Illumina Infinium chemistry are put through a data analysis pipeline, generating genotypes. Specifically, the raw . idat files generated by HiScan scanners are loaded into GenomeStudio to generate quality control metrics for each project. Next, final genotypes for a project are generated by first clustering SNPs using a project's high quality samples, and then applying a technical SNP filter to get final SNP cluster definitions. Historically, using this cluster definition, the final genotypes have been generated using GenomeStudio. As projects have gotten larger over time, it can take weeks to generate these genotype files and, thus, this approach has become time-prohibitive. For example, processing a set of 19,000 samples run on the two-million-SNP MEGA Array, and analyzed with GenomeStudio would have taken more than a month using a blade server with 192 GB of RAM and 32 processor cores. To generate these data more rapidly, the software development group at CIDR has created a massively parallel data generation tool. Following reclustering and final cluster assignments for each variant, this tool generates genotyping release data by directly parsing the GTC files generated via Beeline, Illumina's standalone genotype caller. Using these data, the tool normalizes the raw intensity data and calculates several metrics, including Log R Ratio and B Allele Frequency. The per-sample data generation tasks are embarrassingly parallel, and the entire process scales with available resources. Using the same blade server, data were generated for the 19,000 MEGA array samples in fewer than 24 hours. In addition to generating individual genotyping reports on a per sample basis, this tool also generates several supplementary PLINK files containing the same genotypes as the individualized genotyping reports. These files include two PLINK Ped files (one with pedigree information and one without), a PLINK Map file, and strand translation files from top/bottom to forward/reverse or A/B formats.

1716F

Integration of Bina RAVE for genomic data analysis at the Center for Inherited Disease Research. *K. N. Hetrick, B. Myers, S. M. L. Griffith, D. Snyder, K. F. Doheny, L. Watkins.* Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality genotyping and sequencing services and statistical genetics consultation to investigators working to discover genes that contribute to disease. With the ever increasing volume of NGS data being generated and constantly evolving ways to analyze them; software development, bioinformatics and IT personnel, as well as local HPC resources can find themselves straining to maintain the same level of quality and efficiency with fast turn-around times and expertise. There are a bevy of commercial platforms that aim to lessen this burden with a multitude of those aimed at uploading and analyzing data in the cloud. However due to privacy concerns with sample data, cloud-based solutions are not an option for all facilities. CIDR assessed the Bina RAVE (Read Alignment, Variant Calling and Expression) module (Bina Technologies, Redwood City, CA) deployed on a single Bina Rack local analysis appliance (4 nodes, 64 cores) with optimized workflows for whole genome/exome (WGS/WES) SNV/indel calling, WGS CNV calling, tumor/normal, and RNA analyses. For our initial evaluation, we focused on comparing the results of a WGS sample sequenced on an Illumina® HiSeq™ 2000 at 38x depth analyzed on the Bina appliance using bwa mem and GATK 3 HaplotypeCaller in GVCF mode to an analysis generated outside of the appliance using the same workflow and software versions. Between the two analyses, there was a 99. 9% overlap of the unfiltered SNV calls, with 98. 3% of passing SNVs overlapping. The workflow finished in 9 hours on the Bina appliance with complete BAM and VCF outputs, and included CNV analyses that we had not otherwise implemented. This lessened the burden on our IT and software development staff (e. g did not need to reconfigure our HPC system, develop a software framework to optimize complex/distributed analyses needed to complete WGS analysis in under a day, etc). Throughout and after the evaluation period, more features were added to the appliance software. In addition, we jointly called and filtered ~800 WES samples on the appliance using the GATK 3 joint calling and Variant Quality Score Recalibration workflow in 2 hours. We are currently integrating their API to automate submission of samples for analysis on the appliance as well as automate submission of other downstream analyses on our cluster after the Bina analyses are done. We also plan to evaluate their other analysis workflows (e. g. tumor/normal).

1717W

Efficient approach to correct read alignment for pseudogene abundance estimates. *C. J. -T. Ju, Z. Zhao, W. Wang.* Computer Science, University of California, Los Angeles, Los Angeles, CA.

RNA-Sequencing has been the leading technology to quantify expression of thousands genes simultaneously. The data analysis of an RNA-Seq experiment starts from aligning short reads to the reference genome/transcriptome or reconstructed transcriptome. However, current aligners lack the sensitivity to distinguish reads that come from homologous regions of an genome. Pseudogenes arise from duplication of a set of protein coding genes, and have been considered to be degraded paralogs in the genome due to their lost of functionality. Recent studies have provided evidence to support their novel regulatory roles in biological processes. With the growing interests in quantifying the expression level of pseudogenes at different tissues or cell lines, it is critical to have a sensitive method that can correctly align ambiguous reads and accurately estimate the expression level among homologous genes. Previously in PseudoLasso, we proposed a linear regression approach to learn read alignment behaviors, and leverage this knowledge for abundance estimation and alignment correction. In this paper, we extend the work of PseudoLasso by grouping the homologous genomic regions into different communities using a community detection algorithm, followed by building linear regression models separately for each community. The results show that this approach is able to retain the same accuracy as PseudoLasso. By breaking the genome into smaller homologous communities, the running time improves drastically with an increase number of genes.

1718T

Ferret: a user-friendly Java tool to extract data from the 1000 Genomes Project. *S. Limou[1], A. Taverner[2], G. Nelson[3], C. Winkler[1].* 1) Molecular Genetic Epidemiology Section, Basic Research Laboratory, Basic Science Program, NCI-Frederick, Leidos Biomedical Research, Inc. , Frederick National Laboratory, Frederick, MD; 2) Quantitative and Computational Biology program, Princeton University, Princeton, NJ; 3) Center for Cancer Research Informatics Core, Leidos Biomedical Research, Inc. , Frederick National Laboratory, Frederick, MD.

The 1000 Genomes (1KG) Project provides a near-comprehensive resource on human genetic variation in worldwide reference populations. 1KG variants can be accessed through a browser and through the raw and annotated data that are regularly released on an ftp server. Two years ago, we developed Ferret as a Perl script to easily extract genetic variation information for a locus of interest from these large and complex data files. We now provide Ferret as a user-friendly Java tool with novel features: (1) the input can be a locus, gene(s) or SNP(s) of interest; (2) Ferret retrieves genotype data for 1KG SNPs and indels; (3) Ferret computes SNP, indel and CNV allelic frequencies for 1KG populations and (4) optionally for the Exome Sequencing Project populations. By converting the 1KG data into files that can be imported into popular pre-existing tools (*e. g.* PLINK and HaploView), Ferret hence offers a straightforward way, even for non-bioinformatics specialists, to manipulate, explore, and merge 1KG data with the user's dataset, as well as visualize linkage disequilibrium pattern, infer haplotypes, and design tagSNPs. Ferret is publicly available at: https://ccrod. cancer. gov/confluence/display/BCGC/BCGC+Software. You may contact us at ferret@nih. gov.

1719F

GAMETES 2. 0: Expanding the complex model and data simulation software to generate heterogeneous datasets, custom models, and quantitative traits. *J. H. Moore, R. U Urbanowicz, P. Andrews.* Genetics, University of Pennsylvania, Philadelphia, PA.

Increasing acknowledgement of the complexity of common diseases, particularly with regards to complex multivariate patterns of association, has led to an increased interest in the development of new analytical methodologies, algorithms, and software able to detect, model, and characterize such patterns. In order to properly develop, test, and evaluate such methodologies, a variety of representative simulated datasets for simulation studies are required. Previously we developed the GAMETES software for the rapid, deterministic generation of strict, purely epistatic single nucleotide polymorphism (SNP) models based on user defined parameters such as heritability, minor allele frequencies, prevalence, and the order of interaction (e. g. 2-way, or 3-way). GAMETES 2. 0 expands the capabilities of this software, allowing users to (1) combine multiple genetic models for the simulation of datasets with patterns of genetic heterogeneity, (2) generate custom 2 or 3-way SNP models with a report of the model's characteristics and predicted relative detection difficulty, (3) generate datasets with a quantitative trait/endpoint as opposed to a binary discrete class endpoint. Quantitative endpoints are generated from GAMETES genetic models by using the penetrance values of specific genotype combinations as a centroid for selecting a continuous trait value for each subject in the dataset. We test this new simulation software by generating simulation studies with heterogeneity and quantitative traits respectively and demonstrate that we can identify these simulated patterns using advanced machine learning approaches (i. e. QMDR and EXSTRACS), and feature selection approaches (ReliefF, SURF, SURF*, and MultiSURF).

1720W

Gene: a gene-centered information resource at NCBI. *T. Murphy, D. Maglott, K. Pruitt, the Gene Team, and the RefSeq Curation Team.* NCBI, NIH, Bethesda, MD.

The National Center for Biotechnology Information (NCBI) Gene resource (www. ncbi. nlm. nih. gov/gene/) integrates gene-centric information for over 11 million genes from over 12 thousand taxa spanning the tree of life. Records include a diverse set of content including citations, nomenclature, genome annotation, gene products and their attributes, phenotypes, sequences, interactions, variation details, maps, expression, homologs, protein domains and crosslinks to a variety of external databases. Genome annotation and products originate from NCBI's RefSeq resource, which includes extensive manual curation for human and mouse genes and genome annotation from NCBI's Eukaryotic Genome Annotation Pipeline. Gene web pages also serve as a central hub to other NCBI resources, including browsers (Map Viewer, Variation Viewer, 1000 Genomes, Epigenomics), RefSeq genome, transcript and protein sequence records, ClinVar, dbSNP, PubMed, HomoloGene, and many others. This presentation will focus on the sources and types of content available on human gene records, including data tracks available in the Graphical viewer, support for annotation on multiple assemblies (GRCh38. p2, GRCh37. p13, and CHM1_1. 1), annotation from both NCBI and Ensembl, and crosslinks to orthologous records from hundreds of other vertebrates. Plans for ongoing development in reporting expression data and improved web presentations will be discussed. All data is available via web pages, FTP, and NCBI's e-utils programming utilities.

1721T

BioBin: A comprehensive tool for the biologically-inspired binning and association analysis of low frequency variants in sequence data. *A. Okula[1], J. Wallace[1], A. Frase[1], M. D. Ritchie[1,2].* 1) Center for Systems Genomics, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA; 2) Geisinger Health System, Danville, PA.

Advances in sequencing technology and decreases in cost have presented an opportunity for the investigation of low frequency and rare variants beyond traditional GWAS-based approaches. Rare variant analysis, however, is challenging as statistical power for detection is low. As a means of boosting power, many methodologies have been specifically developed to facilitate rare variant analysis. Many methods focus either on the creation of a relevant set of variants or on the statistical analysis of an arbitrary set of variants. In this work, we expand the framework of BioBin providing a comprehensive bioinformatics tool for the automated and biological inspired binning and association analysis of rare variants. BioBin was initially developed to perform multi-level binning of rare variants based on user-designated features such as genes, pathways, ECRs, and regulatory regions. BioBin follows a flexible, biologically informed binning strategy using an internal biorepository known as LOKI, or the Library of Knowledge Integration. LOKI combines over a dozen databases from the public domain including NCBI Entrez Gene, KEGG and Protein Families. These databases provide variant details, regional annotations and pathway interactions used to generate bins of biologically-related variants, thereby increasing the power of detection for any subsequent test. To facilitate statistical analysis, we have added an extensible testing infrastructure to BioBin and have implemented select burden and dispersion tests, allowing the user to perform regression, wilcoxon and/or SKAT on binary or quantitative phenotypes subsequent to variant binning. By integrating the tests into BioBin directly, we avoid file conversion issues for specific tools, saving analysis time and making the process more robust. BioBin is also capable of multiple phenotype analysis allowing the execution of a rare variant binned phenome-wide association study (PheWAS). Because each phenotype can be analyzed independently, BioBin uses parallel processing to increase the speed of a PheWAS binning analysis through a user-specified number of processors. To highlight the speed of analysis, we were able to perform a gene-based binning SKAT test to analyze 30 quantitative phenotypes in 8,000 individuals using whole-exome sequence data in 5 hours using less than 12GB of memory. Overall, BioBin is a powerful and versatile tool for the biological binning and analysis of low frequency variants for complex phenotypes.

1722F

Geneious R9: a bioinformatics platform for biologists. *K. Qaadri[1], C. Olsen[1], H. Shearman[2], R. Moir[2], M. Kearse[2], S. Markowitz[2], J. Kuhn[2], S. Dunn[2], T. Paulin[2], A. Cooper[2].* 1) Biomatters, Inc. , San Francisco, CA; 2) Biomatters Ltd, Auckland, New Zealand.

Biomatters' Geneious R9 is a bioinformatics software platform that allows researchers the use of industry-leading algorithms for their genomic and protein sequence analyses. R9 offers a comprehensive suite of functions, including peer-reviewed tools, that enable researchers to be more efficient with their bioinformatic workflows.

Geneious is comprised of an extensive tool suite for next-generation sequence analysis, sequence alignment, molecular cloning, chromatogram assembly, and phylogenetics. This major version release includes tools for RNA-Seq expression analysis, support for scaffolding using paired data in *de novo* assemblies, a function to identify CRISPR sites and support for Golden Gate assembly. Additional features for this software release include the 16S Biodiversity tool and Sequence Classifier plugin. The 16S Biodiversity tool identifies high-throughput 16S rRNA amplicons from environmental samples using the RDP database, and visualizes biodiversity as an interactive chart using a secure web viewer. The Sequence Classifier plugin taxonomically classifies an organic sample by how similar its DNA is to your own database of known sequences using a BLAST-like algorithm with multiple loci and trees to assist with identification. New plugins include Blast2GO (Gene Ontology enrichment), FreeBayes (haplotype-based variant calling), Augustus (gene prediction), and FLASH (read merging).

R9 enables real-time dynamic interaction with sequence data and empowers biologists to produce stunning publication quality images to increase the visibility of their research. By utilizing Geneious R9, biologists can easily streamline their sequence analysis workflows. This poster aims to demonstrate the new features and benefits of the highly integrated Geneious R9 tool-suite.

1723W

Novel Approach for Correction of Cell-Type Heterogeneity Improves Power of Epigenome-Wide Association Studies. *E. Rahmani[1], Y. Baran[1], N. Zaitlen[2], C. Eng[2], D. Hu[2], J. Galanter[2,3], S. Oh[2], E. Burchard[2,3], E. Eskin[4,5], J. Zou[6], E. Halperin[1,7,8].* 1) Blavatnik School of Computer Science, Tel-Aviv University, Tel Aviv, Israel; 2) Department of Medicine, University of California San Francisco, San Francisco, California; 3) Department of Bioengineering and Therapeutic Science, University of California San Francisco, San Francisco, California; 4) Department of Computer Science, University of California, Los Angeles, California; 5) Department of Human Genetics, University of California, Los Angeles, California; 6) Microsoft Research New England, Cambridge, Massachusetts; 7) International Computer Science Institute, Berkeley, California; 8) The Department of Molecular Microbiology and Biotechnology, Tel-Aviv University, Tel Aviv, Israel.

Recent work applying epigenome-wide association studies (EWAS) suggest an important role for DNA methylation as a mechanism involved or associated with disease. The standard statistical analyses applied in EWAS test for an association between probed CpG site with the phenotype of interest, similar to the approach taken in genome-wide association studies (GWAS). The analysis of EWAS conducted in primary tissue such as blood is complicated by the fact that tissues are complex mixtures of diverse cell types, each with their own methylation profile. When cell types are correlated with the phenotype of interest, tissue heterogeneity will lead to a large number of false discoveries in standard analyses, similar to the case of population structure in GWAS. In order to overcome this challenge, different analysis methods have been proposed to account for the cell type composition. These methods are either based on reference data in which methylation profiles have been carefully measured for sorted cells, or they implicitly model cell type composition without the use of reference profiles. In this study we show that current methods do not adequately correct for false positive rate and they are sub-optimal in terms of power. We introduce a new highly efficient reference-free method for EWAS analysis. An unsupervised approach is critical since methylation is affected by many factors (e. g. , age, sex and genetics), and a reference-based method is exposed to bias from potential differences between the reference and the study population due to poor matching of these factors. Unlike GWAS were many SNPs are potentially informative, only a small number of the CpG sites are significantly different between cell types, therefore using more sites in a factorization method such as principal components analysis (PCA) may sometimes add more noise than a signal. Motivated by this observation, our method is based on an unsupervised feature selection step followed by a sparse PCA. We show, based on both simulated and real data (the GALA II study), that estimates of the cell type composition using our method are better than both reference-based estimates and PCA. We also show that applying our method to EWAS results in better control for false positives as well as a significant power increase compared to existing methods. We further demonstrate the utility of our method on a previous Rheumatoid Arthritis EWAS, revealing a new unreported methylation site associated with the disease.

1724T

Network-assisted Method of Genome-wide DNA Methylation Association Studies in Cancers. *P. Ruan[1], J. Shen[2], R. Santella[2], S. Zhou[1], S. Wang[3].* 1) School of Computer Science, Fudan University, Shanghai, Shanghai, China; 2) Department of Environmental Health Science, Mailman School of Public Health, Columbia University, New York, NY, USA; 3) Department of Biostatistics, Mailman School of Public Health, Columbia University, New York, NY, USA.

DNA methylation plays important roles in the carcinogenesis process. Many genome-wide DNA methylation association studies have been conducted that have identified aberrantly methylated genes in cancer. Most of these studies focus on analyzing methylation one site at a time; the result may be studies are underpowered to identify cancer related genes as genes interact with each other in carcinogenesis process. Incorporating prior biologic information such as biology network has been proven to be a useful way to examine the joint effect of genes. Here, using DNA methylation data, we present a network-assisted method to search for modules that are associated with cancer by incorporating the association signals that capture differences in both means and variances between the two experimental conditions, such as between tumor and adjacent normal groups with a human protein-protein interaction network. The method searches for the modules that are enriched with genes having stronger association signals. Application to DNA methylation data of hepatocellular carcinoma (HCC) from Columbia University Medical Center (CUMC) and the Cancer Genome Atlas (TCGA) project and breast invasive carcinoma (BRCA) from TCGA demonstrate that our method finds several genes that are known to be related to cancer but are missed by methods ignoring the prior biology information or signals in DNA methylation variation.

1725F

Comparison of preprocessing methods for the Luminex xMap® technology. *A. Schillert[1], T. K. Rausch[1], H. -D. Zucht[2], A. Ziegler[1,3,4,5], P. Schulz-Knappe[2].* 1) Inst Med Biom and Statistics, Universitaet zu Luebeck, Universitaetsklinikum Schleswig-Holstein, Campus Luebeck, Germany; 2) Protagen AG, Dortmund, Germany; 3) Zentrum fuer Klinische Studien, Universitaet zu Luebeck, Luebeck, Germany; 4) Deutsches Zentrum fuer Herz-Kreislauf-Forschung, Standort Hamburg/Kiel/Luebeck, Luebeck, Germany; 5) School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa.

As a multiplex alternative to ELISAs the Luminex xMap® technology allows to analyze up to 500 different analytes simultaneously. As with most high throughput technologies data preprocessing is crucial to eliminate technical variation. A sequence of transforming the data and then combining it via a normalization method is widely accepted. Multiple methods for transforming and normalizing the data have been proposed for gene expression data. However, it remains unclear which combination of transformation and normalization methods yields the best results in the context of proteomics. Through literature search we identified six transformation and seven normalization methods. To test the performance of these methods we used data from a Luminex xMap® experiment of 42 samples and 384 antigens. Intensity values equal of zero were coded as missing. During quality control we excluded antigens with missing data for more than eight samples. We performed median imputation to replace missing values. We then applied a transformation followed by a normalization method. As quality diagnostics we used three statistics for the distribution of the intensity values - tail length, skewness, and coefficient of variation. Additionally, we generated Bland-Altman, Mean-Sd, and Volcano plots. These plots were then evaluated by 15 trained raters in a blinded setting. From the statistics and the rating results we constructed a score between 0 and 12 to identify useful combinations of transformation and normalization methods. The quality controlled data contained 380 antigens. Of the 37 useful combinations of transformation and normalization methods 10 achieved a score of 11 or 12. The transformations were log2, asinh and one of two versions of boxcox followed by either loess or quantile normalization or robust spline normalization after a boxcox transformation. Data preprocessing has a large impact on the validity and power of any statistical analysis using high-throughput data. This is also true for Luminex xMap®data. Our investigations show that for Luminex data transformation is mandatory and normalizing greatly improves the data quality. Suitable combinations are available.

1726W

LAMPLINK: An additional function for PLINK to detect statistically significant genetic interactions. *A. Terada[1,2,3], K. Tsuda[1,3], J. Sese[3].* 1) Computational Biology and Medical Sciences, The University of Tokyo, Chiba, Japan; 2) Research Fellow of the Japan Society for the Promotion of Science; 3) Biotechnology Research Institute for Drug Discovery, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan.

PLINK is widely used software to analyze genome-wide association study (GWAS) data. Its major functions list single nucleotide polymorphisms (SNPs) statistically significantly associated with a target phenotype. While PLINK has uncovered thousands of disease causal SNPs, it cannot detect epistatic effect, in which each chromosomal position is not highly associated with a target trait but their combination shows statistically significant association with the trait. Detecting epistatic effect is difficult not only for PLINK but also for other GWAS analysis software because they use low sensitive multiple testing procedure. Recently, a novel multiple testing procedure, called Limitless Arity Multiple-testing procedure (LAMP) [Terada et al, PNAS, 2013], was proposed for combinatorial effects discovery. It can enumerate any SNP combinations statistically significantly associated with a target trait, whereas the family-wise error rate (FWER) theoretically controls under the equal level to the Bonferroni correction. We incorporated the LAMP with PLINK as an option of multiple testing procedures, named LAMPLINK. Our demonstration using dataset including more than 100,000 SNPs and 364 individuals showed that LAMPLINK detected combinations of three or more SNPs as causal SNPs of Alzheimer's disease, all of which were overlooked by the Bonferroni correction. LAMPLINK can contribute to solve the missing heritability problem.

1727T

Calcordance: A Genotyping Concordance Tool. *B. D. Tibbils, S. M. L. Griffith, M. Z. Mawhinney, I. A. McMullen, M. Jewell, A. B. Robinson, C. M. Ongaco, M. E. Hurley, J. M. Romm, L. Watkins, Jr. , K. F. Doheny.* Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality genotyping and sequencing services and statistical genetics consultation to investigators working to discover genes that contribute to disease. Concordance (in genotyping) involves determining the number of identical SNP calls between two independently generated data sets for the same individual. In this particular use case, one data set comes from a fixed set of reference information and the other is from a data set submitted by the user. Calcordance (a portmanteau of calculation and concordance) is a software application created to generate concordance information for the controls of a genotyping project as quickly as possible with a very large set of reference data. Originally, HapMap genotypes were used as reference data used for concordance at CIDR. These data are comparatively small, already divided into CSV files by sample name and only contain the necessary information: SNP name and allele zygosity. At that point, the strategy was simply to grab the file with the sample name matching that submitted by the user and go through the two data sets, line by line, performing the calculations. Once a much larger reference data set was introduced, this was no longer a reasonable solution. The larger data set, genotypes from the 1000 Genomes Project, is divided by chromosome. Each file is between 10 and 60 gigabytes and the files are in the VCF format. Retaining the same strategy of looping through each line of each file could take hours or even days with this data set, so the data needed to be restructured. To compress these data, three relationships were introduced: chromosome to sample name, sample name to SNP and SNP to allele zygosity. Allele zygosity is always one of two possibilities: homozygous or heterozygous. This data is represented as a single bit. Using a compact data structure we associated each SNP for a given sample to its allele zygosity bit. By serializing this structure to the file system for repeated use, this information could now be quickly accessed during runtime. This data structure was created for each sample and all samples for a particular chromosome were stored together. This increased our efficiency and brought our processing time from two or three days down to less than an hour for a large data set.

1728F

A web portal for rapid imputation of summary statistics in association studies. *J. N. Torres[1], H. Shi[2], B. Pasaniuc[3], G. L. Wojcik[4], C. R. Gignoux[4], C. D. Bustamante[4].* 1) Biomedical Informatics, Stanford, Stanford, CA; 2) Bioinformatics, UCLA, Los Angeles, CA; 3) Human Genetics, UCLA, Los Angeles, CA; 4) Genetics, Stanford, Stanford, CA.

An important component of genome-wide association studies (GWAS) is genotype imputation to capture untyped variation. Many popular imputation packages to impute individual-level genotypes are based on Hidden Markov models (HMM), which are often computationally demanding, especially in large GWA studies. Due to growing volume of genomic data available as well as changing privacy and logistical constraints on data, individual genotypes may not be readily available or genotype imputation not computationally feasible. As a substitute, methods that can leverage summary association statistics are becoming more popular. Here, we highlight a user-friendly imputation approach built upon the ImpG-Summary algorithm to test for associations at untyped SNPS when only summary association statistics are available. Summary association statistic imputation is performed using a multivariate Gaussian model for linkage disequilibrium in a window with an added penalty term for the statistical noise introduced by a finite sized reference panel to reduce false positives. We present a web-based service written in Python that allows for summary association imputation with easy file formatting and user-defined metrics of window size, buffer length regions, and specified ridge regression penalty to account for sampling noise, as well as choice of reference haplotypes to be used. We evaluate various reference panels for diverse populations from the 1000 Genomes Project Phase 3 data release and are using admixture deconvolution to elucidate ancestral LD patterns for improved performance in various populations. This service allows for a fast and simple-to-use utility to test associations using a method that approaches the accuracy of common HMM based approaches but in orders of magnitude less time. In practice our version of the algorithm with pre-computed LD can impute chromosome 20 on the Multi-Ethnic Genotyping Array from 17k sites 20 up to over 1. 7M sites from 1000 Genomes Phase 3 data in an average of 138 seconds. As the size and scope of genetic research increases we anticipate this being a useful screening resource for rapid follow-up and increasing the utility of publicly available data.

1729W

A new 'front' in rule-based data mining for complex, heterogeneous, and noisy association analyses. *R. J Urbanowicz, J. H Moore.* Gen, University of Pennsylvania, Philadelphia, PA.

Biological and statistical phenomena such as epistasis, genetic heterogeneity, and phenocopy can mask the relationship between genotypic, epigenetic, and environmental risk factors/markers and phenotypes of interest. It has often been suggested that such phenomena may account for a substantial portion of so called 'missing heritability' across genetic association studies of common human heritable diseases. The modern data mining toolkit for association analyses mostly includes approaches that labor under restrictive assumptions such as the number of predictive variables, the application of a specific genetic model, linearity, or homogeneity in order to function quickly, effectively, and reliably. Previously, we developed a rule-based machine learning algorithm called ExSTraCS, an Extended Supervised Tracking and Classifying System, for assumption-free classification, prediction, and knowledge discovery designed to be particularly advantageous in detecting, modeling and characterizing complex, noisy, multivariate, epistatic, and heterogeneous patterns of association. The key to this flexibility is that ExSTraCS, like other Learning Classifier System algorithms learns piece-wise effective generalizations of the problem space. In other words, human interpretable rules are evolved to individually capture subspaces of the overall pattern and collectively applied to form the predictive 'model'. One major challenge is to be able to compare and rank the 'value' of these evolved rules in a way that emphasizes both the accuracy and the correct coverage of the dataset in order to reduce overfitting and promote solution interpretability in noisy or heterogeneous problems. In the present study we introduce a Pareto-front-inspired methodology for the calculation of rule-fitness within ExSTraCS that provides a reliable, multi-objective global value metric. Specifically, as rules emerge and are evaluated, a non-dominated front of points defined by the accuracy and correct coverage of a rule is updated and applied to calculate the fitness of every existing rule as a function of the distance from the non-dominated front, and the relative area under the front. We find that this methodology significantly improves performance, interpretability, and allows for dramatic and simple rule compaction, across a spectrum of complex noisy simulation studies concurrently modeling epistatic and heterogeneous patterns with assorted heritabilities and sample sizes.

1730T

A new change-point model based method for copy number variation detection. *F. Xiao[1], X. Min[2], H. Zhang[2].* 1) University of South Carolina, Department of Epidemiology and Biostatistics, Columbia, SC; 2) Yale School of Public Health, Department of Biostatistics, New Haven, CT.

Copy number variation (CNV) is an important type of structural variation and is associated with human complex disorders, such as autism, growth retardation and HIV progression. Duplication or deletion on any of the two copies of genomic segments results in CNV of this region. To detect CNVs in human genome, several change-point based techniques have been proposed. However, these methods usually present high computational complexity, given that the data points are repeatedly used in the process of determining change-points along the same sequence. Moreover, some practical issues arise when these methods are applied to real CNV data, such as handling the heavy-tailed distribution and identifying the biologically meaningful copy number states. In this study, we propose a modified screening and ranking algorithm (modSaRa). This algorithm is suitable for high-throughput genetic data due to its low computational complexity. The aforementioned issues in CNV detection are also addressed. First, modSaRa is robust to the violations of the normal assumption. Second, a novel approach using a normal mixture model coupled with a modified BIC criterion is proposed for filtering false positives and further clustering the potential CNV segments to copy number states. We compare modSaRa with an alternative method, circular binary segmentation (CBS). In both simulated and real data studies, modSaRa outperforms CBS in detecting CNVs with higher sensitivity and specificity.

1731F

Varpipe: Whole Genome Variant Analysis in Minutes. *A. Yao[1,2], E. C. Yeh[1], J. P. Su[1], Y. D. Chiu[1], D. H. Chung[1], W. J. Lin[1], L. Tang[1], F. C. Hsiao[1], K. C. Li[1], H. T. Wu[1], T. C. Chen[1], C. K. Liu[1].* 1) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 2) National Center for Genome Medicine, Academia Sinica, Taipei, Taiwan.

Analyzing variants in a human genome from raw data with 40X depth usually takes days. Setting up optimal analytical process to shorten analysis time at different scales is decidedly non-trivial. In addition, barriers such as limited resources and lack of user-friendly system are hindering the wider use of next generation sequencing (NGS) in human genome research. To address these issues, we developed Varpipe (http://hipipe. ncgm. sinica. edu. tw/varpipe), a web-based NGS analysis pipeline, which significantly reduces computation time and offers researchers a very efficient, scalable, accessible, and friendly system for whole genome sequencing projects. Varpipe uses a common computational pipeline that includes BWA for alignment, Sambamba for sorting and duplicate marking, Freebayes for variant calling, and SAMtools for final result concatenation. To improve performance, Varpipe adopts in-memory streamline processing and buffering to reduce disk input/output, local hard disk storage to reduce network traffic, and genome splitting to achieve good parallelization. Analysis of a test genome, 40X NA12878D human data set (Garvan Institute of Medical Research, DNAnexus and AllSeq), by splitting into 100 sub-regions showed that Varpipe reproduces 99. 97% SNPs and 99. 85% indels and completes analysis within 7 h using a single 64-core server (Dell R815). We further demonstrated that Varpipe, through distributed computing and splitting the genome into 380 sub-regions, completes variant analysis of the test genome in only 56 min with nearly identical results for both SNPs (99. 95%) and indels (99. 82%) on a 10-node 640-core cluster. Furthermore, our web-based Varpipe greatly reduces learning curve by providing carefully designed user interface and pre-configured analysis flow facilitating fast adaptation for users with minimum IT or bioinformatics background. In conclusion, not only can Varpipe analyze a genome in minutes, it also can provide good scalability for researchers to carry out efficient whole genome sequencing analysis at various scales easily.

1732W

FASTQuick: Comprehensive real-time quality assessment of ultra-high-throughput sequence data. *F. Zhang[1], H. M. Kang[1,2].* 1) Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Biostatistics, University of Michigan, Ann Arbor, MI.

The rapidly accelerating throughput of sequencing technologies allows us to sequence more than18,000 genomes a year at 30x coverage, producing 1 quadrillion sequenced bases from an Illumina HiSeq X Ten system. Efficient and thorough quality assessment of sequenced genomes in ultra-high-throughput scale is crucial for successful large-scale genetic studies, because delayed quality assessment in HiSeq X Ten can result in up to $100,000 of financial loss per day. Here we propose to develop a rapid and accurate set of algorithms and tools to produce comprehensive quality metrics directly from raw sequence reads prior to alignment. Our methods offer orders of magnitude faster turnaround time than existing alignment-based methods, while providing a truly comprehensive set of quality metrics, including estimates of genetic ancestry and contamination.

The key idea of our method is focusing on small regions of genome nearby known polymorphisms. Our method first extracts flanking sequences (250-1,000bp) around known variant sites and rapidly filters out more than99% of unalignable reads using Eland-like hashing technique. The filtered reads are aligned using computationally optimized version of bwa. As a result, compared to the widely used 1000 Genomes alignment pipeline, we now are able to reduce the computational time to generate thorough quality metrics on 30x genome from 160 hours to 43 minutes. Our results show that the quality metrics estimated our method is highly concordant to the quality metrics generated from the full-alignment pipeline.

Because our proposed methods collect information about genetic variation between individuals, it also provides accurate estimates of DNA contamination and genetic ancestry, in addition to the variety of typically quality metrics, such as base quality distribution, depth distribution, GC bias, and insert size distribution. By rapidly producing all these information together, our tool will help us detect potential problem in ultra-high-throughput sequence reads at a real-time speed, ensuring high quality sequence production and preventing unexpected loss in time and/or cost.

1733T

Identifying the underlying causal variants in associated regions in multiethnic meta-analysis. *A. B. Zhu[1], K. S. Burch[1], F. Hormozdiari[1], E. Eskin[1,2].* 1) Department of Computer Science, University of California, Los Angeles, California, 90095; 2) Department of Human Genetics, University of California, Los Angeles, California, 90095.

Meta-analyses combining the results of multiple genome-wide association studies (GWAS) have identified thousands of loci involved in dozens of human complex traits. A central challenge in the analysis of these data is to identify the actual causal variants which are responsible for these observed association signals. In many cases, there are multiple causal variants at each locus, and these variants themselves have different effect sizes in each GWAS study. Identifying these causal variants is complicated by complex patterns of linkage disequilibrium which may differ in each study. In this project, we present CAVIAR-Meta, a method for statistical fine mapping in meta-analyses. Our approach attempts to identify the minimum set of variants that, with high probability, contains all of the causal variants. Our method considers the possibility that each causal variant has a different effect size in each study by assuming a model closely resembling the random effects model, which is widely utilized in meta-analyses. We demonstrate the utility of our method through analysis of both simulated and real meta-analysis data.

1734F

Challenges in variant annotation for clinical genomic testing. *J. Yen, S. Garcia, S. Chervitz, B. Linebaugh, A. Montana, M. Morra, J. West, R. Chen, D. Church.* Personalis, Menlo Park, CA. , USA.

Sequencing based clinical testing is undergoing a revolution. While small gene panels have previously dominated the landscape, whole exome (WES) and whole genome (WGS) based strategies have recently emerged as valuable clinical tools, especially for diseases of unknown etiology or where there is extensive genetic heterogeneity. This expansion of WES and WGS based strategies in both scope and scale presents challenges for generating standard variant syntax, which is critical for building accurate and reliable resources for the clinical and scientific community. For example, the same indel can be represented in alternative ways in a VCF file. Analyses of dbSNP 142 identify over 350,000 variants that are represented in a non-normalized way (using software described by Tan et al. , 2015). Over 550 of these variants are also in ClinVar. In many cases, the normalization procedure reveals that the same variant can be represented in multiple ways, each with incomplete annotation. In addition to VCF file representation, challenges also arise in the translation of these variants to standard nomenclature recommended by the Human Genome Variation Society (HGVS), which was developed when sequence based testing was transcript rather than genome based. The same variant can have very different HGVS representation even on the same RefSeq transcript. In evaluating tools for generating HGVS nomenclature, we demonstrate challenges in providing variant syntax in a standardized, unambiguous format. We evaluated snpEff, Variation Reporter and Variant Effect Predictor (VEP) on a test set of 115 variants representing 5 variant classes with complex syntax representations at the genome, transcript and protein level, and their effect impact. While missense variants generally annotated correctly both at the transcript and protein level (all accurately described at least 28/32 of our test missense variants on all levels), the tools perform less far less robustly on frameshifting indels (one described 0/19 test variants correctly at the coding level). A closer look at the outputs reveal complexities in comparing preferred with non-preferred nomenclature, syntax representation across most indels and non-coding regions and where genomic reference differs from the RefSeq transcript. The results of our tests have significant implications for the search and annotation of variants during clinical testing and serve to inform the ongoing adoption and refinement of available tools.

1735W

Improving the annotation of splice-disrupting loss-of-function variants. *D. P. Birnbaum[1,2], K. J. Karczewski[1,2], D. G. MacArthur[1,2].* 1) Analytic and Translational Genetics Unit, Massachusetts General Hopsital, Boston, MA; 2) MacArthur Lab, Broad Institute, Boston, MA.

Disruption of transcript splicing underlies many cases of genetic disease, but identifying variants that alter splicing remains a challenging process. Much work has been done on developing probabilistic models for predicting how sequence variation will impact mRNA splicing, including tools such as MaxEntScan, Spliceman, and ESEfinder, but these tools have generally not been calibrated on large data sets of human genetic variation, and often cannot be scaled to the annotation of the many variants identified by modern large-scale sequencing studies. Here, we integrate splice variant prediction into the Ensembl Variant Effect Predictor (VEP) annotation framework. Specifically, we have written an improvement to LOFTEE, a loss-of-function annotation plugin to VEP, to identify the probability of splice disruption for variants outside the essential splice site, and also to check whether a putatively splice-disrupting variant may be rescued by a nearby alternative splice site. The plugin uses a probabilistic model to evaluate the usability of a potential "rescue" splice site by its local sequence context.

We validate these methods by applying them to variants from the Exome Aggregation Consortium (ExAC) as well as databases of known disease variants. Finally, we directly investigate the role of variants on splicing by intersecting exomes with matched RNA-Seq data from over 500 individuals from the GTEx consortium. The resulting tool can be readily scaled to the detection of candidate pathogenic splice variants in large datasets of human sequence variation.

1736T

Sequencing depth of coverage as a quality control metric in whole exome sequencing fails to identify multiple regions in which variant calling and genotyping cannot be accurately performed. *A. Naik, B. Kelly, P. White.* The Research Institute at Nationwide Children's Hospital , Columbus, OH.

Whole Exome Sequencing (WES) aims to identify potentially pathogenic variants in the protein coding regions of the genome, where it is estimated that 85% of the mutations responsible for Mendelian diseases are found. However, not all regions of the genome targeted are captured well, resulting in exonic regions that lack enough high quality read coverage to enable variant calling and accurate genotyping. A report of all poorly sequenced exons provides an important QC metric in clinical applications, where conventional Sanger sequencing may be needed to fill in for exons within clinically relevant genes. The conventional strategy is to use depth of coverage (DP) as the criteria for this purpose but we hypothesize that using genotype quality (GQ) would more accurately describe exonic regions missed by the assay. GQ is the Phred-scaled confidence that the predicted genotype is correct, considering the likelihood values of all possible genotypes. To create a report of missing exonic regions with both DP and GQ, we used the Churchill pipeline to efficiently perform variant calling, emitting genotypes for the entire genome in a genome Variant Call Format (gVCF) file. In combination with a manufacturer's targeted region BED file, the gVCF file is utilized to identify all regions that were targeted, but inadequately sequenced, at both no-confidence and low-confidence thresholds for DP and GQ. These regions are then annotated using NCBI's RefGene to obtain the specific targeted exons in genes not adequately and completely captured. The end-product is a detailed report of these exons, including annotations from OMIM, ClinVar and HGMD. Tests were run on the popular Coriell NA12878 reference sample captured using the Agilent SureSelectQXT Clinical Exome chemistry, sequenced to a depth of 100X. A total of 7,792 exons included at least one no-confidence base using GQ, as opposed to 6,809 exons using DP. This difference, 983 (3. 6% of the total exons targeted), represents exons with sufficient depth, but the quality of that depth was not such that confident genotypes could be called. Given GQ more adequately defines regions that cannot be accurately genotyped; it is our recommendation that DP not be used as a criterion for describing exonic regions missed by the assay.

1737F

Genotype Imputation Informed by both Kinship and Linkage Disequilibrium. *R. Wasiolek, J. Sinsheimer, K. Lange.* Biomathematics, UCLA, Los Angeles, CA.

Genotype imputation is increasingly important as more sequencing and genotyping technologies are developed and routinely employed. For example, GWAS analysis methods generally require large data sets with full SNP coverage. Genotype imputation fills in missing SNPs so that different genotyping sets can be used in the same study and untyped markers can be tested. Unfortunately, the majority of genotype imputation algorithms are slow. Another limitation is that these methods impute assuming individuals are unrelated, which leads to Mendelian inconsistencies when using data from families. A recently developed method by Chi et al employs a fast MM algorithm to impute missing SNP data based on matrix completion and linkage disequilibrium between neighboring SNPs by using a low rank approximation to the data but still assumes unrelated individuals. We extend this algorithm for family data by exploiting the kinship calculated using known pedigree relationships or from genetic relationship matrices. Our algorithm uses an MM algorithm and matrix completion to minimize an objective function that penalizes discordant genotypes among close relatives. Post-minimization, the singular value decomposition of the imputation matrix is taken and truncated after the *r* largest singular values to capture linkage disequilibrium. The values of r and the penalties are determined by cross-validation on a withheld subset of the observed data. Our preliminary results indicate that our algorithm is more accurate than the Chi et. al. MM algorithm, returning lower error rates and substantially fewer Mendelian inconsistencies with only minor losses in computational speed. Our algorithm also remains competitive in overall error rate and leads to substantiallyfewer Mendelian inconsistencies when compared with a second more accuratealgorithm that employs a Nesterov accelerated gradient method. Additionally, the inclusion of pedigree data allows us to employ a projection to nearest Mendelian consistency post-imputation which resolves more than 95 percent of the remaining Mendelian errors and further improves accuracy. Thus when imputing genotypes when family data are available, we recommend including kinship information to reduce Mendelian inconsistencies without sacrificing accuracy obtained from the LD. Our algorithm is currently implemented in Matlab.

1738W

Heterogeneity of Exome Methodology Causes Variability in CNV Predictions. *C. S. Hong[1], L. N. Singh[1], J. C. Mullikin[2,3], L. G. Biesecker[1,2].* 1) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Comparative Genomics Analysis Unit, Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 3) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

As exome sequencing (ES) becomes widely used in research and clinical settings, the effort to accurately identify copy number variations (CNVs) from the ES has been increasing. While many algorithms have been published to discover CNVs from exomes, no study has examined how variation in ES methodology affects the variability of CNV calls. In this study, we hypothesized that three confounders (independently-derived datasets, sample sizes, and heterogeneity in capture kits) would lead to variable CNV predictions, measured by the true positive rate. Four CNV tools were tested: eXome Hidden Markov Model (XHMM), Copy Number Inference From Exome Reads (CoNIFER), EXCAVATOR, and Copy Number Analysis for Targeted Resequencing (CONTRA). The CNV callers were greatly influenced by all three hypothesized attributes of exome data. The call reliability was strongly associated with the data source (p-value < 0. 05). The sample size was positively correlated with the number of CNV calls (χ> 0. 38) and also was either positively or negatively correlated with the true positive rate, dependent of CNV callers. In addition, the reliability (TP) was strongly associated with the capture kits used (p-value<0. 05). The results demonstrate that due to the intrinsic attributes of exome data, each data set is unique and that the reliability of CNV calls based on one study cannot be generalized for others. We conclude that training and testing the CNV callers on multiple independent datasets, in addition to improvements in algorithm, is a must in further advancing the wide applicability of CNV discovery in exome studies.

1739T

Comparison of a Literature Search Algorithm and Curated Publication Database with the Literature Content of Other Locus Specific Databases. *L. Esterling, D. DiFrancesco, P. Nix, JP. De La O, B. Coffee, S. Manley.* Myriad Genetic Laboratories, Inc. , Salt Lake City, UT.

Background: Effectively searching the scientific literature for publications providing evidence for the pathogenicity of a variant is critical in variant classification. This evidence may be crucial to the classification of a novel variant as either pathogenic or benign. Searches for relevant citations are complicated by the use of alternative nomenclatures for variants, gene names, and reference sequences. **Methods:** To ensure the most exhaustive search possible, our laboratory has developed an automated literature search algorithm coupled with a curated, searchable publication database of over 9500 citations crucial to the classification of variants affecting 25 genes included in a pan-cancer panel. Our automated literature search includes broad parameters to identify all references for a variant, which include all combinations of alternate gene names and variant designations. These references are reviewed by scientific experts, annotated, and linked to individual variants found in each publication. More than 80,000 variant references with evidence pertaining to classification have been identified and have been compiled by actively monitoring new literature and applying this automated literature search algorithm for newly identified variants. Using these tools, information pertaining to a variant's classification, as well as literature regarding allelic and surrounding variants, can be retrieved instantly during the classification process. This method has been used to facilitate the classification of over 27,000 single nucleotide and copy number variants in the past six years. To assess how comprehensive and effective this method is compared to other Locus Specific Databases, we compared variant references identified via our method to publications referenced in locus specific databases for 1554 previously classified and newly identified variants. **Results:** A total of 852 citations were identified for these 1554 variants using our method and locus specific databases combined. Our method identified 334% more citations than the locus specific databases combined. **Conclusions:** These results confirm that our literature search method and algorithm is more comprehensive than publicly available databases and illustrates the significant amount of time and resources that need to be dedicated to variant classification to provide physicians and patients the most accurate test results for clinical decisions.

1740F

Augmenting public databases with ultra-deep clinical targeted sequencing data. *E. A. Evans[1], M. D. Rasmussen[1], J. Maguire[1], K. R. Haas[1], H. P. Kang[1], D. G. MacArthur[2], I. S. Haque[1].* 1) Counsyl, South San Francisco, CA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA.

Clinical interpretation of sequence variants requires a deep understanding of the distribution and frequency of genetic variation in the general population. ClinVar solicits information from clinical labs to report variant-level statistics and clinical interpretations, but lacks comprehensive frequency information on controls. Research data sets (e. g. , 1000 Genomes and the Exome Aggregation Consortium (ExAC)) provide control information for most human genes but have gaps in coverage for known disease genes and are usually sequenced at a depth that limits analysis to SNPs and indels. In contrast, targeted clinical sequencing data has narrower genomic coverage, but can be designed to fully cover the genes of interest with sufficient depth to call CNVs and rearrangements. Here we describe the analysis of data from >100,000 samples tested using a clinical panel of up to 102 Mendelian disease genes. We used hybrid capture followed by sequencing on the Illumina HiSeq 2500 with manual call review (guaranteed mean depth >150x; min 99% bases covered >20x). Our analysis augments standard tools (BWA/GATK) with methods for haplotype, CNV, and high-homology-region calling. We clustered all samples into one of 15 groups using self-reported and genetic ancestry. We present the first sensitivity analysis for literature pathogenicity curation in clinical gene-panel sequencing. We classified all detected variants in 41,530 patients by literature review consistent with ACMG guidelines. 23% of patients carried >=1 variant on a whitelist chosen for an existing genotyping-based screen. Adding known-deleterious variants detected an additional 9% of patients; adding likely-deleterious variants detected another 6. 1% of patients; and adding predicted-deleterious variants detected a further 2% of patients.

We present a catalog of novel CNVs found in Mendelian disease genes. While current sequencing databases are limited by their underlying depth (on our panel ExAC mean coverage ranges from 3. 29 to 77. 9), our ultra-deep data enable new findings.

Finally, we describe the integration of this deep clinical sequencing data into the ExAC browser, more than doubling the sample size in 100 clinically relevant genes and making this data available to the wider community. This is the first example of a clinical lab contributing a large amount of control variant frequency data to a public sequencing database, and outlines a process by which commercial data can be made accessible in an open system.

1741W

From genes to functional elements - enriching RefSeq annotation of the human genome. *C. M. Farrell, K. D. Pruitt, T. D. Murphy, D. R. Maglott, T. Goldfarb, S. H. Rangwala, RefSeq Group.* National Center for Biotechnology Information, NLM/NIH, Bethesda, MD.

The RefSeq database at NCBI (www. ncbi. nlm. nih. gov/refseq/) provides reference sequences for transcripts, proteins and genomes, and its diverse uses include genome annotation, gene identification, variation calling, and functional characterization. Human RefSeqs are used in clinical practice, with the RefSeqGene/Locus Reference Genomic (LRG) collaboration providing a foundation for reporting variants according to the Human Genome Variation Society (HGVS) standard. To date, the RefSeq group has primarily focused on defining the sequences of genes, pseudogenes and gene products, with each gene having a record in NCBI's Gene database (www. ncbi. nlm. nih. gov/gene/). However, genes occupy only a small fraction of the genome, and the sequences of non-genic functional elements need to be defined. Mutations in these elements can lead to human diseases, with many genome-wide association studies (GWAS) pointing to variants in non-annotated intergenic regions. Initiatives such as the ENCODE and Roadmap Epigenomics projects have provided ways to predict the presence of functional elements on the reference genome, but specialized research knowledge and the use of customized genome browsers or tracks are required to assess such elements. Thus, these elements may not be apparent to all users in biomedical research. The RefSeq group has therefore initiated a project to provide reference sequences for functional elements in both human and mouse, and to annotate them on the reference genome alongside our current gene annotations. The scope of this annotation is for elements that are experimentally validated in the literature, including known gene regulatory elements (e. g. , enhancers), elements involved in higher-order genome organization (e. g. , boundary elements), and elements that are otherwise considered to be of functional importance (e. g. , recombination hotspots). Curated records will also be provided for these elements in NCBI's Gene resource to include summaries, functional attributes, and various metadata from the literature and public databases. These functional elements will enrich the current scope of our genome annotation, they will be particularly useful for GWAS and variant interpretation outside of gene boundaries, they will be highly visible and accessible to a wide user base, and are expected to be valuable to research in general. This work was supported by the Intramural Research Program of the National Library of Medicine, NIH.

1742T

Curation of reference sequence (RefSeq) standards to support clinical applications and basic research. *K. Pruitt, C. M. Farrell, M. J. Landrum, D. R. Maglott, T. D. Murphy, RefSeq Curation Group.* NCBI, National Institutes of Health, Bethesda, MD.

The National Center for Biotechnology Information (NCBI) RefSeq project (www. ncbi. nlm. nih. gov/refseq/) provides reference sequence and annotation standards for genomes, transcripts, proteins, gene regions, and pseudogenes. The RefSeq project provides functional and structural annotation of viral, prokaryotic, and eukaryotic RefSeq genomes and supports myriad other investigations ranging from expression studies to clinical reporting of genetic mutations to variation calls. RefSeq records for human (and mouse) include genome annotation for reference and other assemblies, >50,000 (>25,000) curated transcripts and >37,000 (>22,000) proteins plus model transcripts and proteins inferred from protein and transcript data (including RNA-seq) aligned to genomes. The curated RefSeq dataset is of high quality following 15 years of sequence and gene curation by NCBI scientific staff and participation in international collaborations which provide additional scientific review of the data. RefSeq collaborates with the Genome Reference Consortium (GRC), which maintains and updates the human and mouse reference genome sequence (www. ncbi. nlm. nih. gov/projects/genome/assembly/grc/); RefSeq curators request GRC review of sequence regions that conflict with aligned transcripts or publications and review some GRC proposed sequence updates that may impact gene annotation. RefSeq proteins and their genome annotation coordinates are further reviewed by partner members of the Consensus CDS (CCDS) collaboration which aims to stabilize and reconcile conflicts between the human and mouse reference genome protein-coding gene annotation that is provided by the NCBI and Ensembl genome annotation pipelines (www. ncbi. nlm. nih. gov/CCDS/), both of which include curated content from the RefSeq and GENCODE projects. The CCDS project is engaged in ongoing review of conflicts and must reach consensus agreement to change annotation coordinates for proteins having a CCDS ID. In addition, the RefSeqGene/Locus Reference Genomic (LRG) collaboration defines genomic sequence regions for human genes which serve as a stable foundation for reporting mutations in HGVS format, for establishing conventions for numbering exons and introns, and for defining the coordinates of other variations (www. ncbi. nlm. nih. gov/refseq/rsg/). Over 5,400 RefSeqGene sequence records (576 with an LRG ID) are available. The presentation will describe RefSeq, our collaborative approach to providing it, and avenues for data access.

1743F

n_1n_2TrackAnnotator: Software for detection and annotation of tracks comprising user-defined two different nucleotides in genome sequence. *H. N. Singh, M. R. Rajeswari.* Biochemistry, All India Institute of Medical Sciences, New Delhi, Delhi, India.

Chromosomal DNA exists mainly as a right-handed double helix form known as B-DNA. However, other conformations, such as triplexes, tetraplexes, slipped structures with hairpin loops, cruciforms etc. are also known that were found to be primarily associated with gene regulation processes such as replication and transcription blocking as well as genetic rearrangements. Identification and annotation of these repeat sequence which leads to form non-B DNA structures, is the decisive task in order to understand the mechanisms of the genetic instability and disease correlations. Therefore, we developed **"n1n2TrackAnnotator",** a standalone software for genome-wide analysis to identify and annotate tracks containing two different nucleotides in the genome sequences. It provides a complete open-source pipeline on Windows platform comprising all steps from importing of raw data (genome sequence, annotation file: General Feature Format), pre-processing of data, annotation of the identified sequences. n1n2TrackAnnotator is written as a standalone tool using C# 4. 0 object oriented programming language at . NET Framework 4. 0 for Windows platform. In particular, it implements a knowledge based window-shift algorithm and supports NCBI/Genome database information patterns. .

1744W

A Bayesian test to identify variance quantitative trait loci. *B. Dumitrascu[1], G. Darnell[1], J. Ayroles[1,2], B. Engelhardt[3,4].* 1) Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ; 2) Department of Organismic and Evolutionary Biology, Harvard University,Cambridge, MA, USA; 3) Department of Computer Science, Princeton University, Princeton, NJ, USA; 4) Center for Statistics and Machine Learning, Princeton University, Princeton, NJ, USA.

Identifying genotypic variants that are associated with variation in quantitative traits is a primary focus of statistical genetics. However, few studies successfully address the problem of genome-wide identification of variants that have an effect on the variance of a trait instead of the trait mean across a population. Up to now, there has not been a general and robust approach to test for these variance controlling QTLs, or vQTLs. In this study, we develop BvQTL, a Bayesian framework for identifying vQTLs using a heteroskedastic linear regression model. We develop fast approximation algorithms for the statistical testing framework to scale the test to trivial genome-wide application. We show through simulations that our method outperforms classical statistical tests for covariates that affect variance such as the Levene and Brown-Forsythe tests in diverse genomic situations, including at loci with low minor allele frequency or mean effects. In particular, we show that BvQTL is a robust framework that produces fewer false negatives at a consistent false positive rate and performs well on small samples and on imputed genotypes. As empirical validation of our method, we apply BvQTL to data from the Cardiovascular and Pharmacogenetics (CAP) study to test for dependence between variance in gene expression levels and genotypes. We show that our approach works for arbitrary quantitative traits by finding vQTLs in methylation study data, representing genetic variants associated with variance of methylation levels at specific CpG sites, from 255 genotype-annotated samples from the HapMap2 study. We expect that, in the near future, genome-wide studies for quantitative trait association will include tests for vQTLs as part of the standard procedure.

1745T

An improved infrastructure and user interface for the NHGRI-EBI Genome-Wide Association Study (GWAS) Catalog. *J. A. L. MacArthur[1], T. Burdett[1], F. Cunningham[1], L. Gil[1], P. N. Hall[2], E. Hastings[1], L. A. Hindorff[2], H. A. Junkins[2], A. K. Klemm[2], J. Morales[1], D. Welter[1], H. Parkinson[1].* 1) European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom; 2) Population Genomics, NHGRI, NIH, Bethesda, MD.

The GWAS Catalog (Welter et al. , 2014) is a manually curated, publicly available resource of all published GWAS and trait-associations, containing over 2,000 studies and 17,000 SNP-trait associations, as of June 2015. It was originally created by the National Human Genome Research Institute (NHGRI) in 2008 and has been collaboratively produced and developed by the NHGRI and European Bioinformatics Institute (EMBL-EBI) since 2010. In March 2015 the GWAS Catalog was given a new website (http://www. ebi. ac. uk/gwas/) and infrastructure at EMBL-EBI to deliver an improved user experience and long-term support to the Catalog into the future. The new website includes a search interface driven by an ontology-enriched Solr index. Search results are displayed in interactive tables with dynamic links to related GWAS Catalog data and to external resources, such as Ensembl and Europe PubMed Central. Search results can also be sorted, filtered and downloaded. The redesign allows for the improved capture of complex GWAS study design and association data, such as the independent representation of loci in SNPxSNP interaction studies. A new curation platform has been developed, supporting the accurate recording of study design and results, improved quality control processes and scaling of data curation into the future. The increased curatorial capacity will improve our ability to tackle the current backlog of eligible studies. Several enhancements have been made to the scientific content of the Catalog: newly released ancestry information for 2011 studies onwards is available, while mapping information has been added for all GWAS Catalog SNPs to Ensembl genes, in addition to NCBI genes. Informatics improvements have facilitated the integration of GWAS Catalog data and will enable the development of tools to identify causal variants and understand functional relationships between variants and phenotypes. In the future we aim to allow users to define regions of association using linkage disequilibrium data and to explore genes and regulatory elements within these regions. The improved curation, modelling, querying and integration of GWAS Catalog data has vastly improved the utility of this resource in identifying and understanding disease loci. The improvements to curation efficiency and scalability will ensure that the resource is sustainable and can meet the demands of increasing data complexity and volume.

1746F

Meta-analysis of Complex Diseases at Gene Level by Functional Regression. *Y. Wang[1], W. Chen[2], D. E. Weeks[2], H. Ren[3], Y. Li[4], I. Lobach[5], C. I. Amos[6], J. H. Moore[7], M. Boehnke[8], M. Xiong[9], R. Fan[10].* 1) Food and Drug Administration (FDA), Silver Spring, MD; 2) University of Pittsburgh; 3) Regeneron Pharmaceuticals, Inc; 4) University of North Carolina, Chapel Hill; 5) University of California, San Francisco; 6) Dartmouth Medical College; 7) University of Pennsylvania; 8) University of Michigan, Ann Arbor; 9) University of Texas - Houston; 10) Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH).

We develop functional regression models (FRMs) to perform meta-analysis of multiple studies to evaluate the relationship of genetic data to quantitative or dichotomous traits adjusting for covariates by using functional data analysis techniques. Unlike the previously developed MetaSKAT methods, which are based on mixed effect models to model the major gene contributions from loci as a random effect, FRMs are fixed effect models and genetic effects of multiple genetic variants are assumed to be fixed. Based on the FRMs, we developed test statistics to test for an association between a complex trait and multiple genetic variants in one genetic region. We then performed extensive simulations to evaluate the empirical type I error rates and power performance of the proposed models and tests. The proposed test statistics control the type I error very well and are conservative, and have higher power than MetaSKAT in most cases. Specifically, we show: (1) the proposed test statistics have higher power than MetaSKAT for quantitative traits no matter some causal variants are rare and some are common or all causal variants are rare, (2) the proposed test statistics have higher power than MetaSKAT for dichotomous traits when some causal variants are common and some are rare, and (3) when the causal variants are all rare (i. e. , minor allele frequencies less than 0. 03), the proposed test statistics have similar or slightly lower power than MetaSKAT for dichotomous traits. For quantitative traits, one paper is accepted by ``Genetics", whichis available in Early Online, June 9, http://www. genetics. org/content/early/recent. In our unpublished work, the proposed methods were applied to analyze genetic data of 22 gene regions of type 2 diabetes data from a meta-analysis of eight European studies. These 22 genes are all from literature and each of them contains SNPs/variants which are related to the trait by single SNP/variant analysis in various studies. Hence, they can be treated as candidate genes and can be used to test the performance of gene based tests. The proposed methods detected significant association for 18 genes (p-values $< 3. 10 \times 10\text{-}6$) and tentative association for 2 genes (p-values around 10-5) and no association for 2 genes, while MetaSKAT detected none. The models and related test statistics can analyze rare variants or common variants or a combination of the two, and can be useful in whole genome-wide and whole exome association studies.

1747W

Churchill 2. 0: Making the Ultra-Fast Analysis Pipeline for Clinical and Population-Scale Genomics Faster, More Efficient, and More Comprehensive. *B. Kelly, J. Fitch, H. Kuck, P. White.* The Research Institute at Nationwide Children's Hospital, Columbus, OH.

Next generation sequencing (NGS) has revolutionized genetic research, enabling dramatic increases in the discovery of functional variants in syndromic and common diseases. Declining sequencing costs are enabling the study of many 1,000's of samples, and this exponential growth in data generation has created a computational bottleneck. Current analysis approaches can take days to complete, resulting in bioinformatics overheads that exceed sequencing costs and represent a significant limitation. Churchill is a computational approach that overcomes these challenges, fully automating the steps required to take raw sequencing data through the complex and computationally intensive processes of alignment through variant discovery. Compared with alternative analysis pipelines, Churchill is faster, deterministic and 100% reproducible. Through novel parallelization techniques, Churchill enables efficient whole genome sequencing analysis in less than two hours, ideal for clinical applications where turnaround time is paramount and data quality cannot be sacrificed. Churchill also enables population-scale analyses to be performed cost-effectively, efficiently distributing the analysis steps and processing data sets that would otherwise have been too large to analyze at a reasonable cost and in a practical time frame. Published in January 2015, Churchill is now being utilized by a large community within both the research and clinical domains. Since then we have made significant enhancements to the original release of the Churchill pipeline. Churchill 2. 0 expands the novel balanced subregional parallelization strategy to structural variant (SV) detection resulting in significantly reduced SV analysis time. Churchill 2. 0 has been redesigned to be highly modular, allowing users to incorporate different aligners, variant callers, and other components, creating custom pipelines that take advantage of Churchill's parallelism and speed. Churchill 2. 0 is more robust and fault tolerant, providing meaningful QC metrics and allowing a user to restart an analysis from the point at which an error occurred. As we look to the future, the era of the $1,000 genome will exponentially increase the size and number of genomic studies. Churchill 2. 0 makes significant and necessary improvements on an already fast, accurate and deterministic pipeline, enabling comprehensive, fully-automated analyses and eliminating the computational bottleneck in clinical and population-scale genomics.

1748T

NCBI resources for taking advantage of the GRCh38 reference genome assembly. *V. A. Schneider, V. Ananiev, N. Bouk, C. Clausen, M. DiCuccio, B. Holmes, A. Kuznetsov, P. Meric, T. Murphy, L. Phan, M. Ward.* IEB, NIH/NCBI, Bethesda, MD.

The Genome Reference Consortium (GRC) released GRCh38, the current version of the human genome reference assembly, in December 2013. The updated assembly contains many improvements, such as gap closures, path corrections and individual base updates, as well as new features, such as modeled centromere sequence representations. It also contains a large cohort of alternate loci scaffolds, which provide alternative sequence representations for more than 200 genomic regions. Despite a general recognition that the assembly update represents an improved substrate for analysis, the need to remap datasets to new coordinates, uncertainty about what regions have changed and unfamiliarity with alternate loci are challenges that have slowed the adoption of GRCh38 by scientists engaged in both basic and clinical research. Several NCBI tools exist to help users make the transition to GRCh38 and take advantage of the alternate loci. We will present features of NCBI resources such as Variation Viewer and MapViewer, used for genome browsing, that enable users to identify and view assembly differences and regions with alternate sequence representation. We will also discuss new browser features associated with track management and display, as well as support for viewing user uploaded data that improve the user experience with NCBI browsers in general. The NCBI Remapping Service allows users to project features not only between different assembly versions, but also between chromosomes and alternate loci scaffolds. We will describe recent updates to the Remapping Service that improve variant re-mapping and discuss features of the NCBI assembly-assembly alignment algorithm that were developed to specifically support the structure of the GRCh38 assembly with its alternate loci. We will illustrate how assembly-assembly alignment reports available from the remapping service provide users with a convenient catalog of differences between the current and previous reference assembly versions. Together, these and other NCBI resources will help researchers take advantage of the improvements offered in the current reference genome assembly.

1749F

High-accuracy imputation for *HLA* class I and II genes based on high-resolution SNP data of Japanese references and its application on Steven-Johnson Syndrome and Narcolepsy with cataplexy. *S. Khor[1], W. Yang[2], M. Kawashima[1], S. Kamitsuji[2], X. Zheng[3], N. Nishida[1,4], H. Sawai[1], H. Toyoda[1], T. Miyagawa[1], Y. Hitomi[1], M. Honda[5,6], M. Ueta[7], S. Kinoshita[7], N. Kamatani[2], K. Tokunaga[1].* 1) Department of Human Genetics, Graduate School of Medicine,, University of Tokyo, Tokyo, Tokyo, Japan; 2) StaGen Co. , Ltd. , Statistical Genetics Analysis Division, Tokyo, Japan; 3) Department of Biostatistics, University of Washington, Seattle, WA, United States; 4) The Research Center for Hepatitis and Immunology, National Center for Global Health and Medicine, Ichikawa, Japan; 5) Department of Somnology, Tokyo Medical University, Tokyo, Japan; 6) Sleep Research Project, Department of Psychiatry and Behavioral Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan; 7) Department of Frontier Medical Science and Technology for Ophthalmology, Kyoto Prefectural University of Medicine, Kyoto, Japan.

Statistical imputation of classical human leukocyte antigen (*HLA*) classes is becoming an indispensable tool for fine-mappings of disease association signals from case-control genome-wide association studies (GWASs). However, most currently available *HLA* imputation tools are based on European reference populations and are not suitable for direct application to non-European populations. Among the *HLA* imputation tools, the HIBAG R package is a flexible *HLA* imputation tool that is equipped with a wide range of population-based classifiers; moreover, HIBAG R enables individual researchers to build custom classifiers. Here, two datasets, each comprising data from healthy Japanese individuals of difference sample sizes, were used to build custom classifiers. *HLA* imputation accuracy in five *HLA* classes (*HLA-A*, *HLA-B*, *HLA-DRB1*, *HLA-DQB1*, *HLA-DPB1*) increased from the 82. 5-98. 8% obtained with the original HIBAG references to 95. 2-99. 5% with our custom classifiers. A call threshold (CT) of 0. 4 is recommended for our Japanese classifiers; in contrast, HIBAG references recommend a CT of 0. 5. Moreover, our classifiers could be used to identify the risk haplotypes for Japanese narcolepsy with cataplexy, *HLA-DRB1*15:01* and *HLA-DQB1*06:02*, with 100% and 99. 7% accuracy, respectively. Finally, the risk *HLA* alleles (*HLA-A*02:06* and *HLA-B*44:03*) for multiple population Steven-Johnson Syndrome (SJS) can be correctly imputed with an accuracy of 59/59 and 28/29 in 117 patients who carry the risk *HLA* alleles respectively. Therefore, these classifiers can be used to supplement the current lack of *HLA* genotyping data in widely available GWAS datasets.

1750W

PhenomeCentral: a portal for phenotypic and genotypic matchmaking of patients with rare genetic diseases. *O. J. Buske[1,2,3], M. Girdea[1,2,3], S. Dumitriu[3], B. Gallinger[3,4], T. Hartley[5], H. Trang[3,4], A. Misyura[3], T. Friedman[1], C. Beaulieu[5], W. P. Bone[6], A. E. Links[6], N. L. Washington[7], M. A. Haendel[8], P. N. Robinson[9], C. F. Boerkoel[6], D. Adams[6], W. A. Gahl[6], K. M. Boycott[5], M. Brudno[1,2,3].* 1) Computer Science, University of Toronto, Toronto, Ontario, Canada; 2) Genetics and Genome Biology Program, The Hospital for Sick Children, Toronto, Canada; 3) Centre for Computational Medicine, The Hospital for Sick Children, Toronto, Canada; 4) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario Canada; 5) Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario, Canada; 6) Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, Maryland, USA; 7) Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA; 8) Department of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University, Portland, Oregon, USA; 9) Institute for Medical Genetics and Human Genetics, Charité-Universitàtsmedizin Berlin, Berlin, Germany.

The discovery of disease-causing mutations typically requires confirmation of the variant or gene in multiple unrelated individuals, and a large number of rare genetic diseases remain unsolved due to difficulty identifying second families. To enable the secure sharing of case records by clinicians and rare disease scientists, we have developed PhenomeCentral (https://phenomecentral. org), a restricted access portal for sharing de-identified patient phenotype and genotype data to find similar cases around the world. Each record includes a detailed phenotypic description (specified using HPO terms) and relevant genetic information (exome data and/or candidate genes). PhenomeCentral identifies similar patients in the database based on semantic similarity between clinical features, automatically prioritized genes from whole-exome data, and candidate genes entered by the users, enabling both hypothesis-free and hypothesis-driven matchmaking. Additionally, users can opt to include patient records in matchmaking across the Matchmaker Exchange, enabling them to see similar patients not only within PhenomeCentral but also within other patient databases (including GeneMatcher and DECIPHER). Users can then contact other submitters to follow up on promising matches. PhenomeCentral incorporates data for over 1,250 patients with rare genetic diseases, contributed by the FORGE and Care4Rare Canada projects, the US NIH Undiagnosed Diseases Program, the EU Neuromics and ANDDIrare projects, as well as numerous independent clinicians and scientists. Though the majority of these records have associated exome data, most lack a molecular diagnosis. PhenomeCentral has already been used to identify causative mutations for several patients, and its ability to find matching patients and diagnose these diseases will grow with each additional patient that is entered. The coordination with the Matchmaker Exchange further increases the number of potential matches. With additional sites planning to join the Matchmaker Exchange in the near future, storing deep phenotype and genotype data for all patients in PhenomeCentral will help ensure the maximum potential for matchmaking for these rare disease patients.

1751T

NCBI's database of Genotypes and Phenotypes: dbGaP. *M. Feolo, S. Pretel, Z. Wang, M. Bihan, S. Sherry, NCBI's dbGaP staff.* National Center for Biotechnology Information, National Institutes of Health, USA.

The explosion of data from genome scale studies is well underway and the need to archive these data is growing rapidly. The National Center for Biotechnology Information's database of Genotypes and Phenotypes (dbGaP) is an NIH sponsored repository charged to archive, curate and distribute information produced by genome scale studies investigating the interaction of genotype and phenotype. The data submitted to dbGaP includes: individual level molecular and phenotype data; analysis results; medical images; general information about the study; and documents that contextualize phenotypic variables, such as research protocols and questionnaires. The molecular data includes array-based, sequence-based or imputed genotypes; expression, as well as next-generation sequencing (NGS) performed to produce whole exome; whole genome; RNA seq; and epigenomic data. NGS sequence submitted to dbGaP, mostly in the form of BAM files, are archived and accession by NCBI's Sequence Read Archive (SRA). Phenotypic data pertaining to Clinical or Research study participants and their samples consists of any combination of cross-sectional or longitudinal demographic; clinical; laboratory; exposure; or treatment variables. Medical images such as CT; MRI; and retinal scans of the eye are also stored and distributed by dbGaP. Analysis results include summary-level statistical measures of the association between the phenotypic variables and the molecular data. All documents, analysis, phenotypic and molecular data are validated for consistency and format though a computational and human curatorial process. Validated submission are then accessioned with stable, unique identifiers that support update of the same study with successive genetic investigations. These identifiers make it possible to cite the data that were downloaded from dbGaP and used to generate new research results, in a very specific and stable way. The dbGaP provides unprecedented access to a vast array of genome scale studies, both funded by National Institutes of Health, and other funding agencies worldwide. Through controlled access, researchers from across the globe may obtain complete statistical analyses, subject-level phenotype and molecular measures of over 900,000 subjects. This presentation will describe the data content, access procedure and highlight several new features of dbGaP.

1752F

Development of a bioinformatics application for the assessment of genetic risk of complex diseases. *K. L. Valdés[1], F. F. González-Galarza[1,2], R. D. Arellano-PerezVertti[1], E. Ramírez-Ramírez[1], A. Y. García-Marín[2], D. Delgadillo-Guzmán[1], J. Morán-Martínez[1], J. E. Gaytan-Arocha[1], J. R. Argüello[1,2].* 1) Faculty of Medicine, Autonomous University of Coahuila, Torreon, Mexico; 2) Institute of Science and Genomic Medicine, Torreon, Mexico.

In recent years, the number of studies in the genomics field has increased at a rapid pace. Well-structured genome-wide association studies (GWAS) have determined a significant number of SNPs associated to different complex diseases and traits. However, there is still a gap on the translation of this information into medical practice. For instance, the role of low density lipoprotein (LDL), high density lipoprotein (HDL), and triglycerides in cardiovascular disease is well established. However, in the clinical context, the genetic component of these complex traits is commonly disregarded. In this study, we designed a bioinformatics application for the analysis of dyslipidemia-associated SNPs which were extracted from the Catalog of Published GWAS. To test our software, we analyzed one hundred randomly-selected individuals from the North Area of Mexico that were genotyped using the Illumina microarray technology with approximately 600,000 SNPs. Using an additive genetic risk score model, we calculated an accumulated odds ratio and an accumulated p-value for each individual and estimated a genetic risk factor for three complex traits: elevated levels of LDL and triglycerides, and decreased levels of HDL. Here, we provide a user-friendly bioinformatics application for the analysis of genetic risk which could be applied in clinical setting to different health conditions involving complex traits such as dyslipidemias. Ongoing investigations are being performed for futher validation of these algorithms, including the non-genetic components, such as clinical and environmental factors.

1753W

An empirical recombination for demographic inference and IDB detection. *T. Y. Wang, J. H. Loo, M. Lin.* Mackay Memorial Hospital, New Taipei City, Taiwan.

Genome-wide data facilitate the investigation of genomic relatedness between individuals within or across populations, providing an insight into demographic histories. Genomic regions of identity by decent (IBD) in individuals, co-inherited from common ancestors, can be detected and analyzed to reveal genetic relatedness for demographic inference. Many methods have recently been developed to detect IBD regions, aiming at detecting identical regions that are statistically unlikely to occur without common ancestors. Some employ coalescent or probabilistic models to identify such IBD regions with significantly low frequencies of occurrence; the others use non-coalescent and non-probabilistic models to detect IBD regions with long lengths, which serve as proxies for low frequencies. However, due to high computational cost of coalescent or probabilistic models, the first ones are usually not fit to large datasets, and because of no short IBD regions detected, the last ones cannot provide comprehensive information for demographic inference. We propose an empirical approach that is able to infer demographic histories and to detect IBD regions simultaneously. This approach comprises of an empirical model of recombination and an IBD detection algorithm. The empirical model builds coalescent trees with recombination events based on genomic similarities of individuals, and the detection algorithm incorporates the information of coalescent trees with recombination events to identify IBD regions. These two procedures can be executed iteratively till no new IBD regions found and no new changes in coalescent trees. In addition, the two procedures can be in parallel in each iteration to improve computational efficiency. We applied our method in simulated data and two real datasets: the 1,000 genomes and the HLA alleles in Taiwan populations. First in simulation analysis, our method is able to infer demographic histories and to detect short IBD regions with high accuracy while maintaining high computational efficiency. Second in the 1,000 genome dataset, our approach not only reveals recent demographic events based on long detected IBD regions, but also ancient histories from short IBD regions. Finally in the HLA alleles in Taiwan populations, we demonstrate the pure utility of the empirical recombination model for recent demographic inference. Therefore, our proposed method is capable of detecting IBD regions efficiently and making demographic inference comprehensively.

1754T

Smart multiple testing correction in eQTL studies. *D. B. Duong[2], B. Han[1], E. Eskin[2].* 1) Asan Institute for Life Sciences, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Republic of Korea; 2) Computer Science , UC Los Angeles , Los Angeles , CA.

In expression quantitative trait loci (eQTL) studies, it is of tremendous interest to identify eGenes, the genes whose expression levels are associated with particular genetic variants. The common procedure for detecting eGenes is to test association of each nearby variant (cis variants) to the expression level, and select the minimum p-value among all p-values. As we perform multiple tests, the minimum p-value has to be corrected for multiple hypothesis testing. A widely used procedure for this correction is the permutation test. Recently, we proposed an approach called eGene-MVN that can obtain the same correction results as the permutation test in a dramatically reduced time (Sul et al. , AJHG 2015). However, this approach only improves the efficiency of the permutation test, but does not improve the p-values of the permutation test. The limitation of both the permutation test and eGene-MVN is that they do not incorporate prior information of the variants. For example, we have prior knowledge that variants near transcription start site (TSS) are more likely to regulate expressions. We propose a new multiple testing correction approach for eQTL studies, called eGene-Smart. Our approach learns how large an effect a variant would have for regulation from the observed data in a similar strategy to cross-validation. Then based on these learned prior parameters, our approach weights each variant differently in multiple testing correction. Specifically, the corrected p-value becomes more significant if the up-weighted variant is driving the association signal rather than the down-weighted variants. In simulations, we show that our approach gives better power than the previous approaches, while appropriately controlling the overall false positive rate. We apply our approach to the Genotype-Tissue Expression (GTEx) data and show that our approach dramatically improves eGene p-values for significantly associated eGenes. We expect that our approach will be widely used in future eQTL studies to improve p-values and effectively identify eGenes.

1755F

Kaviar: a comprehensive public catalog of human variant and genotype frequencies. *T. Farrah[1], D. E. Mauldin[1], N. Clegg[1], M. Robinson[1], J. Vockley[2], J. Niederhuber[2], L. Hood[1], G. Glusman[1].* 1) Institute for Systems Biology, Seattle, WA; 2) Inova Translational Medicine Institute, Falls Church, VA.

Kaviar, first released in 2010, is now the largest publicly available catalog of whole genome human allele frequencies. Obtaining established allele frequencies of observed genetic variants is a key step in genome interpretation, and allele frequency accuracy depends on the size and diversity of the source data. Kaviar combines data from 37 sources including 5000+ whole genomes sequenced by the Inova Translational Medicine Institute (inova. org/itmi) and the Institute for Systems Biology (familygenomics. systemsbiology. net). It provides allele frequencies for 168 million SNVs, twice as many as the 1000 Genomes Project. In addition to the private whole genomes, Kaviar includes whole genome data from the 1000 Genomes Project, the Alzheimer's Disease Neuroimaging Initiative (ADNI), the Simons Diversity Project, the Wellderly project, the UK10K, Genomes of the Netherlands, and the Personal Genomes Project, and also incorporates the 63,000 exomes from the Exome Aggregation Consortium (ExAC).

Kaviar uses a comprehensive pipeline to integrate sources in a manner that increases concordance, including normalization of all variants to enforce parsimony and left alignment. Based on 194 genomes sequenced on both Illumina and Complete Genomics platforms, we remap variants that are called frequently on one platform but never called on the other. Kaviar also reports population-specific frequencies computed by inferring the continent(s) of origin for each data source.

Kaviar now reports *genotype frequency* in addition to allele frequency, which can be useful in identifying genomic regions that require more inspection. Variants deviating from Hardy-Weinberg equilibrium may suggest errors in the genome reference, in sequencing, in analytics, or population structure. Such deviations may also point toward selection signatures. A variant never seen as homozygous may have a recessive effect. Differences in genotype distributions among different data sources may uncover differences in sequencing technology, differences in variant representation, or reference biases.

We propose Kaviar as the community's standard resource for human allele and genotype frequencies. Kaviar is available via web interface, downloadable VCF files, and GA4GH beacon at db. systemsbiology. net/kaviar.

1756W

Single Sample Imputation from Next Generation Sequencing (NGS) Exome data can improve genotypes in low-coverage regions. *H. Sicotte[1], N. Prodduturi[1], J. A. Johnson[2], Y. Lin[2], P. A. Decker[1], J. E. Eckel-Passow[1], R. B. Jenkins[3], M. P. Goetz[4,5], R. Weinshilboum[4], J. C. Boughey[6], L. Wang[4], K. R. Kalari[1], S. K. McDonnell[1], M. de Andrade[1], J. P. Kocher[1].* 1) Department of Health Sciences Research, Mayo Clinic, 200 First Street SW, Rochester, MN, 55905; 2) Mayo Clinic Research and Education Support Systems, Information Technology Rochester, Mayo Clinic, 200 First Street SW, Rochester, MN, 55905; 3) Laboratory Medicine and Pathology, Mayo Clinic College of Medicine, Mayo Clinic, 200 First Street SW, Rochester, MN, 55905; 4) Department of Molecular Pharmacology & Experimental Therapeutics, Mayo Clinic, 200 First Street SW, Rochester, MN, 55905; 5) Medical Oncology, Mayo Clinic, 200 First Street SW, Rochester, MN, 55905; 6) Department of Surgery, Mayo Clinic, 200 First Street SW, Rochester, MN, 55905.

Background: In the research setting, the high cost of NGS whole genome sequencing (WGS) has driven the use of low-coverage sequencing or whole exome sequencing (WES). However, WES genotypes have large site to site variation because of the uneven coverage of the targeted exons, or complete lack of coverage outside of the target region. We wanted to determine the usefulness of genotype imputation to obtain genotypes in region of poor or non-existent coverage in WES data. NGS data is different from the typical input of imputation (microarray genotypes) and may require different quality control steps to insure accurate imputation. Also, because NGS data is usually delivered in small or single-sample batches, we wanted to study single-sample imputation to best fit imputation in current NGS processing workflows. **Results:** Single-sample imputation is made possible by a recent development in imputation, namely phasing using a haplotype reference. We added a single-sample imputation option to EZimputer, our imputation workflow based on Shapeit2 and Impute2. We evaluated the accuracy of imputation from exome genotypes using genotypes from an illumina OMNI array on germline samples from the Breast Cancer Genome Guided Therapy (BEAUTY). We found that both variable and homozygous reference variants from the imputation reference, such as the 1000 genomes, need to be used as input genotypes to the workflow. We found that imputation accuracy is improved by filtering genotypes used in the imputation. We found that single-sample imputation against the reference is more accurate than multi-sample imputation for small sample size. We also evaluated two tools (HARSH and Shapeit2) that take advantage of NGS-specific information, namely co-occurrence of variants on the same sequencing reads or fragments. Overall, these two tools did not improve the imputation when using our high coverage exomes as input. Lastly, we show that there is an imputation metric that can precisely predict the probability that any given imputed genotype (per sample) is accurate. **Availability and Implementation:** EZimputer is under the GNU public license athttp://www. mayo. edu/research/departments-divisions/department-health-sciences-research/division-biomedical-statistics-informatics/software/bioinformatics-software-packagesandhttp://code. google. com/p/ezimputer/.

1757T

Genetic and clinical predictors for asthma risk assessment among children in Taiwan. *J. Tseng[1,2], M. Su[1], C. Tsai[1], Y. Lee[1].* 1) Institutive of Epidemiology and Preventive Medicin, College of Public Health, National Taiwan University, Taipei, Taiwan; 2) Department of Pediatrics, Children Hospital, Changhua Christian Hospital, Changhua, Taiwan.

Background Asthma is the single most common chronic childhood disease affected by multiple genetic and environmental factors. Although many genetic variants have been noted to be associated with childhood asthma, up to the present, no risk assessment models that incorporate genetic and clinical predictors for asthma occurrence are currently available. **Methods** We analyzed 10 single nucleotide polymorphisms (SNPs) identified from 9 asthma-associated genes among subjects who participated in three children's cohort studies in Taiwan (TCHS, GBCA and TCCAS). The parents of all participants were interviewed regarding the information of asthma risk factors. Modified Asthma Predictive Index (mAPI) and genetic risk scores (GRSs) of 10 SNPs were used in prediction model assessment. The performance of prediction ability was assessed by discrimination and reclassification statistics, and calibration. Cross validation with leave-one-out cross validation method was also conducted. **Results**In total, 640 asthma cases and 1921 control subjects were included in current study. After controlling for age, sex and BMI, two environmental factors, moist and water damage of wall, were significantly associated with childhood asthma and all these covariates were incorporated into baseline model. The prediction ability of childhood asthma improved greatly in combined model composed of GRSs and mAPI compared with clinical model, demonstrating an increased AUC from 0. 748 to 0. 780 ($p \leq$ 0. 0001). Besides, the improvement in discrimination (IDI: 0. 022, $p \leq$0. 0001) and reclassification (continuous NRI: 0. 344, $p \leq$ 0. 0001) also showed while comparing clinical model and combined model. Considering the 18% general risk of asthma in children, the combined model concluded sensitivity, specificity, positive predictive value, and negative predictive value as 78. 8%, 59. 4%, 39. 5%, and 89. 3%, respectively. **Conclusion**We have successfully constructed a prediction model composed of genetic and clinical factors in asthma risk assessment. Adding GRS into clinical factor-based prediction model would increase the prediction ability substantially in predicting the occurrence of childhood asthma.

1758F

A cloud based in-silico research platform for disease genomics: G-DOC *Plus*. *K. Bhuvaneshwar, Y. Gusev, A. Belouali, V. Singh, R. M. Johnson, L. Song, A. Alaoui, M. Harris, S. Madhavan.* Innovation Center for Biomedical Informatics, Georgetown University, Washington, DC.

Our flagship web platform G-DOC was first deployed in 2011 enable the practice of a translational-based approach to research and medicine in cancer. In fall 2014, we released G-DOC *Plus* to enable integration of multiple data types to understand mechanisms of cancer and non-cancer diseases for precision medicine. G-DOC*Plus*was developed using our in-house architectural framework to support over 1100 users and 10,000 patient and cell line samples. Our data collection includes whole genome sequencing (WGS) data from the 1000 Genomes Project and Complete Genomics; multi-omics data from the NCI-60 data collection; numerous breast, GI, and pediatric cancer studies; and non-cancer studies. In Spring 2015, G-DOC *Plus* became the official host of the well-known REMBRANDT data (Repository of Brain Neoplasia Data). Itincludes a broad collection of bioinformatics and systems biology tools for analysis and visualization of many 'omics' types including DNA, mRNA, miRNA, metabolites, DNA copy number analysis using our original Chromosomal Instability Index (CIN) algorithm in addition to WGS and clinical data. Users can also explore medical images, somatic mutations and perform pathway enrichment analysis using Reactome database. G-DOC *Plus* enables researchers to easily mine existing datasets to gain fundamentally new insights into complex diseases in an effort to improve survival and therapeutic outcomes. In this submission, we report in-silico analysis case studies done using clinical, gene expression and copy number data on four different types of brain tumors in the REMBRANDT dataset. We compared Astrocytoma (low-grade glioma) with Gliobastoma (GBM, high-grade glioma) patients. The most under expressed gene in this analysis, RHOF, was six fold under-expressed in the GBM group. This gene is known in literature to be down regulated in GBM patients through the over expression of their activators. We also examined copy number data using the CIN tool in G-DOC *Plus* and found higher chromosomal instability in the Astrocytoma group in chromosome 8q, which is also well known in literature. G-DOC *Plus* tools have been leveraged in cancer and non-cancer disease studies for hypothesis generation; multi-omic in-silico analysis and population genetics analysis. The long-term vision for G-DOC *Plus* is to extend this systems medicine platform to hospital networks to provide more effective clinical decision support. G-DOC *Plus* is available at: https://gdoc. georgetown. edu.

1759W

Can functional data assess genetic risk? – A polygenic risk score approach. *M. Butkiewicz[1], W. S. Bush[1], M. A. Pericak-Vance[2], W. K. Scott[2], J. L. Haines[1].* 1) Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) John P. Hussman Institute for Human Genomics University of Miami Miller School of Medicine, Miami, FL.

Polygenic genetic risk scores (PGRS) have been previously applied to characterize genetic effects for groups of SNPs in GWA studies where individual SNPs may not reach a significance (p-value) threshold. However, because most of the SNPs have no known functional impact and likely serve as surrogates for the actual functional variant, their explanatory effect may be blunted, reducing the accuracy of the score. To circumvent this limitation and as a proof of principle, we hypothesize that a polygenic functional risk score (PFRS) based on functional expression quantitative trait loci (eQTL) data can function like a PGRS because it can capture actual functional effects. Using quantified expression levels of mRNAs instead of GWAS related p-values, a SNP is chosen only if characterized as eQTL. If true, then known associated genes should generate a better predictive set of eQTL SNPs than genes not associated with that trait. In this study, we analyzed subject samples within the Amish assessed for age-related macular degeneration (AMD) comprised of 112 cases and 463 controls. Functional eQTL data was accessed through the Fantom5 database. We started with known AMD-associated genes. Genotype data was parsed to include only SNPs that had matching eQTL data. The resulting SNPs, weighted by the eQTL effect size estimate (beta value), were aggregated by gene to assemble a gene-weighted feature dataset for each subject. Next, machine learning algorithms were trained to distinguish cases from controls. The resulting 5-fold cross-validated classifier, achieved a prediction accuracy average of 81%, with a data split of 80% for training and 20% for testing. Subsequently, a different, distinct set of genes associated with Alzheimer's disease (AD) was chosen and the training protocol was repeated. Based on the same subject samples but using a different target gene set, the resulting classifier only achieved a reduced average prediction accuracy of 58%. These preliminary results demonstrates that classification performance differs depending on the gene set used, and supports our hypothesis that functional eQTL data can capture correlations between trait-associated SNPs and the trait itself, here AMD and AD. Next, we will identify larger case/control cohorts and compare our approach to existing PGRS. With systematic selection of new SNP sets, we anticipate that this ongoing study will reveal novel eQTL SNP combinations correlated with traits specific to human cell types.

1760T

Polycomb Repressive Complex 2 epigenomic signature defines age-associated hypermethylation and gene expression changes. *M. Dozmorov.* Biostatistics, Virginia Commonwealth University, Richmond, VA.

Although age-associated gene expression and methylation changes have been reported throughout the literature, the unifying epigenomic principles of aging remain poorly understood. Recent explosion in availability and resolution of functional/regulatory genome annotation data (epigenomic data), such as that provided by the ENCODE and Roadmap Epigenomics projects, provides an opportunity for the identification of epigenomic mechanisms potentially altered by age-associated differentially methylated regions (aDMRs) and to find regulatory signatures in the promoters of age-associated genes (aGENs). Here we show that aDMRs and aGENs identified in multiple independent studies share a common Polycomb Repressive Complex 2 signature marked by EZH2, SUZ12, CTCF binding sites, repressive H3K27me3 and activating H3K4me1 histone modification marks, and a "poised promoter" chromatin state. This signature is depleted in Pol II-associated transcription factor binding sites, activating H3K79me2, H3K36me3, H3K27ac marks, and an "active promoter" chromatin state. The PRC2 signature was shown to be generally stable across cell types. When considering the directionality of methylation changes, we found the PRC2 signature to be associated with aDMRs hypermethylated with age, while hypomethylated aDMRs were associated with enhancers. In contrast, aGENs were associated with the PRC2 signature independently of the directionality of gene expression changes. These results suggest that the PRC2 signature is the common epigenomic context of genomic regions associated with hypermethylation and gene expression changes in aging.

1761F

Population scale human genome analysis on the cloud. *J. Fitch[1], B. Kelly[1], P. White[1,2].* 1) The Research Institute at Nationwide Children's Hospital, Columbus, OH; 2) Department of Pediatrics, The Ohio State University, Columbus OH.

Advancing sequencing technologies have made population scale whole genome sequencing a possibility. However, current strategies for analysis of this data rely upon approaches that have limited scalability, lack reproducibility and are complex to implement, requiring substantial investment in specialized IT solutions. To overcome these challenges we developed a highly accurate and deterministic analysis solution, named Churchill, which fully automates the analytical process required to perform the complex and computationally intensive process of alignment, post-alignment processing, local realignment, recalibration and genotyping. Our parallelization strategy enables division of each analysis step across multiple compute instances, enabling whole genome variant discovery and genotyping to be completed in under 90 minutes. In addition to rapid analysis of a single sample, Churchill optimizes utilization of available compute resources and scales in a near linear fashion, enabling population scale genome analysis to be performed cost-effectively using cloud resources. To demonstrate this, we analyzed the 1000 Genomes Project dataset of 2,504 whole genome and exome sequenced individuals. Starting from FASTQ raw input data, we were able to fully automate the complete analysis process, ultimately performing multi-sample variant calling and generating population allele frequencies in seven days. We identified approximately 75 million variants in common with those released in the 1KG Phase 3 analysis of this dataset, with highly similar minor allele frequencies, demonstrating the efficacy of our rapid approach. In addition, through utilizing a highly sensitive method for variant detection that calls SNPs and indels simultaneously via local re-assembly of haplotypes in an active region, in combination with joint genotyping across the entire population, we discovered 5. 6 million indels not reported in the original analysis. Our approach demonstrates the feasibility of generating population allele frequencies specific to a given unified analysis approach, critical for accurately filtering datasets for discovery of rare pathogenic variants. Moreover, through use of on demand cloud computing resources, our method represents a solution for the genomics computational bottleneck and will keep pace with the magnitude of data generated by population scale sequencing.

1762W

Functional long-range regulatory interactions harbor co-occurring ancestry specific variants. *D. L. Gibbs[1], C. Humphries[1], N. Chambwe[1], T. Knijnenburg[1], B. Bernard[1], J. Roach[1], J. Vockley[2], J. Niederhuber[2], I. Shmulevich[1].* 1) Institute for Systems Biology, Seattle , WA; 2) Inova Translational Medicine Institute, Inova, Fairfax, Virginia.

Populations are defined by genetic isolation. This isolation results in differences in genetic variants and different patterns of gene expression. Variants falling within regulatory regions can affect the control of gene expression by altering binding motifs. These influences include long-range three-dimensional chromatin interactions crucial to the control of expression. This work examines the idea that variant pairs, co-segregating in populations, could affect the three-dimensional regulatory structure and alter expression patterns. Using whole genome sequences, we mapped common variants (>1% MAF) from 791 women and 788 men of diverse backgrounds into publicly available ChIA-PET derived chromatin interaction data. Statistical testing determines if variant pairs appear at a higher rate and associate with altered gene expression for a given population. In total, 78,982 variants fell within the genomic interactions; 2,215 interactions contained variants on both sides of the interaction. Significant variant pairs allowed for a comparison between populations, showing varying degrees of agreement. All population groups were observed to have unique variant pairs, but the European group was found to have the largest number of unique variant pairs with 74, followed by the African group with 19 unique variant pairs. Using UCSC and ENCODE annotation databases, such as RegulomeDB, variant pairs were annotated for gene structure and regulatory elements. While many variant pairs are found in insulator regions, 157 pairs are found in combinations of active promoter and strong enhancer regions. Lastly, using RNAseq data derived from women in the study, variant pairs are assessed for impact on the control of gene expression. This work helps to explain observed differences in gene regulation across populations, and contains the potential for clinical application.

1763T

Network-based meta-analyses of multiple gene expression profiles with BMD variations in females. *H. He[1], S. L. Cao[2], Y. Zhou[1], L. Zhang[1], T. H. Niu[1], Y. P. Wang[2], H. W. Deng[1].* 1) Biostatistics and Bioinformatics, School of Public Health and Tropical Medicine, New Orleans, LA; 2) Department of Biomedical Engineering, Tulane University, New Orleans, LA, 70118, USA.

Existing microarray studies of bone mineral density (BMD) are important parts of the efforts to understand the pathology of osteoporosis and have identified a number of candidate genes. However, these studies were limited to small sample size and usually analyzed individually. We proposed a network-based meta-analysis approach that combined meta-analysis across six microarray datasets and functional module identification from human protein-protein interaction (PPI) to highlight differentially expressed genes (DEGs) and a functional module that may play important roles in BMD regulation. Expression profiling studies were identified by searching PubMed database, GEO and ArrayExpress. Two meta-analysis approaches were applied across the preprocessed datasets. The first nonparametric meta-analysis combined p-values by Fisher's method from individual experiments to identify genes with a large effect size in all datasets. The second approach combined effect sizes from each dataset into a meta-effect size to estimate the amount of change in expression across all datasets. Genes with *Q* test's p-value less than 0. 05 or *I2* value higher than 50% were assessed by Random Effects Model, otherwise by Fixed Effects Model . Functional modules were identified through the integrated analysis of microarray data in the context of large biological protein–protein interaction (PPI) networks by using Fisher's combined p-values. KEGG pathway enrichment analysis on module genes was performed by a hypergeometric test. Six gene expression datasets were identified and raw CEL files were downloaded from GEO and ArrayExpress. In total there were 249 female subjects. A consensus module was identified containing 58 genes (nodes) and 83 edges. KEGG pathway enrichment analysis of 58 module genes revealed that genes were enriched in several important pathway like Osteoclast differentiation and B cell receptor signaling pathway. Meta-analyses by combining effect size and by combining p-values were both applied to the individual gene level to prioritize more important candidate genes. Five candidate genes (*ESR1*, *MAP3K3*, *PYGM*, *RAC1* and *SYK*) were identified and their associations with BMD were all validated by a BMD meta-analysis study. In summary, our network-based meta-analysis not only identified important DEGs but also discovered functional modules biologically related to osteoporosis pathology. Our study may provide importantly potential therapeutic targets for osteoporosis.

1764F

AnthOligo: Automated design of hybridization oligonucleotides for region-specific extraction of large contigious DNA fragments. *P. Jayaraman[1], K. Mackiewicz[3], P. M. Clark[3], M. Sarmady[1], D. Monos[2,3].* 1) Bioinformatics Group, Division of Genomic Diagnostics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pathology and Lab Medicine, Perelman School of Medicine, University of Pennsylvania; 3) Immunogenetics Lab, Division of Genomic Diagnostics, Dept. of Pathology and Lab Medicine, Children's Hospital of Philadelphia.

Region-specific extraction(RSE) of DNA is an increasingly utilized solution-based technique for enrichment and targeted resequencing of defined genomic regions of interest by next-generation sequencing(NGS). The captured contiguous fragments of DNA are particularly amenable to PacBio Single Molecule Real-Time Sequencing(SMRT) platforms, which can produce elongated reads, enabling long-range haplotype phasing across complex regions of the genome, such as the MHC. AnthOligo is a web application developed to automate the integrated process of designing capture oligonucleotides for such emerging sequencing platforms. This program facilitates the design of oligos for any locus of interest based upon user-defined constraints. The final set of optimal oligos resulting from such a complex process needs to satisfy numerous requirements (melting temperature, self/cross dimerization, entropy and putative off-target hybridization loci etc.). While countless tools exist to perform a subset of the desired tasks, they do not automatically retrieve genomic sequence and design the oligos. AnthOligo integrates several tools including Primer3, UCSC Genome Browser utilities and UNAFold to design oligos and analyze their properties. For given regions of interest, it extracts the reference sequence from UCSC Genome Browser (hg19) and breaks it down to 6Kb windows. To ensure optimal design of capture oligos, the 6Kb window is shifted towards the 3' end of the target locus at 3Kb intervals, producing overlapping windows. For each unique window, the Primer3 application is run based on user-defined configurations to generate raw oligos. UCSC BLAT, hairpin, and self/cross-dimer analyses are then run to further refine the oligos based on uniqueness, temperature, and entropy parameters. To select final list of oligos, a sliding window is used to select oligos that are ~8Kb apart from each other to ensure significant coverage of the target region. Previously designed oligos were used to successfully validate AnthOligo results. By streamlining the process of selecting an optimal set of oligos, this application is estimated to reduce considerable hands-on time. Program parameters can be customized for each individual run, making AnthOligo flexible for a wider range of applications (e. g. the design of internal oligos to be used as baits for gene panel analysis or as probes for large-scale CgH array processes). AnthOligo is publicly available as a web application at http://antholigo. chop. edu.

1765W

miR-96 targets a network of transcription factors and cell cycle regulators associated with prostate cancer progression. *M. D. Long[1], P. K. Singh[2], M. J. Campbell[1].* 1) 1Department of Pharmacology and Therapeutics, Roswell Park Cancer Institute, Elm & Carlton Streets, Buffalo, NY 14263, USA; 2) New York State Center of Excellence in Bioinformatics and Life Sciences, State University of New York at Buffalo, 701 Ellicott St, Buffalo, NY 14203 USA.

MicroRNAs often target gene expression to multiple members of the same network. The nuclear receptor (NR) superfamily of transcription factors forms one such network that is relevant to prostate cancer (PCa) progression and therapy. Previously, we used bootstrap permutation approaches on TCGA PCa cohorts (prad_mskcc and prad_tcga) to reveal that NR superfamily members were more commonly down-regulated at the mRNA level than predicted by chance. The retinoic acid receptor gamma (RARγ) was the most commonly lost NR (reflecting prostatic metaplasia in Rarγ knockout mice, reported by others). Rather than mutations or deletions correlating with NR expression loss, a cohort of 65 miRNA targeting NRs were elevated in expression, more than predicted by chance. The most significant negative correlation was between miR-96 and RARg mRNA. Both miR-96 transfection and *RARG*-3'UTR luciferase assays validated that miR-96 regulated RARγexpression. Expression analyses revealed miR-96 gain and RARγ loss during murine PCa development and elevated miR-96 levels significantly associated with worse disease free survival in the prad_tcga cohort. To identify miR-96 targets we used a biotinylated miR-96 (bi-miR-96) pull down approach. Non-malignant RWPE-1 prostate cells and LNCaP prostate cancer cells were transfected with bi-miR-96 or *c. elegans* bi-cel-miR-67 prior to microarray analyses. These analyses revealed both unique and shared targets between the two cells. Focusing on LNCaP cells, we identified nearly 400 genes enriched significantly with bi-miR-96. Hypergeometric tests revealed these genes were significantly enriched in miR-96 targets predicted by miRWalk (p<1. 3e-16) and GSEA (P<2e-27). Kolmogorov–Smirnov tests established these genes were significantly negatively correlated with miR-96 in prad_tcga cohort (p< 5. 2e-7). PCR validation confirmed miR-96 interacted with established (*FOXO1*, *RARG*) and novel targets (*SLC6A9*, *GCNT1*). Annotation of these genes revealed that the most enriched class of genes was transcription factors (~10%) including *CEBPA*, *ARID3B* and *SMAD4.* GO term analyses revealed a very significant enrichment for cell cycle regulation and reflecting this, miR-96 mimics increased the growth of LNCaP cells by reducing cells in G1 phase. Together these findings suggest that miR-96 is significantly associated with aggressive prostate cancer in part by regulating RARγ and a network of transcription factors associated with cell cycle control.

1766T

Examining lost reads to survey the microbiome and immune components of the human body across 43 human sites from 175 individuals. *S. Mangul[1], N. Strauli[2], H. Yang[1], R. Hernandez[2], R. Ophoff[3], E. Eskin[1,3], N. Zaitlen[4].* 1) Computer Science Department, University of California, Los Angeles; 2) Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco; 3) Human Genetics, Los Angeles, University of California, Los Angeles; 4) Department of Medicine, University of California, San Francisco.

In this work we aim to profile the lost reads to study immune system function across tissues directly from RNA-seq data. We use 1641 RNA-Seq samples corresponding to 175 individuals and 43 sites from GTEx project : 29 solid organ tissues, 11 brain subregions, whole blood, and two cell lines, LCL and cultured fibroblasts from skin. . We use the unmapped reads to obtain a detailed profile of the microbial and immune components of the human body. We filtered out 54. 68%±7. 28% of the unmapped reads, which were low-quality and/or low complexity as well as 33. 18%±4. 82% of the unmapped reads that mapped to the human reference sequence (MegaBLAST, edit distance 10). The remaining high-quality unique reads are used to perform a survey of the microbiome and immune components. To profile the microbiome we used phylogenetic marker genes to assign candidate microbial reads to the bacterial and archeal taxa. A total of 713 taxa were assigned with Phylosift, with 8 taxa on the phylum level. No microbial organisms were observed in heart, pituitary and adrenal gland. All other tissues contain at least one bacterial or archeal phyla (0. 79+-0. 55 phyla per sample). We observe two viruses harbored in multiple tissues. EBV virus is present in 20% of the skin samples and 66% of the liver samples but it is not present in any of the brain samples. Enterobacteria phage phiX174 virus is present in 20% of the skin samples but is not present in liver and brain tissues. To profile B and T cells across tissues candidate immune reads were mapped to the V(D)J regions of the Ig loci. We compared normalized read counts of V(D)J mapped reads between tissues in order to determine the related abundances of immune cells. As a positive control we confirm that abundances of V(D)J reads in LCLs are significantly greater than thyroid and brain. Consistent with its role as an interface to environmental exposure, lung also contained and increased abundance of V(D)J reads relative to thyroid and brain (p-value < 10-6). We next measured the genetic diversity of V(D)J variants by examine number of gene alleles expressed per sample We observed a lower diversity of gene alleles in the lung and thyroid compared to LCLs (p-value < 10-16). Examining immune and microbial genes in GTEx can help define typical profiles for a healthy tissue. It is essential to monitor microbial and immune diversity, and this work may eventually help diagnose immune and microbiome imbalance in a tissue specific manner.

1767F

The feasibility test of fetal fraction estimation in targeted sequencing using fragment size in the case of Duchenne Muscular Dystrophy (DMD). *J. Park[1], D. Y. Park[1], B. Lee[1], T. Yun[1], J. Lee[1], J. H. Chae[2,3], B. C. Lim[2,3].* 1) Molecular diagnostics R&D Team, IVD Business Division, SK telecom, Seongnam, Kyunggi, South Korea; 2) Departent of Pediatrics, Seoul National University College of Medicine, Seoul National University Children's Hospital, Seoul, Korea; 3) Institute of Reproductive Medicine and Population, Medical Research Center, Seoul Natinal University, Seoul, Korea.

Background:In noninvasive prenatal testing (NIPT), targeted deep sequencing is used to detect micro insertion and deletion for screening Mendelian disorders. In this study, fetal fraction estimation method based on DNA fragment size was applied to targeted sequencing data. The feasibility of DNA fragment size based fetal fraction estimation was already shown in whole genome shallow sequencing (Stephanie Yu et. al). From the comparison of other conventional methods, DNA fragment size based method shows comparable results. Thus, DNA fragment size based method is more cost effective and can be applied for quality control of data. This method works well without any extra gene information and SNP information. Method:Customtargeted panel(Agilent SureSelect) was designed to cover the entire DMD gene region. Targeted sequencing was performed using 9maternal plasma DNA at varying gestational weeks from 5 mothers. Fetal genomic DNA obtained from either chorionic villi sampling or amniocentesis and genomic DNA from 5 mothers was used to validate the results. Using mean read depth of two Zinc Finger genes with a minimum mapping quality score of 20 and base quality score of 20, the fractional fetal DNA concentration was calculated as previous study (S. Yoo et. al). Also a size ratio indicating the relative proportions of short and long DNA fragments was calculated for each sample for the comparison ofZinc Fingermethod (Stephanie Yu et. al.). For more information, we compare the AB allele frequency ratio of hetero SNP sites of mother to confirm the fetal DNA was shorter than maternal DNA. Linear regression was used to compare two methods. Results:- From the comparison of AB allele frequency ratio based fetal fraction estimation and size ratio based one, the proposed method shows positive correlation (r=0. 85, p=3. 896x10-8). Similar result is also obtained from the comparison of Zinc Finger based estimation (r=0. 67, p=0. 09). In addition, our linear model is similar to the previously published whole genome sequencing data (Stephanie C. Y. Yu et. al). Thus fragment size based fetal fraction estimation can be extendible to targeted sequencing. ReferenceStephanie C. Y. Yu et. al. Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. PNAS 111:23 p8583–8588 (2014)Seong-Keun Yoo et. al. Noninvasive Prenatal Diagnosis of Duchenne Muscular Dystrophy: Comprehensive Genetic Diagnosis in Carrier, Proband, and Fetus. Clinical Chemistry 61:6 (2015).

1768W

Predicted effect of single nucleotide polymorphisms affecting skin/hair coloration and/or 4-6 KHz hearing loss in young adults on secondary protein structure of melanocortin-1 receptor (MC1R). *C. B. Pudrith[1], S. L. Phillips[1], V. C. Henrich[2].* 1) Communication Sciences and Disorders, University of North Carolina at Greensboro, Greensboro, NC; 2) Center for Biotechnology, Genomics, and Health Research, University of North Carolina at Greensboro, Greensboro, NC.

In a recent study (Phillips et al. 2015), we found that young adult musicians carrying 1 of several minor alleles in the melanocortin receptor 1 (MC1R) gene had bilateral notching (BN), a 15 db or greater loss of hearing acuity at 4-6 KHz in both ears. A case-control study (N=252) included 84 subjects with BN, along with 2 control groups (unilateral notching and no notching). Eight MC1R SNPs were identified for all 252 subjects, and 25 (9. 9%) individuals had at least 1 minor allele. Of these 25 carriers, 18 (72%) had BN. Bilateral notching is a possible phenotype for identifying genetic variants associated with susceptibility to noise induced hearing loss. Because there are numerous SNPs with a relatively low frequency within the MC1R gene, a statistical analysis for any single polymorphism could not be performed. The purpose of this study is to determine if non-synonymous SNPs (NS-SNPs) in MC1R including those that are prevalent in BN subjects, also are predicted to affect the secondary structure of MC1R, using a protein structure-modeling program, Raptor X. The analysis included 5 SNPs associated with BN in the Phillips study (rs1110400, rs117952179, rs2229617, rs34090186, and rs34158934), one that was not (rs885479) and 8 other previously studied MC1R SNPs. The predicted structure of the wild type sequence was compared to the predicted structure of each polymorphic sequence. The variance between each sequence pair was calculated at each amino acid site, and then totaled throughout the protein. The overall variances for the NS-SNPs had a bimodal distribution. Of the 14 NS-SNP tested, 4 were not predicted to affect MC1R secondary structure (Group 1) and 10 did (Group 2). Two of the 5 NS-SNPs associated with BN (rs2229617 and rs34090186) led to a predicted change in secondary structure and all 8 SNPs from previous studies predicted a change. An unexpected finding was that the specific amino acids affected by Group 2 substitutions were nearly identical. This indicates that MC1R has several polymorphic sites that lead to similar secondary structural changes, regardless of their location in the MC1R sequence. It also suggests the possibility that some SNPs of MC1R can affect hearing acuity even though a predicted change in MC1R secondary structure is not discernible by computer modeling. This work was supported in part by the National Institutes of Deafness and other Communication Disorders grant R21DC009296-01.

1769T

Carrier Risk Estimation using Single Nucleotide Polymorphism-based Measures of Relatedness. *R. Shraga, M. Gold, A. Manoharan, S. Yarnall.* Recombine, Inc. , New York , NY.

Background: Genome wide measures of human relatedness have been used in a wide variety of applications. The goal of this study was to evaluate localized genetic relatedness in specific clinically relevant regions. We focused on DNA regions surrounding genes related to Cystic Fibrosis, Sickle-Cell Anemia, and Smith-Lemli-Opitz Syndrome. **Methods:** We created three different cohorts of patients by selecting carriers of the 3 disorders and adding randomized control patients who were known non-carriers of the same ethnicity. In order to study the relatedness of each group of patients, we selected single nucleotide polymorphisms (SNPs) in the region of the genes implicated in each disease. We used PLINK to prune the initial set of SNPs and kept those that are in approximate linkage equilibrium. We created genetic similarity matrices for each cohort using six different SNP-based measures of relatedness. These measures differ in how they score heterozygous matches and account for minor allele frequency (MAF). Additionally, we used different windows for SNP selection ranging from 1 to 10 Mbp upstream and downstream of the gene. We compared average relatedness within the case groups, within the control groups, and between the case and control groups. This research led to the development of an initial carrier risk classifier for Cystic Fibrosis mutation DeltaF508, which modifies the k-Nearest Neighbor (KNN) algorithm to use our SNP-based relatedness measure instead of Euclidean distance. We tested the performance of our classifier using repeated random sub-sampling cross validation. **Results:** We found that carriers of all 3 disorders appear more similar to each other than they do to controls. This observation was more pronounced in the measures that give a stronger weight to SNPs with low MAF. We also found that inter-group relatedness between sets of patients that are carriers of different mutations for the same disease varies widely. Using the modified KNN classifier, we randomly sub-sampled our cohort into training and testing sets 20 times, and found that 95. 28% of cases get classified as at risk compared to 25. 54% of controls. **Conclusion:** These results indicate it may be feasible to classify carrier risk even when direct observation of a mutation is not possible, especially if our observations hold true for rare variants. This may be useful for improving sensitivity in cases where direct testing is difficult or error prone.

1770F

Childhood Asthma Clusters Reveal Different Clinical Characteristics and Gene Signatures. *M. W. Su[1, 2], C. H. Tsai[1], Y. F. Wu[1], B. L. Chiang[3], Y. H. Yang[3], Y. T. Lin[3], L. C. Wang[3], J. H. Lee[3], C. C. Chou[3], W. C. Wu[1], Y. L. Lee[1, 2].* 1) Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan; 2) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan; 3) Department of Pediatrics, National Taiwan University Hospital, Taipei, Taiwan.

Childhood asthma is a heterogeneous disorder that comprises distinct disease subphenotypes with complex pathophysiology. Asthma subphenotypes would evoke different clinical characteristics and various responses to medications. We want to identify different clusters of childhood asthma and achieve better therapeutic strategies. We applied k-means cluster analysis of twelve objective clinical examinations to 351 children with asthma in the Taiwanese Consortium of Childhood Asthma Study (TCCAS). Using microarray of peripheral blood mononuclear cell (PBMC) from children with different asthma subphenotypes, we develop a gene expression signature that predicts neutrophil-predominant asthma by logistic regression model. Five clusters of childhood asthma were identified and can be characterized by either predominant eosinophilic or neutrophilic, or even mixed eosinophilic and neutrophilic inflammatory pattern. A six-gene signature discriminates children with neutrophil-predominant asthma from other subphenotypes, including aquaporin 9 (AQP9), lipocalin 2 (LCN2), carcinoembryonic antigen-related cell adhesion molecule 6 and 8 (CEACAM6 and CEACAM8), orosomucoid 1 (ORM1), and peptidoglycan recognition protein 1 (PGLYRP1). The performance was evaluated by leave-one-out cross-validation, and the area under ROC curve (AUC) was 0. 93. Clustering can be helpful to identify clinically relevant patients and develop personalize asthma therapeutic strategies. Gene expression signatures of asthma subphenotypes not only improve our knowledge about childhood asthma, but provide guidance on future therapeutic strategies.

1771W

Reducing Off-Target Rate in Targeted Sequencing. *L. Teng, C. L. Hsieh, C. Lin, H. Y. Chuang.* Illumina, 5200 Illumina Way, San Diego, CA.

Targeted sequencing involves isolation of genomic regions of interest in a sample library, enabling cost-effective and systematic detection of germline and somatic variants. There are a wide variety of targeted approaches, ranging from biochemistry assays to NGS platforms and analysis tools. One of the major obstacles shared across the various approaches is the significant amount of sequenced reads eaten by off-target genomic regions. Targeted sequencing generally achieves only 40%~60% efficiency [1]. The off-target reads not only wastes sequencing yield but also potentially compromises variant calling for somatic mutations of low frequency. In this study, by utilizing peak calling techniques commonly used in ChIP-seq [2], we found that under the same experimental condition, a large majority of the off-target reads were systematic, highly reproducible across different samples and replicates. Many regions that were highly enriched with the systematic off-target reads were inside gene bodies. The similarity between the off-target regions and capture probes suggested that a great portion of off-target reads were likely pulled down by non-specific probes. To improve the on-target rate, we designed specific blockers or probes that bind to the identified off-target genomic regions. Different enrichment protocols which make use of the off-target blockers or probes were explored and the result showed promising reduction on the off-target rate.

1772T

Pathway Based Targeted Sequencing and Meta-Genomic Analysis of Preterm Birth. *A. Uzun[1,2], J. Schuster[1,2], B. Mcgonnigal[1,2], C. Schorl[3], A. T. Dewan[4], J. F. Padbury[1,2].* 1) Pediatrics, Women and Infants Hospital, Providence, RI; 2) Pediatrics, Brown Medical School, Brown University,Providence, RI; 3) Molecular, Cellular Biology Biochemistry, Brown Medical School, Brown University,Providence, RI; 4) Department of Chronic Disease Epidemiology, Yale School of Public Health, Yale University, New Haven, CT.

Preterm birth is a major public health problem. It occurs in 1 in 8 women. Classical estimates from twin studies suggest that the heritability of preterm birth is up to 40%. Pedigree analysis puts all of the risk on the maternal side. Previously, we analyzed a genome-wide case/control study, with approximately total of 2,000 term preterm mothers from the Danish National Birth Cohort to test the validity of the genes for Preterm birth. We identified 329 genes nested within 69 gene sets and 132 haplotype blocks that are highly associated with preterm birth. **1)** We performed targeted sequencing on these genes and genomic regions on 23 women with 2 generations of preterm birth, 9 women with 3 generations of preterm birth and 16 term controls. We found over 23,200 variants. **2)** We applied zygosity testing where we compared the variants of cases and controls for their abundance of homozygosity and/or heterozygosity. Statistical comparisons yield around 400 significant ($p<0.05$) variants on 200 genes. **3)** Meta-genomic analysis: We built a gene list for each mother (total of 48) where variants with $p<0.1$ were included and applied Gene Set Enrichment Analysis (GSEA) for each. Next, we transformed the gene set results for each mother into binary gene set profiles and applied an iterative binary bi-clustering algorithm (iBBiG) to identify groups of gene sets that are coordinately associated with groups of mothers and their phenotypes (preterm, term) across the GSEA results. We identified 5 significant ($p<0.05$) and overlapping clusters (modules) of shared genes and variants associated with preterm birth. These 5 modules included 34 genes with 217 variants were contributed. This meta-genomic approach is readily adaptable to incorporation of sequencing results from other projects on preterm birth.

1773F

Determine the contribution of genomic structure variation to the etiology of Tourette syndrome using whole-exome sequencing. *Y. Zhang[1], T. Fernandez[2], R. King[2], M. State[3], J. Tischfield[1], G. Heiman[1], J. Xing[1], The TIC Genetics Collaborative Group.* 1) Rutgers, the State University of New Jersey, Department of Genetics and the Human Genetics Institute of New Jersey, Piscataway, NJ, USA; 2) Yale Child Study Center and Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA; 3) Department of Psychiatry, University of California, San Francisco, USA.

Copy number variation (CNV) is a major component of human genetic variation. CNVs have been associated with many neuropsychiatric disorders, such as Schizophrenia, Autism, Autism spectrum disorder, and obsessive–compulsive disorder (OCD). Tourette syndrome (TS) is a neurodevelopmental disorder characterized by repetitive, stereotyped, involuntary movements and vocalizations called tics. Despite the strong evidence that genetics play a significant role in Tourette syndrome etiology, little is known about the genetic basis of TS. For this reason, whole-exome sequencing (WES) data has been generated by the Tourette International Collaborative Genetics (TIC Genetics) study to identify disease causal variants. To date, the TIC Genetics study has completed WES on 325 simplex families. We hypothesize that rare genic CNVs contribute to TS. Compared to microarray-based approach, WES offers several advantages for CNV detection, including higher sensitivity for smaller CNVs and higher enrichment for genic CNVs. To identify CNVs, we evaluated existing WES-based CNV computational approaches, including FishingCNV, CoNIFER and EXCAVATOR. We are also developing a novel computational approach to identify *de novo* genic CNVs in TS patients. Overall, our analysis will shed light on the genetic causation of TS related to genomic structural variation.

1774W

Accelerating pancreatic cancer drug screening by leveraging genomics to select better *in vitro* models. *R. Kusko[1], Y. Cha[1,2], J. Perez-Rogers[1,3], A. Lysaght[1], B. Weiner[1], S. Kolitz[1], F. Towfic[1], J. Funt[1], K. Fowler[1], B. Vardarajan[1,4], M. Artyomov[1,5], B. Zeskind[1].* 1) Immuneering, Cambridge, MA; 2) MIT, Cambridge, MA; 3) Boston University, Boston, MA; 4) Columbia University Medical Center, New York, NY; 5) Washington University in in St Louis, St Louis, MO.

The projection that pancreatic cancer will be the second leading cause of cancer related death by 2020 compounded by the numerous clinical trial failures precipitates the need for increased progress in new medicine development. Cell lines used for screening pre-clinical compounds prior to animal or human testing are selected for ease of access and literature prevalence. However, the constellation of genomic derangements in cell lines commonly used for *in vitro* studies may not be representative of pancreatic cancer tumors. In this study, we leveraged a wealth of publically available data from The Cancer Genome Atlas (TCGA) and the Cancer Cell Line Encyclopedia (CCLE) to predict optimal cell lines. We estimated the popularity of each CCLE profiled pancreatic cancer cell line using literature search and compared this to genomic similarity metrics. To start, we compared median per gene CNV values in TCGA pancreatic cancer tumors to pancreatic cancer cell lines in CCLE. Contrary to our expectation, the top five cell lines by CNV correlation with TCGA pancreatic tumors represented only 6% out of all literature search hits for all CCLE pancreatic cancer cell lines, indicating the availability of more optimal cell lines from a genomics perspective. Additionally, we used publically available algorithms to filter targeted sequencing data from TCGA and CCLE pancreatic cancer cell lines. Using clustering based on the presence or absence of these 7 mutations which passed filter, we showed that some pancreatic cancer cell lines readily clustered amongst TCGA pancreatic tumors (such as L33), while others occupied discrete branches of the dendrogram exclusive of most TCGA tumors (such as PK1 and PANC1). Lastly, we conducted an integrative analysis comparing the overall mutational burden in pathways between pancreatic cancer cell lines and tumors. Cluster analysis revealed that the majority of cell lines are highly similar to each other, but a handful of cell lines such as L33 cluster in branches mostly occupied by tumors. These results together imply that while some cell lines mimic pancreatic tumor genomes closely, others represent genomic constellations not commonly observed in patients. In summary, our work reports that many popular pancreatic cancer cell lines harbor distinct genomic aberration patterns from pancreatic cancer tumors in patients and highlights the emerging role of genomics in advancing the clinical success of therapeutic trials.

1775T

Using genome-wide genotyping and genome-wide gene expression profiles to perform integrative analyses of neutrophilic asthma. *C. H. Tsai[1], M. W. Su[1], B. L. Chiang[2], Y. H. Yang[2], C. C. Chou[2], Y. J. Chang[3], Y. F. Wu[1], Y. L. Lee[1,3].* 1) Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, Taipei, Taiwan; 2) Department of Pediatrics, National Taiwan University Hospital, Taipei, Taiwan; 3) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.

Background:Asthma is a heterogeneous airway disease. Neutrophil predominance in asthmatics has been reported to be associated with poor lung function and disease outcomes, as well as sudden asthma fatalities. Genome-wide association studies (GWAS) and genome-wide gene expression have been used to detect susceptibility gene for asthma. However, the results of GWAS generally are weak effects and wide variations in gene-expression microarrays. Therefore, the objective of this study is to integrate genome-wide genotyping and genome-wide gene expression profiles to investigate the genetic markers and pathway in neutrophilic asthma. Methods:Absolute neutrophil counts in peripheral blood from 535 asthmatics were used to find the cut-off point to discriminate neutrophilic asthma from Taiwanese Consortium of Childhood Asthma Study (TCCAS). We examined the whole genome genotyping and expression profiles from 30 cases of neutrophilic asthma and 29 non-neutrophilic asthma of peripheral blood mononuclear cells. Differentially expression gene were be defined, then Gene Set Enrichment Analysis (GSEA) and Ingenuity Pathway Analysis (IPA) were be applied to find potentially biological pathway. We preformed Gene Set Association Analysis (GSAA) and expression quantitative trait loci (eQTL) to integrate multiple genomic analysis platforms. Results:Higher neutrophil counts were found to be associated with poor pulmonary function, higher lung illness and without controlled asthma. We identified 592 differentially expression genes (FDR<0. 05) which were associated with defense and immune responses, including biology of immune system process. Production of nitric oxide and relative oxygen species in macrophages was the most significantly canonical pathway by IPA. Using integrative analysis, we found 14 leading edge genes by GSAA and 20 SNP-gene pairs by eQTL. A total of 3 genes, *BCL6 (B-Cell Lymphoma 6)*, *PPARG (Peroxisome proliferator-activated receptor gamma)* and *CDA (Cytidine Deaminase)*, were also reported on GSEA or IPA analysis. Conclusion:Our results indicate that neutrophil predominance is an important subphenotype in asthma. We could identify SNPs that affect gene expression level altering biological function in neutrophilic asthma. It is a practicable approach to combine SNP and gene expression information to investigate the important genetic markers and biological pathway on complex diseases.

1776F

Detecting the influence of genetic variants on personality. *K. Wolffhechel[1], H. Jarmer[1], S. M. van den Berg[2], M. H. M. de Moor[3], J. J. Hottenga[3], D. I. Boomsma[3].* 1) Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Kongens Lyngby, Denmark; 2) Department of Research Methodology, Measurement and Data Analysis, University of Twente, Enschede, Netherlands; 3) Department of Biological Psychology, VU University Amsterdam, Amsterdam, Netherlands.

Personality has for many decades been known as partly hereditary, but the genetic basis for the development of our personality is still mostly unknown. In this work we attempt to classify personality traits from genetic variants using state-of-the-art machine learning algorithms on 60-item NEO-FFI measurements and over 8 million SNPs from 2946 unrelated Dutch participants. First, we choose pathway genes known to influence brain function and in a repeated nested 5-fold cross-validation we then select by genome-wide association for pathways associated with personality types and use these as predictors to classify to the two sides of a given personality trait. We find a classification better than random to the two ends of the scale for some of the Big Five traits. Next steps will be to get a better understanding of the found pathways and to look into possible gene-environment effects not captured by our model. Getting a clearer understanding of how exactly genetics influence our personality development would give us a huge advantage in predicting and perhaps preventing the onset of genetically caused personality disorders.

1777W

Diagnostic Role of Exome Sequencing in Immune Deficiency Disorders. *S. E. Brenner[1], A. N. Adhikari[1], J. P. Patel[2], A. Y. Chan[3], D. Punwani[3], H. Wang[3], A. Kwan[3], T. A. Kadlecek[3], M. J. Cowan[3], M. Mollenauer[3], J. Kuriyan[1], S. M. Fu[4], U. Sunderam[5], S. Rana[5], A. Chellappan[5], K. Kundu[5], A. Mulder[6], F. H. J. Claas[6], J. A. Church[7], A. Weiss[3], R. A. Gatti[8], J. M. Puck[3], R. Srinivasan[5].* 1) University of California, Berkeley, CA, USA; 2) Children's Hospital of Los Angeles, Los Angeles, CA, USA; 3) University of California, San Francisco, CA, USA; 4) University of Virginia School of Medicine, Charlottesville, VA, USA; 5) Innovation Labs, Tata Consultancy Services Hyderabad, India; 6) Leiden University Medical Centre, Leiden, The Netherlands; 7) University of Southern California, Los Angeles, CA, USA; 8) University of California, Los Angeles, CA, USA.

We developed an analysis protocol for individual genome interpretation and used its distinctive features to diagnose numerous clinical cases. We applied the protocol to exomes from patients with undiagnosed primary immune disorders, with a particular focus on infants with absent or low T cell receptor excision circles (TRECs), a marker for T-cells, at birth. To yield high quality sets of possible causative variants, we used multiple callers with multisample calling and integrated variant annotation, variant filtering, and gene prioritization. Our first cases involved two unrelated infant girls with low TRECs at birth that were not eligible for a bone marrow transplantation because no abnormalities were found in genes related to severe combined immunodeficiency (SCID). We discovered compound heterozygous variants in the *ATM*gene for both the infants. After clinical confirmation, this discovery offered a very early diagnosis of Ataxia Telangiectasia (AT), allowing for avoidance of undue irradiation and live vaccinations as well as revealing a new potential role for TREC screening in AT diagnosis. In another case, the affected siblings had early onset bullous phempigoid, a chronic autoimmune disorder. Our analysis revealed compound heterozygous mutations in ZAP70, a gene associated with profound primary immunodeficiency, the opposite phenotype. Cellular immunological studies indicated that one variant was hypomorphic and the other was hyperactive. These combined to yield a novel presentation, adding to the existing phenotype repertoire of ZAP70 in humans. Our analysis protocol focuses on genomic features that may be overlooked by other methods. In the case of a female with severe influenza pneumonia, our annotation tool, Varant, flagged variants in *PRKDC* apparently occurring after the genomically-encoded stop codon. This stop codon in the reference genome was correctly identified as premature due to a rare single base deletion. But accounting for the proband being normal at this position, we correctly annotated the proband's two variants as nonsynonymous mutations likely causative for the phenotype. Our protocol has been similarly revealing in other SCID and CID cases including Nijmegen Breakage Syndrome and BCL11B, which highlight unique features of the analysis framework that facilitate genetic discovery. These help provide crucial information to offer prompt appropriate treatment, family genetic counseling, and avoidance of diagnostic odyssey.

1778T

OMIM at 50: the nexus for knowledge on human genes and genetic disorders. *A. Hamosh[1], C. Bocchini[1], J. Amberger[1], N. Sobreira[1], F. Schiettecatte[2], A. F. Scott[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins Univ Sch Med, Baltimore, MD; 2) FS Consulting, LLC. Salem, MA.

The 2015 ASHG meeting launches a one-year celebration of Online *Mendelian Inheritance in Man*, OMIM's 50th anniversary. As of June 11, 2015, OMIM has curated over 7,800 phenotype and 15,000 gene entries. Among these are 3,405 genes recognized as causing 5,500 phenotypes. Between May 1, 2014 and April 30, 2015, there were a total of 2. 4 million visitors to OMIM. org, over 10,000 registered users of the ftp download, and close to 3,000 users of the API. Over the past 50 years, OMIM(http://OMIM. org) has grown from a series of books comprised primarily of descriptions of Mendelian phenotypes to a comprehensive interactive resource of knowledge on human genes and phenotypes that serves clinical and research genetics. OMIM classifies and names genetic disorders described in the published literature with consideration given to both clinical and molecular aspects of disease. For those not familiar with a particular gene or phenotype, the OMIM entry text provides a descriptivesummary and a jumping point to additional references and connections to context-specific information in other resources(for a comprehensive list, see external resources at OMIM. org). Discrete clinical features of a disorder or phenotype are represented in Clinical Synopses and are mapped to HPO and SNOMED CT terms thus facilitating incorporation into computational research. The OMIM Phenotypic Series provide a unique view of the genetic heterogeneity of many Mendelian disorders and facilitate bioinformatic research across networks of related disorders. OMIM provides data to the NCBI Clinical Variation Database, ClinVar, NCBI Gene databases, and controlled vocabularies such as HPO and the UMLS. OMIM is working with the ClinGen Project to establish community standards for phenotyping, disease gene curation, and variant annotation. In addition, OMIM encourages collaboration among clinicians and researchers through MIMmatch, a tool that allows users to follow entry updates and find colleagues with similar research interests. By providing unique expertly curated information on the relationship between genes and disease that is critical to overlapping and diverse communities of clinicians, molecular biologists and genome scientists, as well as students and teachers of these disciplines, OMIM stands at the nexus of clinical genetics, genomics and translational science.

1779F

Investigation of rare variants in complex disorders by using pooled DNA sequencing. *J. Wang[1], T. Skoog[1], H. Laivuori[2,3], P. Gerdhem[4], H. Lohi[5,6], J. Kere[1,6], H. Jiao[1].* 1) Department of Biosciences and Nutrition, Center for Innovative Medicine, Sciences for Life Laboratory, Karolinska Institutet, Stockholm, Sweden; 2) Medical Genetics and Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, Flnland; 3) Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 4) Department of Orthopedics, Karolinska University Hospital and Department of Clinical Sciences, Intervention and Technology (CLINTEC), Karolinska Institutet, Stockholm, Sweden; 5) Department of Veterinary Biosciences, University of Helsinki, Helsinki, Finland; 6) Molecular Neurology Research Program, University of Helsinki, and Folkhälsan Institute of Genetics, Helsinki, Finland.

BackgroundNext generation sequencing technologies have revolutionized genetics studies by providing an opportunity to detect rare variants associated with complex diseases. However, sequencing thousands of individuals in population-based association studies is costly and laborious. Sequencing pooled DNA samples could reduce the effort, but the question is if rare alleles are detected and if minor allele frequencies (MAF) estimated using pooled samples are correct. We aimed at evaluating the accuracy of MAF calculated using pooled DNA sequencing data. **Method**Whole genome sequencing (WGS) was performed with 20 untagged bullterrier DNA samples distributed into two pools, ten in each. A similar pooling strategy was applied on 100 scoliosis and 100 preeclampsia patient samples, ten in each pool using exome-sequencing. We tested different variant calling tools, e. g. SAMtools, GATK and FreeBayes, and compared the discovery of alleles and MAFs called by different software. Furthermore, the MAFs estimated by the three tools were evaluated by the MAFs obtained from individual-based genotyping data using microarray or Sequenom platform. **Result**We did not find large difference among the variants identified by three different tools, suggesting high discovery rate. For 20 bullterrier samples, 67867 of 69061 autosomal SNPs detected from Illumina Canine chip were detected by whole genome sequencing. The pooled MAF of around 50% shared SNPs was identical between WGS and microarray, and 35% had only one allele difference by direct allele counting. The rate of identical allele estimates was larger than 73% for rare variants (MAF≤0. 05). For exome-sequencing, 48 low frequency variants in preeclampsia study and 43 rare variants in scoliosis study were validated by genotyping individually. Except four monomorphic cases, MAFs estimated from pooled DNA sequencing data showed high concordance (R2= 0. 86) with those from genotyping data. **Conclusion**Pooled DNA sequencing data captured with high accuracy the alleles and MAFs as compared to individual genotyping data. The pooling strategy could be a rapid and cost-effective alternative approach for evaluating rare variants in association studies.

1780W

Distribution of unique sequences in the human genome. *K. Misawa[1,2], K. Kojima[2,3], Y. Kawai[2,3], M. Nagasaki[1,2,3].* 1) Department of Cohort Genome Information Analysis, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 2) Graduate School of Medicine, Tohoku University, Sendai, Miyagi, Japan; 3) Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan.

Targeted nucleases are useful tools for genome alteration with high precision. The RNA-guided Cas9 nuclease from the microbial clustered regularly interspaced short palindromic repeats (CRISPR) adaptive immune system can be used as an efficient genome engineering tool in eukaryotic cells including human cells (*1*). CRISPR requires a 20-bp targeting sequence (*2*). When studying large genomes, however, the design of oligonucleotides for CRISPR is complicated by the redundancy of sequences. Such non-unique sequences cannot be CRISPR targets, because they cause off-target reaction. Thus, the distribution of unique sequences in the genome is of interest. We describe the development of a novel method, UF, for detecting unique 20-bp sequences in entire genomes (*3*). UF stands for "Unique Finder. " By using UF, the distribution of unique sequences in the human genome was investigated. It was found that 60% of the human genome is unique on average as shown in the previous study (*4*). It was found that non-unique regions of human genome are concentrated on centromeres and terminal regions of the chromosomes. The proportions of unique sequences are about 80% in the rest part of the genome. ToMMo processed the whole genome sequences of 1,070 healthy Japanese individuals. We are now investigating genetic variations of non-unique sequences in the Japanese population. The program for obtaining unique sequences is available at https:// sourceforge. jp/projects/parallelgwas/releases/. Reference1. P. D. Hsu *et al.* , DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* 31, 827-832 (2013). 2. F. A. Ran *et al.* , Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8, 2281-2308 (2013). 3. K. Misawa, Distribution of unique sequences in the human genome. *Austin J Comput Biol Bioinform.* 2, 1010 (2015). 4. W. Li, J. Freudenberg, P. Miramontes, Diminishing return for increased Mappability with longer sequencing reads: implications of the k-mer distributions in the human genome. *BMC Bioinformatics* 15, 2 (2014).

1781T

Modular transcriptional repertoire, network community structure, and microRNA target analyses in thymic tissue of Down syndrome infants. *C. A. Moreira-Filho[1], S. Y. Bando[1], F. B. Bertonha[1], F. N. Silva[2], L. F. Costa[2], L. R. Ferreira[1], M. Carneiro-Sampaio[1].* 1) Pediatrics FMUSP, University of São Paulo, São Paulo, São Paulo, Brazil; 2) Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, Brazil.

Trisomy 21- driven transcriptional alterations in human thymus were characterized through gene coexpression networks (GCNs) - obtained for differentially expressed (DE networks) and for all valid GO annotated transcripts (CO networks) - coarse-grained community structure and microRNA-target analyses. We used whole thymic tissue (corticomedullar sections) - obtained at heart surgery from Down syndrome (DS) and karyotipically normal individuals (CT) - and a network-based approach for GCN analysis that allows the identification of modular transcriptional repertoires and the interactions between all the system's constituents through community detection. Changes in the degree of connections observed for hierarchically important hubs/genes in CT and DS gene networks corresponded to sub-network changes, i. e. module (communities) changes. Distinct communities of highly interconnected gene sets were topologically identified for DS and CT networks. Coarse-grained community structure revealed distinct functional interactions between gene communities for CT and DS networks. The role of microRNAs in modulating the expression of highly connected genes in CT and DS was revealed through microRNA-target analysis. Our results show that thymus global hypofunction in Down syndrome correlates with distinctive GCN topology and node hierarchy. The comparative genetic and topological analysis of DE networks disclose how significant is the impact of trisomy 21 on thymus functioning. The high modularity of DS-DE network contrasts with its reduced connectivity, thus indicating a certain degree of disorganization of modular interactions caused by gene dosage imbalance. Coarse-grained community analysis confirmed these data. Since the CO networks are more similar in their connectivity and modularity (although not in hubs' hierarchy), the DS-DE network may represent the "ground zero", where trisomy-driven gene dysregulation strongly impacted the normal gene-gene interactions, with the derived "shock-waves" reflecting on the DS-CO network. In conclusion, alterations observed in DS networks regarding connectivity, modularity, and communities' structure reflect chromosome 21 dysregulation and its consequences. As a whole, the results indicate that GCNs' functional and topological modules correspond, and that trisomy 21 may be interpreted as the breakdown and altered reorganization of functional modules. This mechanism is coherent with the network-based model of human disease.

1782F

Longitudinal prediction of the infant gut microbiome with dynamic bayesian networks. *M. J. McGeachie[1], J. E. Sordillo[1], T. Gibson[1], G. M. Weinstock[2], Y. -Y. Liu[1], D. R. Gold[1], S. T. Weiss[1], A. Litonjua[1].* 1) Channing Div Network Med, Dept of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; 2) The Jackson Laboratory for Genomic Medicine, Farmington, CT.

BACKGROUNDSequencing of the 16S rDNA gene allows comprehensive assessment of microbial community composition from human body sites. Emerging evidence suggests that differences or aberrations in the microbiome can lead to various diseases and chronic conditions. Dynamic Bayesian Networks (DBNs) have been used in other settings to successfully model time series data and obtain accurate predictions of future behavior as well as identify salient connections and relationships within the data. METHODSWe modeled extant infant gut microbiome time series data using DBNs to demonstrate ability to predict progression of microbiome taxa and to identify important relationships among those taxa. Previously published and publicly accessible data on 58 preterm infants in the Neonatal Intensive Care Unit who underwent frequent stool collection for a month was used. We constructed DBNs from the data and analyzed predictive performance and network characteristics. RESULTSA DBN model of the infant gut microbial ecosystem explicitly captured relationships observed previously. Relationships identified between age and *Clostridia*, and between *Clostridia*, *Gammaproteobacteria*, and *Bacilli* were among the strongest relationships identified. In general, with increasing infant age, the DBN predicted increasing amounts of *Clostridia*, residual amounts of *Bacilli*, and increasing amounts of *Gammaproteobacteria* that then give way to *Clostridia*. Decreasing amounts of *Bacilli* over time were dependent upon increasing *Clostridia* and *Actinobacteria*, when present. Prediction performance of DBNs with fewer edges were overall more accurate, although less so on harder-to-predict subjects (p = 0. 016). The DBN provided quantitative likelihood estimates for rare, dramatic, sudden shifts in microbiome population ("abruptions") observed in some infants. Iterative prediction, while less accurate that sequential prediction (p < 0. 001), showed remarkable insensitivity to initial conditions and indicated convergence to a mix of *Clostridia*, *Gammaproteobacteria*, and *Bacilli* for all tested initial microbiome populations. CONCLUSIONSDBNs were able to identify important relationships between microbiome taxa and predict future changes in microbiome composition from measured or synthetic initial conditions. DBNs also provided likelihood estimates for sudden, dramatic shifts in microbiome composition, which may be useful in guiding further analysis of those samples.

1783W

Structural and functional investigation of pathogenic mutations of IL10RA gene by computational methods. *B. Babajan[1], M. Kaleemuddin[3], E. Ramu[1], Y. A. Jumana[1,2], S. Noor Ahmadh[1,2].* 1) Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD) , King Abdulaziz University Jeddah 21589 Kingdom of Saudi Arabia; 2) Department of Genetic Medicine, Faculty of Medicine King Abdulaziz University Jeddah 21589 Kingdom of Saudi Arabia; 3) Department of Biochemistry, King Abdulaziz University Jeddah 21589 Kingdom of Saudi Arabia.

ObjectiveThe advent of genome wide association studies, whole - genome and - exome studies has started to uncover known/novel pathogenic variations in interleukin-10 receptor alpha (IL10 RA) gene in several autoimmune diseases. However, the structural and functional impact of these genetic mutations is not well identified. Hence, this study has performed the computational analysis of clinically potential missense and untranslated region mutations of human IL10RA gene. **Methodology**In this study a combination of empirical rule and support vector machine based insilico algorithms were employed to predict the pathogenic potential of non-synonymous mutations of 1L10 gene. Additionally, molecular modeling and secondary structure analysis was performed to confirm their impact on the stability and secondary properties of IL10RA protein. **Results**Besides, the mutations corresponding to p. Y57C, p. T84I, p. Y91C, p. R101W, p. R117C, p. R117H and p. G141R in exonic region; c. *1537T>C, a regulatory region variant was also found to potentially influence the structural and functional deviations of IL10RA activity. Moreover, the molecular docking analysis of IL10RA with substrate (IL10) and mutated forms of IL10RA was found to be three mutated proteins of IL10RA are (p. Y57C, p. R117C and p. G141R) most interesting residues for mutagenesis and affect the selectivity and affinity of IL10RA towards IL10. **Conclusion**Our findings are expected to help in narrowing down the number of IL10RA genetic variants to be screened for auto immune disease association studies and also to design a better competitive inhibitors for mutated forms IL10RA protein. **Keywords**Insilico analysis, IL10 Gene, Pathogenic Mutations.

1784T

Forecasting a renal prognosis of IgA nephropathy using machine learning. *H. Lee[1], J. Noh[2], H. Kim[3], W. S. Yang[3], G. Kim[2], Y. S. Kim[1], D. K. Kim[1].* 1) Internal Medicine, Seoul National University Hospital; 2) Computer Science and Engineering, Seoul National University College of Engineering; 3) Internal Medicine, Asan Medical Center, University of Ulsan College of Medicine.

Previous studies of IgA nephropathy (IgAN) patients have focused on risk factor evaluation. We aimed to develop individual outcome prediction models in IgAN patients using machine learning methods. We reviewed clinical and pathologic characteristics of adult IgAN patients from Seoul National University Hospital (SNUH, n = 1,540) and Asan Medical Center (AMC, n = 1,044) at the time of renal biopsy. They were divided into development (followed up ≥10 years) and prediction (follow up <10 years) sets, respectively. The outcome was 10-year renal survival (10YRS) probability. We developed prediction models from SNUH test set by using logistic regression (LR) with Lasso method, a classification and regression tree (CART), and neural network (NN) using 16 clinico-pathologic variables. We also used bagging, random forest (RF) and boosting for ensemble learning. Finally, those models were validated internally in SNUH prediction set and externally in AMC development and prediction sets. Considering missing data, 1,514 and 847 patients were included from SNUH and AMC cohorts. In the LR model, estimated glomerular filtration rate (eGFR), hemoglobin, proportions of GS and SS, interstitial fibrosis were selected as predictors for ESRD. In the CART model, eGFR 53. 3 ml/min/1. 73m2 was proved to be a watershed for 10YRS, followed by, proportion of GS (≥ 39. 2%) and SS (≥ 28. 5%), tubular atrophy (≥ moderate), hemoglobin (<11. 9 g/dL) and proteinuria (≥ 1. 39 g/g), sequentially (accuracy, 0. 849). In the NN, variables were selected including eGFR, proportions of SS, hemoglobin, proportion of GS, albumin, and interstitial fibrosis, sequentially (accuracy, 0. 862). In addition, the ensemble learners showed good performance (accuracies of bagging, 0. 868; RF, 0. 874; boosting, 0. 862). Those individual learners were validated internally with good performance (sensitivities of LR, 0. 855; CART, 0. 921; NN, 0. 952; bagging, 0. 857; RF, 0. 921; boosting, 0. 921). And finally, we proved the robustness of those models from external validation. Good performances of both development (sensitivities of LR, 0. 847; CART, 0. 867; NN, 0. 855; bagging, 0. 852; RF, 0. 872; boosting, 0. 851) and prediction sets (sensitivities of LR, 0. 980; CART, 0. 902; NN, 0. 941; bagging, 0. 882; RF, 0. 902; boosting, 0. 922) were showed. We developed robustness of prediction models using machine learning for the individual's likelihood of 10YRS in IgAN with both internal and external validation.

1785F

Identification of genetic interactions involved in Dyslexia pathogenesis. *N. Karbalai[1,2], D. Czamara[1,2], B. Pütz[1], C. Wolf[1], T. Kam-Thong[1], A. Gialluisi[11], R. Malik[2,17], F. Ramus[15], T. S. Scerri[7,8], J. Arloth[1], P. Hoffmann[4,5], K. Moll[6], K. U. Ludwig[4,5], C. Francks[11,12], J. Schumacher[4,5], G. Huguet[13,15], A. P. Morris[7,18], F. Fauchereau[13,14], T. Bourgeron[13,14,16], A. P. Monaco[7,10], S. Paracchini[7,9], S. E. Fisher[11,12], M. Nöthen[4,5], G. Schulte-Körne[6], B. Møller-Myhsok[1,2,3].* 1) Max Planck Institute of Psychiatry, Munich, Germany; 2) Munich Cluster for Systems Neurology (SyNergy), Munich, Germany; 3) Institute of Translational Medicine University of Liverpool, Liverpool, UK; 4) Institute of Human Genetics, University of Bonn, Bonn, Germany; 5) Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany; 6) Department of Child and Adolescent Psychiatry, Psychosomatic, and Psychotherapy, Ludwig-Maximilians University of Munich, Munich, Germany; 7) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 8) The Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia; 9) School of Medicine, University of St Andrews, St Andrews, UK; 10) Tufts University, Medford, MA, USA; 11) Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands; 12) Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, Netherlands; 13) Human Genetics and Cognitive Functions Unit, Institut Pasteur, Paris, France; 14) University Paris Diderot, Sorbonne Paris Cité, Paris, France; 15) Laboratoire de Sciences Cognitives et Psycholinguistique, Ecole Normale Supérieure, CNRS, EHESS, Paris, France; 16) FondaMental Foundation, Créteil, France; 17) Institute for Stroke and Dementia Research, Ludwig-Maximilians University, Munich, Germany; 18) Department of Biostatistics, University of Liverpool, Liverpool, UK.

Dyslexia is a highly prevalent disorder in children, characterized by severe deficits in learning to read and/or spell. Several cognitive endophenotypes like phonological awareness and auditory short-term memory were correlated with core dyslexia phenotypes. However, the underlying neurobiological processes of these different cognitive abilities suggest a highly heterogeneous genetic architecture contributing to susceptibility for dyslexia, which can obscure single-locus associations. We therefore went beyond established single-locus methods to analyse higher-order genetic interactions, which are essential components of genetic processes. Within a sample of 862 dyslexics from France, Germany, USA, and UK we conducted a genome-wide two-locus interaction scan via the tool GLIDE. We identified genetic interactions affecting susceptibility for altered cognitive skills (single-word reading, phonological awareness, nonword reading, and spelling), previously ascertained for each individual. Our survey revealed interactions at highly relevant genomic sites, comprising both novel and previously known candidates for dyslexia susceptibility. Strong associations ($p \leq 9*10\text{-}14$) for single-word reading were obtained for interactions between variants on chr9q31. 3 and intronic variants of *FOXP2* (chr7q31. 1). This transcription factor gene has been implicated in a severe form of speech and language disorder, and recent studies have suggested that it may also contribute to etiology of dyslexia. Among others, we also observed significant interactions ($p \leq 9*10\text{-}13$) for phonological awareness between variants on chr18q11. 2, a genomic region linked to various dyslexia endophenotypes, and intronic variants of *NCAM1* (chr11q23. 2). This neural cell adhesion gene is known to be involved in nervous system development and has been associated with risk of ADHD, and Bipolar Disorder. Gene-set enrichment analysis of all loci with an endophenotype association of $p \leq 1*10\text{-}10$ highlighted processes, that were previously implicated in dyslexia, as significantly enriched e. g. axon guidance, locomotory behaviour, development of the cerebral cortex and nervous system. Furthermore, we found an accumulation of transcription factor binding sites and/or DNaseI hypersensitivity sites annotated by ENCODE at the chromosomal regions of our significant loci. Our results suggest multi-genetic mechanisms like allelic interactions play an important role in etiology of dyslexia.

1786W

Systematic and integrative analysis of a gene-set involved in prostate cancer: a bioinformatics study. *J. Pani[1], K. Gautam Singh[2], A. Narayan Singh[3], F. Kumar Pandey[4], H. Narayan Singh[5].* 1) cytogenetics, MTA Infotech, New Delhi, New Delhi, India; 2) Bioinformatics Division, MTA Infotech, New Delhi; 3) IT Division, MTA Infotech, New Delhi; 4) I. T. S Paramedical College, Ghaziabad, India; 5) School of Sciences, Noida International University, Gautam Budh Nagar, India.

The etiology of prostate cancer (PC) is multifactorial, including genetic and non-genetics factors. However, the etiology of the disease is yet to understand completely. Development of any cancer reflects a progressive accumulation of alterations in various genes and their effect on various pathways. Therefore, identification of genes and study of the molecular basis of their corresponding pathways will definitely provide insights into the etiology and prognosis of the disease. List of genes involved in the PC were extracted from gene-disease association databases. Genes were classified and further, pathway enrichment analysis was also done at FDR<0. 01. 27 genes were found to be reported for the disease pathogenesis. These gene products belong to 12 different categories based on their function viz. Signaling molecule (13 genes), Enzymes (11), Receptor (8 genes), Transcription factor (5 genes), Nucleic acid binding (5 genes), Enzyme modulator (4 genes), Defense/immunity protein(3 genes), Extracellular matrix protein (2 genes), Calcium-binding protein (2 genes), Transporter (2 genes), Cell adhesion molecule (1 genes), Surfactant (1 genes), however, some genes belong to more than one categories. Pathway enrichment analyses (FDR < 0. 01) of these genes clearly show their involvement with signaling pathways. These genes belong to various protein function categories; however, signaling molecules are prominent in disease progression. HMGB1 involve in transactivation of androgen receptor; VEGFA implicated in the angiogenesis that occur in PC via VEGF signaling pathway; IL-6, 7, 15 differentially regulates androgen receptor transactivation via various signaling pathways; ERG, a protooncogene involve in translocation leading to cancer; KLK3, also known as prostate-specific antigen and androgen control its expression; KLK4 is androgen-regulated molecule which is reported to be overexpressed in PC; CD82 inhibit metastatic process; dysregulation of TGFB1 activation and signaling may result in apoptosis, IGF1 is a marker of tumor progression, VIP activates adenylyl cyclase in prostate gland; NRG1 can accumulate in specific subnuclear compartments. The present study gives valuable insights into pathogenesis of PC and suggests important interventions to ameliorate this disorder. It reveals that PC is a multifactorial complex disease which involves many genes and multiple pathways; however, androgen-receptor signaling pathway is most prominent mediator of the PC progression.

1787T

Implications for cell type-specific transcriptomics in the developing human brain from next-generation sequencing of subcellular RNA fractions. *A. J. Price[1,2,3], D. Kim[1,2], R. Tao[1], D. R. Weinberger[1,2], A. E. Jaffe[1,2].* 1) Lieber Institute for Brain Development, Baltimore, MD; 2) Johns Hopkins University School of Medicine, Baltimore, MD; 3) McKusick-Nathans Institute of Genetic Medicine, Baltimore, MD.

Post-mortem human brain tissue is an essential substrate for understating the molecular pathology of brain disorders and their risk. Many risk genes for schizophrenia and related disorders are likely dysregulated within individual cell populations, resulting in reduced signal when studying homogenate brain tissue. The most popular solution to the issue of cellular heterogeneity when using human frozen postmortem brain is currently to isolate cell nuclei from homogenate tissue, label the nuclei with cell-type specific fluorescent antibodies and sort out different cell types using fluorescence-activated cell sorting. This solution is well suited for DNA-based epigenetic studies but is limiting in terms of RNA, since a large part of the transcriptome is discarded with the cytosol. To determine the impact of profiling nuclear RNA in lieu of the whole transcriptome, we isolated nuclear and cytosolic RNA from three fetal and three adult human brains, prepared libraries from each RNA fraction for next-generation sequencing using both rRNA-depletion methods (Ribozero) and PolyA selection methods (PolyA+), and performed RNA sequencing on these 24 samples. In line with previous reports, we found that Ribozero library preparation highly enriched the proportion of reads aligning to introns particularly in the nuclear RNA fraction (6. 18% for PolyA+ and 47. 03% for Ribozero). Likewise, reads spanning splice junctions were highly enriched in polyA+ library samples regardless of cell fraction (p<10-6), suggesting enrichment for pre-mRNA in Ribozero nuclear samples. We found that library type had a larger impact on measurement of gene expression than RNA fraction overall, and that differences in gene expression between the nucleus and cytosol were mitigated by using polyA+ library preparation. On a cautionary note, the top differentially expressed genes across brain development only partially overlapped between library type in both the nucleus and cytosol. Investigators should therefore consider these results before deciding on library preparation method when undertaking transcriptome studies using human brain nuclei.

1788F

A Computational Protein Phenotype Prediction Approach to analyze the deleterious mutations of human med12 gene. *R. Elango[1,5], B. Banaganapalli[1,5], M. Kaleemuddin[1,3,5], J. Y. Al-Aama[1,2], I. Khan[4], N. Shaik[2].* 1) Princess Al-Jawhara Al-Brahim Center of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdullaziz University, Jeddah, Saudi Arabia. , Jeddah, Saudi Arabia; 2) Department of Genetic Medicine, Faculty of Medicine, King Andullaziz University, Jeddah, Saudi Arabia; 3) Department of Biochemistry, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia; 4) Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, King Saud University, Riyadh, Saudi Arabia; 5) Equal contribution.

Objective: MED12is a subunit of the multiprotein complex Mediator, an evolutionary-conserved regulator of transcription. Somatic mutations in MED12gene have been described in benign and smooth muscle cell tumors. The underlying basis if any mutation elicit the in protein phenotype changes remains unclear. The computational investigation on mutation effects on protein structure found to be an effective alternate method to traditional *in-vivo* and *in-vitro* approaches. Hence, we have performed the *in-silico* structure analysis of clinically potential missense mutations of human MED12 gene. Methodology: The MED12 gene mutations were collected from different databases and by text-mining. Combination of empirical rule and support vector machine based *in-silico* algorithms were employed to screen the pathogenic non-synonymous genetic mutations based on their nucleotide sequence conservation across species. Both native and conserved region specific 3-D MED12 mutant protein structures were built by integrative *ab-initio* and homology modeling methods. Solvent accessibility, secondary structure analysis, change in stability and other structural deformities of the protein due to mutation were analyzed. Results: The concordant predictions of SIFT, PolyPhen2 and I-Mutant 3. 0 have attributed the pathogenicity to 47 (28. 31%) missense mutations that span the gene. Our analysis showed that protein structures with amino acid substitutions between 1-100 amino acids- aa (Q43P, G44S, G44D, G44F) and 1800-1969aa (N1845T and S1969P) are structurally less divergent (0. 51Å -0. 59Å) compared to the substitutions between 350-550aa (P353H, R356W, E446A, D499H, R521H, R540H, K550N), with high structural divergence (0. 60Å -0. 89Å). Interestingly, these results confirm the protein stability and solvent accessibility findings. Conclusion: Missense mutations showing high structural divergence are likely to be more deleterious to protein and contribute to disease. Our findings are expected to help in narrowing down the number of MED12 variants to be screened for association studies. The *in-silico* based mutation analysis, being helpful in providing information about the nature of mutations, may also function as a first-pass filter to determine the candidates worth pursuing for further research in other coagulation disorder causing genes.

1789W

Diagnostic sequencing in mosaicism and heteroplasmy: allele detection and base calling by graph analysis. *S. Bang[1], J. Chae[1,2,3], K. Hwang[4].* 1) Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea; 2) Department of Pediatrics, Division of Neurology, Seoul National University Children's Hospital, Seoul, Korea; 3) Seoul National University College of Medicine, Seoul, Korea; 4) School of Computing, Soongsil University, Seoul, Korea.

Identifying mosaicism in nuclear DNA and heteroplasmy in mitochondrial DNA are challenges in diagnostic sequencing, as both yield sequences with superimposed peaks of variable height and shape. Many base calling algorithms are not adequately equipped to accurately call such positions, as the tallest peak is usually considered the 'dominant' peak, and smaller peaks are handled as background noise. Visual evaluation of each sequence is generally accurate at these locations, but low- to mid-level superimposition can yield minor peaks that are difficult to distinguish from noise based on height alone. An in-house base calling algorithm has been developed to identify alleles with superimposed peaks more accurately. The algorithm serially interpolates polynomial functions of the trace graphs and represents peaks with best-fit Gaussian distributions, allowing fine evaluation of the morphologic characteristics defining the size and shape of each peak. The rates of detection of superimposition between this algorithm and the base callers PHRED and ABI have been compared for sequences gathered from the mitochondrial DNA of 42 pediatric patients with suspected mitochondrial disease and 57 healthy controls. For 92 positions showing superimposition, our algorithm, ABI, and PHRED correctly identified 60/92 (65%), 11/92 (12%), and 0/92 (0%) positions, respectively. Positive linear correlations between the algorithm's peak profile score ratios and the quality scores given by ABI and PHRED were observed (ABI, r = 0. 667; PHRED, r = 0. 690). Graph analysis provides more accurate base calling at superimposed positions and also shows potential for distinguishing minor peaks from baseline noise, thus improving base calling accuracy not only for heteroplasmy but also for low-level mosaicism.

1790T

Morphling: a likelihood based Mobile Element Insertion (MEI) caller for whole genome sequencing data. *S. Chen[1], G. Abecasis[2].* 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI 48109; 2) Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109.

Mobile Element is a type of DNA sequences than can change their location or copy numbers in the genome. Previous studies have reported that an individual genome contains around 1,500 novel Mobile Element Insertions (MEIs). These MEIs, like other types of structure variation, might be present in regulatory or coding sequences, and might be associated with certain genetic disorders. Thus, identifying MEIs from sequencing data is of great importance to statistical genetics research. Here, we developed an MEI detector, Morphling, which automatically adjusts its parameters from remapping a small portion of reads back to a modified genomic sequence. Sample coverage and mapping status is also taken into account in this adjustment. With the adjusted parameters, Morphling converts split-read and read-pair information to genotype likelihood estimate of each potential MEI event. The genotype likelihood estimate will be of great use in following analysis. In simulated data, Morphling shows better performance compared to other callers, in terms of detection power, false discovery rate and break point accuracy. In real data, Morphling has more stable performance across samples with different coverage and mapping status. Also, Morphling is well-designed perform joint calling on multiple samples, and utilize population-scale information to improve its performance. It is quite useful for population-based studies with lower sample coverage but larger sample size. Currently we're applying Morphling to multiple sequencing studies to obtain a high quality MEI call set. We're also performing PCR validation followed by Sanger sequencing on some samples to evaluate Morphling's performance in real data.

1791F

Benchmarking well know bioinformatics aligners and variant callers using the Pilot Genome (NA12878) and Ashkenazim Father-Mother-Son trio. *A. B. Diallo, A. Carroll, B. Hannigan, S. Ma, M. Kinsella.* DNAnexus, Mountain View, CA.

Next Generation Sequencing cost is decreasing continuously and clinical applications are taking advantage of the technology. Nevertheless the standards to assess confident variant calls are limited, and improving this is important to assessing clinical utility and reproducibility. The "Genome in a Bottle Consortium" is working on the development of widely accepted reference materials and performance metrics to provide a foundation to address theseconcerns. Building on that foundation,here we present benchmark results using well know bioinformatics aligners and variant callers on the Pilot Genome (NA12878) and Ashkenazim Father-Mother-Son trio. We focus on the variant calls results qualityby using different variety of coverage depths of Illumina reads and parameters tuning to understand the costs and benefits of sequencing to higher coverage. We detail tools for comparisons of the VCF results and investigate the benefits of assembling across different methods. The pipelines, datasets, and tools are openly available on DNAnexus.

.

1792W

A realistic simulation for benchmarking germline and somatic mutation detection with long read sequencing. *B. Lau, J. C. Mu, M. Mohiyuddin, L. T. Fang, N. B. Asadi, H. Y. K. Lam.* Bina Technologies, Roche Sequencing, Redwood City, CA.

In response to recent advances in third generation sequencing technologies, an increasingly mature arsenal of bioinformatics software has been developed to analyze long reads with high error rates. For example, Pacific Biosciences (PacBio) long-read sequencing and software suite have demonstrated power to resolve long haplotypes in difficult regions of the genome. Such demonstrations have mostly been sample-specific; however, a controlled simulation-validation study of various toolsets can provide valuable insights into the general applicability of existing methods. To date, simulation has been hampered by a lack of a realistic yet versatile long-read simulator. For example, PBSIM does not generate data in the H5 format required by PacBio processing tools. We also found that the Alchemy simulator generates H5 format for a particular, older chemistry of PacBio and produces spurious artifacts. In this study, we have developed a generic and realistic long-read simulator and an automated framework for systematic evaluation of sequencing and variant calling accuracy, in the context of long-read sequencing for germline and somatic mutations. Our simulator works by first extracting a sequencing error profile from real data followed by read generation using the specified genome reference. The output is stored in both FASTQ and H5 formats, allowing downstream processing by pipelines which work exclusively with either format. The simulator is then used in conjunction with the VarSim framework which synthesizes synthetic genomes based on real variations and evaluates germline and somatic variant calling for diploid organisms. This results in an unbiased evaluation platform for a wide variety of tools when applied to long reads. We report an evaluation based on reads simulated with the error profile of PacBio public CHM1 human data. Our simulator will aid researchers in benchmarking tools to extract the most out of their long-read sequencing data.

1793T

Fast and accurate variant calling of thousands of human genomes for clinical applications. *G. A. Lunter, A. Rimmer, E. L. S. Sanders, H. T. Tonkrova, S. Trygubenko, S. Weller, G. McVean, M. A. Simpson.* Genomics plc, 52 Cornmarket Street, Oxford, United Kingdom.

Whole-genome and -exome sequencing is fast becoming a standard tool in clinical practice, and many clinical sequencing projects are underway aimed at populating large databases with patient data and their associated genomes. To help unlock the potential of these large data collections, Genomics plc is buildingthe infrastructure to analyze genomic data in a range of healthcare settings. Recognising the need for a highly efficient algorithm to enable processing of hundreds of thousands of genomes using reasonable resources while making no sacrifices to accuracy, we present a haplotype-based variant caller written in C++11 that combines high accuracy with very low CPU, memory and I/O usage. Calling a 30X human genome on a single CPU core takes 4 hours, uses 60 Mb of RAM, and requires just a single pass over the input BAM file, considerably better than other existing tools. We tested our algorithm on several community-standard data sets, including the Genome In A Bottle data, Illumina's Platinum genomes and the naturally haploid hydatiform mole data. The tool's sensitivity and specificity for SNPs andindels exceed those of other community-standard callers. We achieve this performance by usingbespoke error models for homopolymers and other repetitive regions, and for error modes typical of Illumina X10 data,the preferred platform for large-scale whole-genome sequencing. In addition the tool has the functionality to integrate a knowledge base of known variation and systematic data problems to further improve sensitivity and specificity. The tool includes integrated modulesof keyinterestforclinical genome sequencing projects, including modules forcalling of germline de-novo mutations andsomatic variants in cancer samples, the ability to genotype a pre-defined set of variants, and theability to call reference bases toconvey the confidence ofthe absence of mutations known to contribute to disease risk. Variants are reported in standard VCF / gVCF format.

1794F

Whole read overlap assembly accurately detects structural variants now in GRCh38. *A. Mangubat[1], B. Drees[1], J. Bruestle[1], N. Spies[2, 3], E. E. Eichler[4], M. Malig[4], S. N. Shekar[1].* 1) Spiral Genetics, Seattle, WA; 2) Stanford University School of Medicine, Stanford, CA; 3) National Institute of Standards and Technology Genome in a Bottle, Gaithersburg, MD; 4) Genome Sciences, University of Washington School of Medicine, Seattle, WA.

Until now, variant calls for structural variants (SVs) in whole human genome data have been prone to imprecision and high rates of false discovery. Here we show how read overlap assembly to detect variantshas the ability to detect indels and SVs (including larger insertions) with considerable precision and few false discoveries in NA12878 and an Ashkenazi Jewish trio. .

In simulated data and in a previous study using real data by English et al (2015, accepted), the method has been shown to have high sensitivity compared to other bioinformatic SV callers and a false discovery rate of less than 5% when detecting SVs.

Using the Anchored Assembly method on Illumina HiSeq NA12878 data, we detect variants previously reported in the results of the 1000 genomes project, as well as variants that have now been incorporated into the new human reference (GRCh38). In addition, we detect a number of novel SVs, which were confirmed using PacBio sequencing of fosmids (Eichler et al, 2015, in preparation). This confirms that the method can be used to detect variants not otherwise normally detected by other bioinformatic methods on Illumina data.

Using Anchored Assembly on the Ashkenazi Jewish trio sequenced by Personal Genome Project (PGP), we used the tool SVViz to validate the SVs by segregation within the family. Of 10 SVs in the offspring selected at random, all showed consistency of calls and logical segregation within the trio. Anchored Assembly detected a 3. 4kb insertion inherited in the offspring that is a match to an alternate allele assembly now in GRCh38. Given the resolution of this insertion, it was possible to identify the 5 SNPs and an indel within the insertion that were inherited from the father as well as a single SNP inherited from the mother. This level of precision makes Anchored Assembly useful for analyses of a family.

Whole read overlap assembly allows for accurate and precise calls of SVs with low false discovery rates. The ability to compare across samples is particularly useful for identifying variants associated with autism and rare genetic disorders.

1795W

A comprehensive genomics resource for assessingvariant calling accuracy. *M. Mohiyuddin[1], J. C. Mu[1], P. T. Afshar[2], X. Chen[3], N. Baniasadi[1], M. B. Gerstein[4], W. H. Wong[3, 5], H. Y. K. Lam[1].* 1) Bioinformatics, Bina Technologies, Redwood City, CA; 2) Electrical Engineering, Stanford University, Stanford, CA; 3) Statistics, Stanford University, Stanford, CA; 4) Computational Biology and Bioinformatics, Yale University, New Haven, CT; 5) Health Research and Policy, Stanford University, Stanford, CA.

Precision medicine based on high-throughput sequencing requires accurate sequencing algorithms. This necessitates the development of high-confidence and comprehensive human variant sets to accurately assess their performance. Recent works have attempted to provide such resources containing variants of high-confidence but they fail to encompass all major types of variants including structural variants (SVs). For example, the current gold sets of the NA12878 genome contain only SNVs and Indels up to couple hundred base-pairs while the HS1011 male genome gold set only incorporates SVs. To address these limitations, we leveraged the publicly available massive high-quality Sanger sequences of relatively high (9x) coverage for the HuRef genome to construct by far the most comprehensive gold set. We cross validated the gold set with deep Illumina sequencing of 140x coverage, population datasets, and well-established algorithms. Since the published variants for HuRef were mostly reported five years ago with the primary aim of obtaining its genetic makeup, we found multiple issues involving compatibility, organization, and accuracy which prevented their direct use in benchmarking. This necessitated our thorough reanalysis and validation of the HuRef genome which resulted in a gold set with high specificity and sensitivity. Our gold set contains around 3. 5M SNVs, 560K Indels, and 5K SVs (including long insertions) up to a hundred thousand base-pairs, which makes it comprehensive not only in terms of the number of variants but also for the kinds of variants over a large size range. As a demonstration of the HuRef gold set's utility, we use it to benchmark the accuracy of several published SV detection tools. Our gold set as well as related data have been released to public repositories. We further enhanced the comprehensiveness of the call set by identifying thousands of tandem repeats using the Sanger reads. We successfully identified the previously found uVNTR (upstream variable number of tandem repeats) in HuRef's MAOA gene which is associated with antisocial behavior, alcoholism, and impulsivity. As a negative control, we confirmed in our HuRef genome there is no expansion of the polymorphic trinucleotide (CAG)n repeat in the HTT gene which is involved in Huntington disease.

1796T

Utilizing the Genome Analysis Toolkit's (GATK) CalculateGenotypePosteriors to refine sequencing genotype calls based on external population and trio information. *E. Pugh, K. Hetrick, P. Zhang, K. Doheny.* Center for Inherited Disease Research, Institute of Genetic Medicine, The Johns Hopkins School of Medicine, Baltimore, MD.

GATK 3. 3-0 can use population allele frequency data and/or trio information to refine the posterior probability of the genotypes generated by HaplotypeCaller. We hoped to use this to reduce the number of false positive de novo variants seen without losing 'real' de novo variants utilizing a set of 767 whole exome (WES) samples (Agilent® SureSelect™ XT Human All Exon v4, Illumina® HiSeq™ 2000/2500) sequenced as part of our work for the Baylor Hopkins Center for Mendelian Genomics (http://www. mendelian. org/) containing a variety of ethnicities and family structures, as well as whole genome (WGS) data for a Hapmap trio. BAM files were generated using GATK 3 and bwa mem 0. 7. 8. gVCF files were created with GATK 3-3. 0 HaplotypeCaller on baited regions (79 Mb), joint called with GenotypeGVCFs, variant sites filtered with Variant Quality Score Recalibration (VQSR) using Best Practices and then genotypes refined with CalculateGenotypePosteriors using different combinations of supporting external population information (1KG and Exome Aggregation Consortium) and/or pedigree information. Concordance and sensitivity to genotypes generated on a SNP array was calculated for each sample. Mendelian errors were calculated using PLINK. For the WGS data, we compared Mendelian error rate before and after genotype refinement observing a decrease from 0. 6% to 0. 08% and 6% to 0. 15%, for SNVs and indels, respectively. For the trio subset of WES families, the combination of trio and frequency refinement reduced the SNV Mendelian error rate from 0. 20% to 0. 03%, while with trio refinement alone it was 0. 002%. Unexpectedly, for nuclear families with two or three children sequenced frequency and trio refinement reduced the error rate but trio refinement alone did not. Trio refinement alone also resulted in lowering a subset of sequencing quality metrics for a subset of samples. Mean sensitivity to heterozygotes and homozygote concordance were similar across permutations at ~98% and ~99. 5% respectively but mean homozygote concordance dropped significantly for 9 samples when only trio refinement was used. A similar pattern was seen for percent of SNPs in dbSNP 138 with a mean of ~98% with 85 samples falling below 96. 5% for trio refinement alone. Our preliminary investigation suggests that the combination of frequency and trio based genotype refinement may help reduce false de novo calls and we are continuing to explore the application of genotype refinement to WGS and WES data.

1797F

Comparison of alignment strategies for allelic expression imbalance. *C. K. Raulerson[1], M. L. Buchkovich[1], A. Ko[2,3], Z. Xu[1,4], Y. Li[1,4], M. Laakso[5], P. Pajukanta[2,3], T. S. Furey[1,6], K. L. Mohlke[1].* 1) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Department of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, USA; 3) Molecular Biology Institute at UCLA, Los Angeles, USA; 4) Department of Biostatistics, University of North Carolina, Chapel Hill, NC; 5) Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland;; 6) Department of Biology, University of North Carolina, Chapel Hill, NC.

Allelic expression imbalance (AEI) analysis identifies genetically-affected cis-acting regulatory elements at loci implicated in complex diseases. Accurate detection of AEI depends on correctly aligning reads to heterozygous sites. Most aligners only consider mapping to the reference allele at a given locus, leading to an artificially inflated mapping rate for the reference allele. Allele-aware aligners consider both alleles at each heterozygous site, providing equal opportunity for correctly mapping reads containing the alternate allele. The extent to which reference-mapping bias affects RNA-seq alignment is not fully understood. To assess the aligner's contribution to detection of AEI, we evaluated a genotype-independent aligner, STAR (allowing two mismatches), and an allele-aware aligner, GSNAP, by comparing reads mapped both to a reference genome and a complement genome. We used genotypes and adipose tissue RNA-seq data (50 bp paired-end, avg. ~45 million reads) from 8 individuals from the METSIM study. We first aligned reads to the hg19 human reference genome using each aligner. We then created a complement reference genome by converting alleles at heterozygous sites to the non-reference allele and realigned the reads. Reads that aligned uniquely to different locations in these two alignments must be incorrectly mapped in at least one of them. STAR and GSNAP mapped a similar average percentage of reads (89 and 88% of reads, respectively), with greater than 99% of reads mapped to the same location in the two genomes. Using STAR, an average of 0. 19% (±0. 27) of reads mapped to different locations in the reference and the complement genomes, while using GSNAP, an average of only 0. 007% (±0. 0015) of reads mapped differently. This represents a nearly 30-fold decrease in detectable incorrect alignments. In our study, GSNAP required 6-fold more analysis time compared to STAR due to GSNAP's inclusion of known genotypes in the analysis. At disease-associated loci, where allele counts inform our understanding of underlying biology, the increased accuracy at heterozygous sites may prove useful in examining the contribution of regulatory variants. Although the analysis time increased using GSNAP, this allele-aware strategy showed a 30-fold increase in accurate alignments to a subset of heterozygous sites and presumably will detect AEI with greater accuracy.

1798W

Decision Tree Machine Learning Approach to Identify Clinically Actionable Copy Number Events. *S. Shams, A. O'Hara, Z. Che, V. Wasnikar.* BioDiscovery, Inc. , Hawthorne, CA.

Clinical case review is focused on identification of potentially pathogenic copy number alterations while excluding common benign copy number alterations (CNA) and frequently observed array-specific artifacts. This includes comparison of the proband sample results with public databases and in-house archived sample results to exclude copy number polymorphisms and to identify previously observed pathogenic changes. Depending on the quality of the results and the complexity of the sample DNA, call classification can become a time-consuming process. We have developed a Variant Interpretation Assistant (VIA) system that uses a decision tree machine learning approach to set up rules for automatic pre-classification of CNA. This sophisticated tool includes pre-classification of CNA based on user specified designations (defaults include benign, likely benign, VUS, likely pathogenic, pathogenic and artifact); The VIA system then uses region specific functions, such as SIMILARITY, OVERLAP, PREVIOUS_SIMILAR_CASES, and POSITION to allow the user to set up lab specific rules. Benign and likely benign regions are pre-classified based on public copy number polymorphism databases and publications, and previous in-house case classification results. Pathogenic and likely pathogenic regions are pre-classified based on overlap with known pathogenic alterations in previously classified in-house case results, as well as public database and publication records. Alterations over a certain length or those which cover specific genes/exons of interest are pre-classified accordingly. VIA tracks results in real time from previously annotated samples to allow for interpretation of newly uploaded samples based on the evolving in-house results archive. Microarray results on mixed array platforms from normal and affected samples were tested with VIA. A set of previously reviewed samples with a mix of benign and pathogenic classified calls was imported into the system. VIA then pre-classified these same samples and the results were compared. All pre-classification results were annotated with the logic used by VIA to keep a complete audit trail of calling. This system was able to correctly identify pathogenic results for affected samples while finding benign results for normal samples. Final results were then auto-reported in ISCN nomenclature. VIA is capable of quickly and reliably classifying CNA from microarray and includes an extensive audit trail for potential regulatory demand.

1799T

Modeling *Pseudomonas* DNA Integrations into Stomach Adenocarcinoma. *K. B. Sieber[1], P. Gajer[1,2], N. Kumar[1], J. C. Dunning Hotopp[1,2,3].* 1) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201; 2) Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD 21201; 3) Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, MD 21201.

The integration of exogenous DNA into the human genome can produce disease causing somatic mutations. For example, human papillomavirus (HPV) caused ~490,000 cases of cervical cancer in 2002. The integration of HPV DNA into the human genome is a key step to promoting carcinogenesis. In contrast to viral integrations, the instances and repercussions of bacterial DNA integrations into the human genome are less clear. Given that there are 10× more bacterial cells in our bodies than human cells, there is ample opportunity for integration of bacterial DNA into somatic genomes. Previously, using publicly available Illumina sequence data from The Cancer Genome Atlas (TCGA), we identified putative integrations of bacterial DNA in cancer genomes, defined as >4× coverage on the human genome of Illumina paired-end reads with one read mapping exclusively in the human genome and the other read mapping exclusively to bacteria. These data support the specific integration of *Pseudomonas* 16S & 23S DNA into four proto-oncogenes. Here, in order to better understand the structure and sequence of these integrations, we have modeled the most likely configuration of these integrations. These models rely on using the Jensen-Shannon Distance to compare the insert size of the models of bacterial DNA integration with the insert size of the sequencing library. These data support that the integrations are into the 5χUTR near the transcriptional start site of the human genes. Using the model of the bacterial DNA integration into *CEACAM5*, we have constructed the bacterial DNA integration into the human *CEACAM5* promoter in a luciferase reporter construct to test if the bacterial DNA integration may alter transcriptional regulation of *CEACAM5.* Preliminary results suggest *CEACAM5* promoter with the bacterial DNA integrations have altered expression compared to the native *CEACAM5* promoter. This is consistent with our previous observation of higher expression of genes with integrations relative to the same genes without integrations in other stomach adenocarcinoma samples. Further experiments are needed to characterize the other putative bacterial DNA integrations before more general conclusions can be drawn about the effects of bacterial DNA integrations on transcriptional regulation.

1800F

SNooPer: a machine learning-based method for somatic variant identification from next-generation sequencing. *JF. Spinella[1], R. Vidal[1], V. Saillour[1], P. Cassart[1], C. Richer[1], M. Ouimet[1], J. Healy[1], D. Sinnett[1,2].* 1) University of Montreal, Centre Hospitalier Universitaire Sainte-Justine Research Center, Montreal, Quebec, Canada; 2) Department of Pediatrics, Faculty of Medicine, University of Montreal, Quebec, Canada.

The advent of next-generation sequencing has allowed unbiased, in-depth interrogation of cancer genomes. Many somatic variant callers have been developed yet accurate ascertainment of somatic variants remains a considerable challenge as evidenced by the consistently weak overlap between algorithms. Statistical model-based algorithms that are currently available perform well under best-case scenarios, e. g. high sequencing depth, homogenous tumor samples, high somatic variant allele frequency (VAF), but only show limited performance with sub-optimal data such as low-pass whole-genome sequencing data. We propose SNooPer, a highly versatile machine learning approach that uses Random Forest classification models to accurately call somatic variants in low-depth sequencing data. SNooPer uses a subset of variant positions from the sequencing output for which the class is known, either true variation or sequencing error, to train the data-specific model. This implicitly requires that a subset of positions be validated on an independent NGS platform or using orthogonal technology. During the training phase, multiple features including measures of quality, coverage and strand bias are extracted from the mpileups files. Features are ranked by information gain and only informative features are used to build the classification model. Using a real dataset of 40 childhood acute lymphoblastic leukemia patients, we show how the SNooPer algorithm outperforms benchmarked algorithms (JointSNVMix2, VarScan2, MuTect), is not affected by limited sequence coverage or low VAFs, and can be used to reduce overall sequencing costs while maintaining high specificity and sensitivity to somatic variant calling, particularly in low-depth sequencing data. While the goal of any cancer sequencing project is to identify a relevant, and limited, set of somatic variants for further sequence/functional validation, the inherently complex nature of cancer genomes combined with technical issues directly related to sequencing and alignment can affect either the specificity and/or sensitivity of most callers. The flexibility of SNooPer's random forest protects against technical bias and systematic errors, and is appealing in that it does not rely on user-defined parameters. The SNooPer source code is freely available for acadamic users, upon request.

1801W

Toward assembly-based variation discovery from highly divergent regions. *S. Tian, H. H. Yan, S. Slager.* Mayo Clinic, Rochester, MN.

Whole exome sequencing (WES) has been widely used to identify variants implicated in disease. However, current variant discovery approaches often rely on mapped reads alone and whose efficiency is limited by the gaps, misassemblies and underrepresentation of alternative loci in the reference sequence. Obviously, their performance deteriorates in regions with high sequence variation and large INDELs. The human leukocyte antigen (HLA) at 6p21. 3, for example, is associated with over 100 diseases and well known for its extreme divergence. Fully cataloging variants from such highly divergent regions (HDRs) is clinically important but challenging in practice. A few assemblers have been developed to identify variants through direct variant calling, microassembly or whole genome assembly. Nevertheless, these tools mainly focus on whole genome sequence and their feasibility on WES data remains largely unknown. Here we utilize WES de novo assembly for variant discovery in HDRs. Using exonic regions from Chr6, seven 100-bp datasets were simulated at 0. 05-15% divergence rates that reflect the whole spectrum of genome variation. We compared three popular variation-aware de novo assemblers, Cortex, Fermi and String Graph Assembler (SGA), on simulated data and the 150-bp NA12878 exome data released by the Illumina Inc. In parallel, three haplotype-based callers, GATK HaplotypeCaller, FreeBayes and Platypus, were tested together with BWA or BWA-MEM mapping. On the simulated data, the three mapping-based approaches showed 97% SNP calling sensitivity at low (<=1%) divergence but performed poorly (median 61%) at high divergence. Compared to the mapping-based methods, Cortex had much lower sensitivities, even at low divergence (median 78%). We implemented key enhancements to SGA that was originally designed for the assembly of consensus sequence. Remarkably, SGA achieved excellence in both sensitivity (93. 7-95. 6%) and speed. It showed superior performance across all the divergence rates and over a wide range of coverage depths. Fermi was clearly less sensitive than SGA, especially when coverage is under 100x. We recapitulated our findings on the NA12878 exome data, in which SGA achieved comparable SNP calling sensitivity in the HLA (93. 8%) and non-HLA (94. 1%) regions and identified large INDELs from the capture regions. We conclude that SGA represents an ideal platform for variant discovery in HDRs using WES data and should be generally applicable to the entire genome.

1802T

SuperVario: a Common Repository for Variant Data. *A. S. Wenocur[1], L. Tian[1], K. Nguyen[1,4], P. M. Abou-Sleiman[1,2,3], H. Hakonarson[1,2,3].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Center for Dynamic Data Analytics, Rutgers University, Piscataway, NJ.

In many bioinformatic analysis pipelines, nucleic acid reads are mapped, then submitted to a variant caller. Typically, the variant data are stored in flat files, difficult to use and query by multiple investigators. A common problem encountered is that CNVs, point mutations, indels, SNPs, structural, and other variants are tagged for specific analyses and are not easily shared or retrieved for new *insilico*experiments. A solution to this is a centralized repository for genomic variants, which can accommodate multiple classes of variant data. We have developed a database of variants, allowing researchers to apply custom filters and visualize by frequency in a specialized genome browser. The initial prototype was an instance of Exac Browser, modified to load VCFs produced at our institution. The next phase of development for this repository shall be to integrate it with a more fine-grained filtering framework, allowing it to be extended by a developer on-demand. Currently, SuperVario is an evolving framework unifying several bioinformatics tools into a user-friendly querying and visualization platform. SuperVario will store variant annotations and clinical metadata in RDF format, using a SPARQL framework for querying. This currently supplements the MongoDB data store used by the Exac Browser; however, these two data stores ultimately shall be unified. Subsequent releases of SuperVario are to include a system for load-balancing on the cloud, and a custom binary data store and index format for compact storage and rapid querying.

1803F

Analysing the technical sensitivity of indel detection for molecular diagnostics. *N. Whiffin[1,2], R. Govind[1,2], S. Wilkinson[1,2], E. Edwards[1,2], P. J. Barton[1,2], J. S. Ware[1,2], S. A. Cook[1,2].* 1) National Heart and Lung institute, Imperial College, London, UK; 2) NIHR Cardiovascular BRU, Royal Brompton and Harefield NHS Trust, London, UK.

While insertions and deletions (InDels) are known to play a large part in the genetics of human disease, they remain difficult to identify accurately from short-read sequencing data. Different InDel callers have small overlap and particularly struggle with repetitive stretches of DNA. This issue is of particular importance in a diagnostic setting where high sensitivity of variant calling is crucial. To elucidate which of the constantly expanding catalogue of calling algorithms should be used in our diagnostic pipeline, we compared the sensitivity and specificity of six commonly used algorithms (GATK UnifiedGenotyper, GATK HaplotypeCaller, Pindel, Platypus, VarScan2 and Scalpel) across 169 genes sequenced in 89 individuals. In total, 510 InDels were identified in protein coding regions, with 29. 8% identified by all six callers, and 15% identified by only one. Of 22 InDels found uniquely in one sample, 10 (45. 5%) were consistently called by all six callers. The remaining 12 InDels were significantly longer than those universally identified by all algorithms (9. 6 and 3. 5 bp respectively; Wilcoxon P=0. 028), reflecting known difficulties detecting larger InDels. Validation by Sanger sequencing revealed 10 of the 12 inconsistently called InDels to be true variants, including two InDels over 20 bps long. Only HaplotypeCaller and Pindel successfully identified these two large InDels. Overall, HaplotypeCaller showed the best performance, calling all 10 validated InDels (100% sensitivity) and only one false positive. Currently, a barrier to using HaplotypeCaller in diagnostics is the commercial licence fee. Here, we find that using a combination of other callers, including Pindel, the same high sensitivity can be achieved. Previous comparisons of indel callers have often been based on only two callers. We present a comprehensive analysis of up-to-date software solutions. Furthermore, we use real, rather than simulated, data and specifically focus on InDel calling in a diagnostic setting. It is hoped this work will guide scientists in both research and diagnostic settings when choosing a variant caller to identify InDels in sequencing data. *This abstract presents independent research commissioned by the Health Innovation Challenge Fund (HICF-R6-373), a parallel funding partnership between the Department of Health and Wellcome Trust. The views expressed in this publication are those of the authors and not necessarily those of the Department of Health or Wellcome Trust. .*

1804W

ScanIndel: a hybrid framework for indel detection via gapped alignment, split reads and *de novo* assembly. *R. Yang[1], A. C. Nelson[2], C. Henzler[1], B. Thyagarajan[2], K. A. T. Silverstein[1].* 1) Supercomputing Institute for Advanced Computational Research, University of Minnesota, Minneapolis, MN; 2) Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN.

Comprehensive identification of indels across the full size spectrum from second generation sequencing is challenging due to the relative short read length inherent in the technology. Different indel calling methods exist but are limited in detection to specific sizes with varying accuracy and resolution. We present ScanIndel, an integrated framework for detecting indels with multiple heuristics including gapped alignment, split reads and *de novo* assembly. Using simulation data, we demonstrate ScanIndel's superior sensitivity and specificity relative to several state-of-the-art indel callers (including Pindel, GATK HaplotypeCaller, Platypus and Scalpel) across various coverage levels (10X, 20X and 50X) and indel sizes (1 bp to 1 kbp). ScanIndel yields higher predictive accuracy with lower computational cost compared to Pindel for both targeted resequencing data from tumor specimens and high coverage whole-genome sequencing data from the human NIST standard NA12878. Thus we anticipate ScanIndel will improve indel analysis in both clinical and research setting.

1805T

ConVarCal: A Reliable and Robust Platform for Next-Generation Sequencing Variants Identification. *Y. Zheng[1], T. F. Yoshimatsu[1], A. Rodriguez[2], R. K. Madduri[2], P. J. Dave[2], O. I. Olopade[1].* 1) Department of Medicine, The University of Chicago, Chicago, IL 60637; 2) Computation Institute and Argonne National Laboratory, The University of Chicago, Chicago, IL 60637.

Background: In the era of next-generation sequencing (NGS), enormous volume of single nucleotide variants (SNVs), insertions and deletions (Indels), and structural variants (SVs) can be identified at both individual and population levels. To minimize the false positive/negative detection rates, to gain better variant annotation, and to curtail the bottleneck inherent to processing of Big Data, we developed ConVarCal (Confident Variant Calling), a platform for NGS variants identification. **Methods:** ConVarCal integrates multiple NGS reads aligners, variant callers and annotator in a malleable manner using the elastic computing capability of Globus Genomics built upon Amazon Web Services. Raw data (e. g. FASTQ) can be safely transferred via Globus Online and automatically fed into ConVarCal for BWA-MEM or Bowtie2 alignment. The resulting BAM files can be processed by highly parallelized variant calling pipelines, including GATK HaplotypeCaller, Platypus, FreeBayes, SAMtools mpileup, and Atlas2. To realize dynamic parallelization of some callers, wrapper scripts using Swift language are implemented. The output VCF files can then be normalized before employing the Consensus Genotyper to select a set of highly confident SNVs and Indels. SVs can also be detected using tools such as Delly and CONTRA. Variants can be further annotated on ANNOVAR. **Results:** Here we elucidate the performance of ConVarCal by analyzing germline BROCA cancer risk panel sequencing data (Illumina 2500 paired-end, 1. 3Mbp targeted regions, ave. depth of 260x) in a test dataset of 200 Nigerian breast cancer patients with family history and/or early age at onset. FASTQ files (ave. 1. 65 GB per sample) were transferred across continental USA in about 3 minutes per sample at each interval. The entire analysis of 200 samples on ConVarCal were completed in one to two days, depending on the configuration and resource availability of cloud computing. Among 5 callers, ConVarCal confidently identified 14 deleterious SNVs and 11 damaging Indels (*BRCA1*: 15, *BRCA2*: 4, *PALB2*: 2, *BRIP1*: 1, *CHEK2*: 1, *NBN*: 1, *TP53*: 1) in 29 subjects, including the 16 previously known mutations detected by Sanger sequencing. **Conclusion:** ConVarCal takes full advantage of Globus Genomics for NGS variant detection in a reliable and robust manner. It is scalable for larger tasks, and modular to adapt more tools. New tools like FermiKit, VarScan2, Lumpy, and MetaSV are currently being optimized to further enhance our platform.

1806F

A comprehensive SomaticSeq workflow to prioritize biologically relevant somatic mutations in cancer. *L. T. Fang, A. Chibber, M. Mohiyuddin, J. C. Mu, G. Gibeling, S. Barr, N. Bani Asadi, H. Y. K. Lam.* Bina Technologies, Roche Sequencing, Redwood City, CA 94065.

Cancers are diseases of the genome. Accurately identifying somatic variants between the tumor and matched normal tissue is critical in understanding cancer biology. There is a plethora of software tools designed to detect many types of somatic changes occurring in tumor-normal sequencing experiments, such as somatic single nucleotide variations (SNVs), small insertions/deletions (INDELs), copy number aberrations (CNAs), and structural variations (SVs).

We previously presented SomaticSeq, a workflow that incorporates an ensemble of algorithms for SNV and INDEL detection. It implements a stochastic boosting machine learning algorithm to drastically improve accuracy of the combined call sets, and have achieved excellent results in a variety of simulated and real data. We obtained F1 score above 95% in partially synthetic DREAM Challenge data and recalls above 90% in publicly available data from European Genome Archive (EGA), better than any individual algorithm that we have incorporated.

In this study, we expandSomaticSeq to detect somatic copy number aberrations and structural variations in tumor-normal whole genome sequencing data. It incorporates methods such as Control-FREEC, OncoSNP-SEQ, Seq2C, CNVkit, and CREST. It also merges the somatic SV/CNA results with SNV/INDEL to reflect different somatic changes to the same gene. Control-FREEC and OncoSNP-SEQ are designed to detect large-scale somatic CNAs (10 kbps to whole chromosomes) in whole genome sequencing, whereas Seq2C and CNVkit are better suited for whole exome or targeted sequencing. CREST is capable of detecting medium-scale SVs (> 50 kbp) with exact breakpoints in whole genome sequencing. Thus, the comprehensive SomaticSeq workflow has tools capable of detecting somatic variants of any scale.

The comprehensive SomaticSeq allows cancer researchers to specifically prioritize for genes with somatic changes of different types, or where multiple changes have occurred within the same gene. For each gene, the researchers can find somatic mutations, copy number aberrations, structural variations, and use these genetic contexts to infer its biological significance. We have used a number of publicly available datasets from TCGA and EGA as test cases to demonstrate its ability and accuracy performance.

1807W

Identification of novel tumor suppressor candidates in familial cholangiocarcinoma using sequencing-based Megabase-scale haplotypes from germline and cancer genomes. *S. Greer[1], L. Nadauld[2], B. Lau[1], E. Hopmans[1], C. Wood[3], J. Bell[1], H. Ji[1,3].* 1) Stanford Genome Technology Center, Stanford University, Stanford, CA; 2) Intermountain Healthcare, Salt Lake City, UT; 3) Dept of Med/Oncology, Stanford University, Stanford, CA.

Cholangiocarcinoma (bile duct cancer) is an epithelial malignancy originating from the bile duct that connects the liver to the small intestine. Only recently has the molecular genetic underpinnings of this cancer been explored with a number of genome sequencing studies of a limited number of primary tumors. In this study, we sought to identify potential susceptibility genes that have not yet been described and whose germline variants are associated with an increased familial risk of cholangiocarcinoma. We characterize the genome wide variants from multiple members of an extended pedigree with an early onset of cholangiocarcinoma and an inheritance pattern phenotypically consistent with autosomal dominance. As haplotype-resolved genomes of pedigrees would provide higher confidence of Mendelian segregation patterns, we demonstrate for the first time the application of experimentally-derived Megabase-scale haplotypes and phased genome analysis to identify Mendelian disease loci. Our approach relies on a novel technology where individual high molecular weight DNA molecules up to and greater than 50 kb in size are used to experimentally derive haplotypes from whole genome sequencing. This phasing technology uses droplet-partitioned barcodes to tag individual DNA sequence reads. Thus, these barcode sequences can be linked to individual DNA molecules of high molecular weight. This approach enables the genetic phasing of large groups of variants (“phase blocks”) on a Megabase scale. This approach also provides a means to validate structural variant (SV) and copy number variant (CNV) calls from whole genome short-read data. Using this phasing approach in combination with traditional whole genome shotgun sequencing, we analyzed the extended pedigree for germline variants as well as some primary cholangiocarcinomas. We identify haplotypes that co-segregate with affected family members and several novel candidate genes with germline mutations with a predicted deleterious effect. These represent novel loci. Somatic variant analysis of the tumor samples, specifically considering SVs and CNVs, was used to identify driver mutations that may work in concert with candidate germline mutations to initiate and maintain tumorigenesis. Overall, our study has significant implications by shedding light on the genetic basis of cholangiocarcinoma, as well as providing a general method for discovering novel variants and determining its haplotyped segregation in pedigrees.

1808T

Towards reliably detecting structural variants with nanopore sequencing. *A. L. Norris[1], R. E. Workman[2], J. R. Eshleman[1], W. Timp[2].* 1) Pathology and Oncology, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins School of Medicine, Baltimore, MD; 2) Biomedical Engineering and Medicine, Center for Epigenetics, Johns Hopkins School of Medicine, Baltimore, MD.

Despite advances in 2nd generation DNA sequencing (e. g. Illumina), structural variants (SVs) are difficult to reliably detect due to the short read length (<300bp). Not only do the reads – or paired-end reads – need to straddle a breakpoint,but repetitive elements often lead to ambiguities in the alignment of 2nd generation sequencing's short reads. Given that repetitive DNA regions (including centromeres, telomeres, and other repetitive elements) encompass over half (56%) of the human genome, this is a significant concern when mapping SVs. We propose to use the long-reads (~20kb) available via 3rd generation sequencing, specifically nanopore sequencing on the MinION, to detect SVs. Nanopore sequencing relies on a similar concept to a Coulter counter, reading the DNA sequence from the change in electrical current resulting from a DNA strand being forced through a nanometer size pore embedded in a membrane. Though 3rd generation sequencing has a high mismatch rate, precluding base substitution and small frameshift mutation detection, the long reads are superior to 2nd generation sequencing for SV detection. We have tested nanopore sequencing for detection of a series of previously characterized SVs, including large deletions, inversions and translocations that inactivate the *CDKN2A* (*p16*) and *SMAD4* (*DPC4*) tumor suppressor genes. Using PCR amplicon mixes, we have demonstrated that nanopore sequencing can detect large deletions, translocations and inversions at dilutions as low as 1:100, with as few as 500 reads per sample. Given the speed, small footprint and low capital cost of nanopore sequencing, nanopore sequencing could become the ideal tool for the low-level detection of cancer-associated SVs needed for molecular relapse, early detection of cancer, and therapeutic monitoring.

1809F

Finding the fastest route to the right answer: optimizing detection of mutations, copy number changes, and loss of heterozygosity in a single sequencing assay. *B. Peter[1], K. S. Jeong[2], A. Ashutosh[2], C. Le Cocq[2], A. Vadapalli[2], D. Joshi[2], D. Roberts[2], J. Ghosh[2].* 1) Agilent Laboratories, Agilent Technologies, Santa Clara, CA; 2) Diagnostics and Genomics Group, Agilent Technologies, Santa Clara, CA.

Although the cost of Next Generation Sequencing has fallen, there remains a need to improve extraction of the most actionable information, in a more cost and time efficient manner. Different genome variations have different requirements for detection. Sensitive and accurate detection of Single Nucleotide Polymorphisms and Variants (SNPs, SNVs) requires deep coverage with a high number of sequencing reads. However, detection of Copy Number Variation (CNV) and of Loss-of Heterozygosity (LOH) requires broad coverage across the genome. Whole Genome Sequencing (WGS) has the potential to detect the full range of aberrations, but the cost, data analysis burden, and slow turnaround time of deep coverage WGS prevent it from being efficient for clinical research. Target enrichment panels can concentrate sequencing data on regions of interest, but any aberrations falling outside those regions are missed. We created OneSeq, a single sequencing assay combining broad and deep coverage, to generate the most relevant information in the shortest amount of time. First we introduce a genome-wide target enrichment strategy which allows deep coverage on specific regions of interest, while detecting CNVs and LOH across the entire genome. This assay compares an experimental sample to a known diploid reference sample to detect copy number changes. Specific regions of interest, such as disease associated genes, can be targeted by custom SureSelect panels added to the catalog SureSelect backbone panel for higher coverage, allowing sensitive detection of mutations and indels. Second, new computational methods vastly improve the analysis pipeline and reduce result turnaround time from days to hours. Starting with the sequence files, the SureCall desktop software performs an extremely parallelized pipeline composed of optimized BWA-MEM (Li H. 2013), pre-processing, SNP calling and other Agilent algorithms. Typical processing time on a Windows workstation (3GHz/12GB RAM) for a 175M read pairs data set (2x4GB) is only two hours, including alignment, file conversions, SNV, CNV, and LOH calling, annotations, and report generation. We validated this new method on samples with various sizes of known chromosomal aberrations, and compared results to data obtained by low-pass WGS and several microarray platforms. The OneSeq assay, using the fast analysis pipeline, accurately detected the confirmed aberrations in every sample, and simultaneously provided information on SNPs and LOH regions.

1810W

Next-Generation Sequencing Carrier Screen for Alpha Thalassemia Identifies Both Common and Rare Variants. *J. R. Maguire, K. M. D'Auria, H. H. Lai, X. Wang, C. S. Chu, I. S. Haque, E. A. Evans, H. P. Kang, D. Muzzey.* Counsyl, South San Francisco, CA.

INTRODUCTION: Alpha thalassemia (AT)—one of the most common recessive diseases—is characterized by deficient production of alpha globin from the *HBA1* and *HBA2* genes, which have identical tandem coding sequences. Carrier status depends not just on the total number of *HBA* copies in the genome, but also on copy number per chromosome. Current screening methods for AT are low-throughput and high cost, plus they either cannot reflect the underlying locus architecture or have a low signal-to-noise ratio that makes them error prone. We developed an NGS-based assay that identifies all of the major large deletional variants—including the phase and copy number of *HBA1* and *HBA2* genes—as well as SNPs and short indels, giving it a detection rate near 100%. The assay is entirely automated and could, in principle, process thousands of AT samples on a single Illumina RapidRun flow cell.

METHODS: We designed both specific and degenerate hybrid-capture probes that tile the *HBA* locus on Chr16. Depth at each *HBA* probe was normalized to the depth at probes out of the *HBA* region, where average copy number is two, yielding a copy-number profile. *HBA* phase and copy number was resolved by comparing the observed profile against the set of expected profiles for all common deletional variants. The copy number at each position informed which allele frequency to expect during SNP- and indel-finding. During validation, we confirmed results using MLPA.

RESULTS: A validation study with normal, carrier, and affected samples showed perfect concordance between our NGS assay and MLPA. We have since run the assay on nearly 150,000 clinical samples with carrier rates consistent with published literature. For carriers with two *HBA* copies, the assay resolves cis vs trans orientation of the *HBA* genes and, in the cis case, identifies the best-matching ethnic origin of the deletion. Further, it reveals silent carriers who nevertheless have four total *HBA* copies, where the duplicated and deleted genes have different chromosomal offsets. Finally, though most samples have four copies of the *HBA* genes, many have far more copies, and we have observed more than six copies in several samples.

CONCLUSIONS: AT carrier status can be assessed at a nearly perfect detection rate with an automated, inexpensive NGS-based assay. The confidence of calls is substantially greater than MLPA can achieve, and obtaining the same confidence from a SNP array would require that fully half of the chip be devoted to AT alone.

1811T

Confirming Variants Discovered by Next Generation Sequencing (NGS) with Sanger Sequencing Using Innovative Bioinformatics Tools. *E. H. Schreiber, S. Berosik, M. Wenz, S. Chang, J. Zhu, J. Zhang, P. Brzoska.* Genetic Sciences Division, Thermo Fisher Scientific, South San Francisco, CA.

Next-generation sequencing (NGS) analyses have revolutionized our understanding of biological processes. In many basic science or applied research studies, substantive insights have been made by comparing the primary DNA sequence of genes in different groups of subjects. In such studies, researchers attempt to identify the role that variations of nucleotide sequence between the groups may play in disease susceptibility, disease progression, or phenotypic variation. The accurate identification of sequence variants is therefore instrumental in insuring successful experimental outcomes. However, the library preparation and chemistries used in NGS technologies are prone to low, but detectable, levels of sequencing errors that can lead to confusing or incorrect interpretations. Therefore, many labs confirm potential variants identified by NGS by using an orthogonal method such as Sanger-based sequencing: polymerase termination with fluorescent dideoxynucleotides followed by sequence collection on automated Capillary Electrophoresis (CE) instruments. This is a robust and inexpensive method with chemistries and analysis tools that are easy to use and well understood – and due to its high accuracy and ease of data interpretation is considered as the gold standard for DNA sequence analysis. It is therefore an ideal system for validating variant calls made on NGS platforms. To streamline the validation workflow we have developed a web-based Primer Designer tool for PCR and Sanger sequencing that provides pre-designed PCR primer pair sequences for almost all exons in the human exome in an easy to search format. Using these primers researchers can readily PCR amplify and Sanger sequence variants of interest discovered by NGS. The resulting Sanger sequencing data (. ab1 files) can then be analyzed and compared against the NGS vcf variant file using a novel software termed NGC (next generation confirmation). The software aligns the NGS variant finding directly with the variant found in the Sanger sequencing trace for visual inspection and produces a consolidated report for matches or mis-matches as well as a joint vcf file with genome coordinates for downstream bioinformatic processing. In conclusion, the Primer Designer and NGC analysis tools enable a seamless workflow for NGS variant confirmation by traditional Sanger sequencing.

1812F

Detecting Single Exon Deletions in Clinical Whole Exome Sequencing. *T. Chiang[1], P. Liu[2], T. Gambin[2], S. White[1], S. Matakis[4], M. Dahdouli[1], P. Panchal[1], Y. Ding[1], C. Shaw[2], S. Wen[2], W. Bi[2], F. Xia[2], N. Zhiyv[2], R. Person[2], M. Walkiewicz[2], J. Scull[2], J. A. Rosenfeld[2], Z. H. C. Akdemir[2], A. Hawes[1], W. Salerno[1], N. Veeraraghavan[1], D. M. Muzny[1,2], S. E. Plon[1,2], J. R. Lupski[2], C. Eng[2], A. L. Beaudet[2], M. Bainbridge[1], E. Boerwinkle[1,3], Y. Yang[2], R. A. Gibbs[1,2].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, 77030; 3) Human Genetics Center, University of Texas Health Science Center, Houston, TX 77030; 4) Baylor Miraca Genetics Laboratories, Baylor College of Medicine, Houston, TX 77021.

Clinical Whole Exome Sequencing (WES) is becoming routine for the identification of pathogenic variants underlying genetic diseases. Copy Number Variation (CNV) is also often associated with human genetic disease, but routine detection of CNVs in WES remains a challenge. Even the conceptually more straightforward problem of reliable detection of homozygous or hemizygous deletions is not fully solved. We therefore developed methods to reliably detect exon deletions from WES data in a clinical setting. Previous read depth-based approaches for calling WES CNVs have either utilized a match-control set of samples as a reference baseline, or employed analysis on a cohort of samples as a single independent study. Here, we observed that the growing collection of samples from our clinical laboratory that were sequenced with identical methods could be used as the baseline sample of reference. Incoming new samples are therefore added to the reservoir and analyzed within this context. We used a dual normalization approach, consisting of PCA-XHMM intermediates and RPKM Read Depth (RD) values to call deletions. The former is centered by exon on a z-score scale whereas the latter is normalized at the sample level. Both methods remove exons with GC and low complexity biases from the analysis. Complete homozygous/hemizygous exon deletions are called using a strict filtering criterion. In sum, 6477 exomes were analyzed and a total of 2440 single exon deletions (1110 genes) were called in 11% of the samples (median:11 genes/sample). OMIM disorders are associated with 195 of these candidate genes (18%). The majority of samples (80%) contained deletions in only one gene. Incorporating the deletion analysis in known disease genes has boosted the diagnostic rate. In addition, new disease gene discoveries can potentially be made from these data. To extend these methods to reliable detection of heterozygous deletions, we profiled a subset of samples and found that while most exons displayed a normal RPKM distribution, a subset of exons displayed a bi-modal distribution, which is suggestive of polymorphic heterozygous deletions or duplication events. They were flagged as candidate CNVs for further investigation. In a profile screen of 1825 exomes, we report 7335 such candidate genes (22% are OMIM genes) in 1683 samples (92%). Current efforts are aimed at fully resolving these events in order to obtain higher specificity and sensitivity of the overall routines.

1813W

BioNano nanochannel-based genome mapping can detect large pathogenic CNVs on chromosome 22q11. 2 and describe genome architecture in humans. *R. R. Haraksingh, A. E. Urban.* Depts. of Psychiatry and Genetics, Stanford University, Palo Alto, CA.

We used nanochannel-based mapping of 150+ kb fragments of genomic DNA (Bionano Irys platform) to detect large, disease associated CNVs in the human 22q11. 2 locus. The protocol involves linearizing, nick-labeling and uniformly stretching extremely long fragments of unamplified genomic DNA into nanochannels, thereby preserving the long-range architecture of the genome without amplification bias. Massively parallel single molecule visualization via optical flow-through imaging of these DNA molecules is used to create a physical map of the genome and resolve CNVs even within segmental duplications (SegDups). Bionano mapping has been used previously for a variety of bacterial, plant, normal human and cancer genomes, but not yet in cases of large CNVs associated with developmental disorders. Detecting large-scale CNVs in the human genome is paramount for biomedical research and clinical cytogenetics. Currently, CNVs are detected using hybridization-based methods and short-read sequencing-based methods. These methods are largely adequate but have limited detection power in highly repetitive regions of the genome and are subject to genome amplification, hybridization, and coverage biases. Further, current methods often do not provide sufficient positional context for large CNVs, in particular novel insertions or CNVs with breakpoints in SegDups. Determining the relative positions of large-scale structural variants in the genome is necessary to elucidate their functional impact. We used the Bionano protocol to assemble genome maps for three patients in the NIMH Molecular Genetics of Schizophrenia cohort. Two of these genomes are known to contain 2. 5 Mbp duplications and the third contains a 2. 9 Mbp deletion at chromosome 22q11. 2. Large CNVs in this locus are strongly associated with morphological abnormalities and neuropsychiatric disorders such as Schizophrenia and Autism. Bionano analysis resolved the structure of the 22q11. 2 locus in all three genomes. DNA molecules that unambiguously mapped to the reference genome were used to determine copy number by a computational approach analogous to read depth analysis to resolve the duplications and deletion. Molecules that did not uniquely map to the reference genome were retained for 'split read' analysis to further determine the genome architecture at this locus. Nanochannel based genome analysis represents a novel tool for detecting large disease-associated CNVs, even in complex regions of the human genome.

1814T

panelcn. MOPS reaches clinical standards as a CNV detection tool for targeted panel sequencing data. *V. Haunschmid[1], G. Povysil[1], J. Vogt[2], K. Wimmer[2], G. Klambauer[1], S. Hochreiter[1].* 1) Institute of Bioinformatics, Johannes Kepler University Linz, Linz, Austria; 2) Division of Human Genetics, Medical University Innsbruck, Innsbruck, Austria.

Targeted panel sequencing is becoming increasingly important as a cost-effective strategy to identify disease-causing variants in clinical and research applications. While various copy number variation (CNV) detection methods exist for whole-genome and whole-exome sequencing data, highly accurate methods for panel sequencing data that are suitable for clinical purposes are still missing. The challenges with this kind of data are the small size and number of target regions as well as their uneven coverage. For clinical applications a method should furthermore be able to detect both short CNVs affecting only single exons or even just parts thereof as well as longer CNVs that affect multiple exons or even an entire gene. We present panelcn. MOPS for copy number detection which extends our previously developed method cn. MOPS to targeted panel sequencing data. The method is well suited for this type of data since it can estimate technical and biological characteristics influencing the read counts of each targeted region by a mixture of Poissons model. The design of the count windows, the read counting procedure, the parameters of the model and the segmentation algorithm have been optimized for targeted panel sequencing. cn. MOPS supplies integer copy numbers together with probabilities which inform users about the reliability of the copy number estimates. We have tested panelcn. MOPS on simulated and real sequencing data. On 240 simulated data sets, that resembled the characteristics of targeted panel sequencing data, panelcn. MOPS has reached an average accuracy of 99. 96%. The real sequencing data was enriched with the TruSight cancer panel that targets 94 cancer predisposition genes including NF1/2, BRCA1/2 and APC. panelcn. MOPS detected 100% of CNVs known from previous MLPA analyses without any false positives. The size of the CNVs ranged from an 80bp deletion starting in the intron and affecting only part of one exon over duplications of several exons to deletions of 350kb affecting the entire gene. These results show that CNVs in targeted panel sequencing data can accurately be predicted with panelcn. MOPS. Consequently additional biotechnologies to detect CNVs, such as MLPA, can be omitted in order to reduce time and costs.

1815F

***i*PsychCNV: A robust method for copy number variation detection on dried blood spots.** *J. H. Thygesen[1,2], M. Bertalan[1], S. Weinsheimer[1], W. Mazin[1], T. Sparsø[1], T. Werge[1].* 1) Mental Health Centre, Sct. Hans and, Roskilde, DK, Research Institute of Biological Psychiatry, Psychiatric Center Sct. Hans, Roskilde, Denmark; 2) Neuroscience of Mental Health Department, Division of Psychiatry, University Collage London.

Dried blood spot (DBS) has been collected in Denmark for over 30 years, offering uncountable possibilities for population genetics studies. However, DBS genomic analysis offers unusual challenges, which current methods for copy-number variation (CNV) detection are not designed to handle. Existing methods predict large number of false positive CNVs on DBS data, making association analysis infeasible. Here we describe a novel methodological approach, *i*PsychCNV, the first tool designed to predict and analyze CNVs from DBS genomic data obtained via Illumina SNP array. *i*PsychCNV outperforms the widely used algorithm PennCNV on three datasets with different source of genomic DNA: whole blood, DBS and mock data (simulating DBS). Direct comparison of matched whole blood vs. DBS data from four samples reveals that PennCNV and *i*PsychCNV have a ratio of 0. 75% and 45% of true positives, respectively. To evaluate the methods specificity and sensitivity, we generated mock data that simulates 800 different CNV combinations found in Infinium PsychArray BeadChip (Illumina) from DBS data. On mock data, PennCNV has poor performance with 0. 65 area under the ROC curve (AUC), when compared to *i*PsychCNV which has excellent performance of 0. 92 ROC AUC. Traditional CNV prediction methods perform poorly because they rely mostly on Log R ratio signal, which on DBS data can have three times higher standard deviation than observed from whole blood. *i*PsychCNV takes full advantage of B allele frequency distribution, whereby false positive CNVs are reduced. The *i*Psych project includes 80,000 DBS samples from five psychiatric diseases and controls; therefore program functions are designed to manage large datasets, like searching for CNV hotspots, which summarize genomic regions that are more relevant for a specific disease. A support vector machine (SVM) model can be constructed using selected variables from CNV hotspot regions, increasing true positive calls. *i*PsychCNV can improve existing CNV call from other programs by using B allele frequency, hotspots and SVM. *i*PsychCNV is a R package easily installed, implemented and efficiently executed using multiple cores. It performs well for different datasets, offering a robust alternative to the existing CNV prediction methods. *i*PsychCNV is publicly available on Github as an open source project: https://github. com/mbertalan/iPsychCNV.

1816W

The NIH Undiagnosed Diseases Program approach to prioritizing variants from genome-scale sequencing. *D. R. Adams[1,2], C. Lau[2], B. N. Pusey[2], W. Bone[2], E. Valkanas[2], C. Tifft[1,2], C. Toro[2], W. A. Gahl[1,2], T. Markello[2].* 1) NHGRI, NIH, Bethesda, MD; 2) Undiagnosed Diseases Program, NIH, Bethesda, MD.

The NIH Undiagnosed Diseases Program (UDP), now its 8th year, uses a wide variety of resources to make diagnoses and nucleate research for individuals with challenging undiagnosed diseases. Genome-scale sequencing has proven to be a valuable tool for both pursuits. Typical UDP study participants have been evaluated extensively before arrival. Prior evaluations often include targeted sequencing, gene panels and/or clinical exomes. As a result, most genome-scale sequencing results produced by the UDP do not have readily interpretable clinical significance. Instead, they form a body of starting points for research into new diseases and atypical disease presentations. Prioritization of these promising but non-diagnostic DNA sequence variants remains a persistent challenge. Our approach to the selection of research candidates incorporates individualization of cases beginning with the selection of target family members. A particular emphasis is placed on phenotyping genetically informative relatives. SNP chip data is used extensively both for copy number variant detection, recombination mapping and other procedures. Once sequencing is complete, called variants are culled by automated analysis followed by manual curation of BAM file data for each finding. The list of manually curated variants is then sorted by senior bioinformatics staff based on known-disease implication of the associated gene and other bioinformatic characteristics. Finally, a multidisciplinary meeting is used to discuss each variant. The meeting includes the clinicians who evaluated the patient, the bioinformatics staff who processed the data and selected members of the UDP translational laboratory who will be tasked with performing follow up validation experiments. Variants are ranked based on multiple parameters including potential phenotypic match, research tractability, resource utilization, and expertise availability (both local and via collaboration). Finally, variants are placed into one of three categories: high interest variants suitable for immediate action, medium interest variants suitable for promotion through match-making websites and low interest variants that will be included in routine dbGaP submissions only. We present a case-based description of our ongoing efforts to optimize this process and hope that it will contribute to a generalized discussion of frameworks for evaluating research-level sequencing findings.

1817T

A combination of targeted enrichment methodologies for whole-exome sequencing reveals novel pathogenic mutations. *F. Miya[1], M. Kato[2,3], T. Shiohama[4], N. Okamoto[5], S. Saitoh[6], M. Yamasaki[7], D. Shigemizu[1], T. Abe[1], T. Morizono[1], K. A. Boroevich[1], K. Kosaki[8], Y. Kanemura[9,10], T. Tsunoda[1].* 1) Center for Integrative Medical Sciences, RIKEN, Yokohama, Kanagawa, Japan; 2) Department of Pediatrics, Yamagata University Faculty of Medicine, Yamagata, Japan; 3) Department of Pediatrics, Showa University School of Medicine, Tokyo, Japan; 4) Department of Pediatrics, Graduate School of Medicine, Chiba University, Chiba, Japan; 5) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 6) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 7) Department of Pediatric Neurosurgery, Takatsuki General Hospital, Osaka, Japan; 8) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan; 9) Division of Regenerative Medicine, Institute for Clinical Research, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 10) Department of Neurosurgery, Osaka National Hospital, National Hospital Organization, Osaka, Japan.

Background: Whole-exome sequencing (WES) is a useful method for identifying disease-causing mutations. However, often no candidate mutations are identified using the commonly available targeted probe sets. In a recent analysis, we also could not find candidate mutations for 20. 9% (9/43) of our pedigrees with a congenital neurological disorder using pre-designed capture probes (SureSelect V4 or V5). One possible cause is that standard WES does not sequence all protein-coding sequences (CDS) due to capture probe design and regions of low coverage, which account for approximately 10% of all CDS regions. **Methods:**We combined a selective circularization-based target enrichment method (HaloPlex) with a hybrid capture method (SureSelect V5; WES) to increase the coverage of CDS regions and identify pathogenic mutations for pedigrees with no candidates through standard WES analysis. We designed a probe set for complementary custom CDS sequencing (CCCS), and applied this combination approach to 7 (SureSelect V5) of the 9 pedigrees. These 7 pedigrees are families with a congenital neurological disorder (microcephaly,lissencephaly,ventricular dilatation,bilateral perisylvian polymicrogyria(BPP) and/or hydrocephaly). **Results:**The combination method predicted a 98. 6% coverage of CDS regions by probe design and produced an observed coverage, with read depth $\geq$10, of 97. 4%. We identified novel pathogenic mutations in one pedigree with microcephaly (OMIM #608716) using this combination method. The mutation was composed of two nonsense variants in *ASPM* gene, c. 8098C>T [p. R2700*] and c. 10168C>T [p. R3390*], identified through WES and CCCS analysis, respectively. The variant call of latter locus was missed in standard WES as the region had low read depth (<10) in all samples. The variants were validated for all six individuals in the family using Sanger sequencing. The parents were carriers of one variant each, the unaffected children showed no mutations and the mutations were identified in both the affected children. **Conclusion:**Our combination method of WES and CCCS contributed to an increase in the coverage of CDS regions (97. 4%) and successfully identified a novel pathogenic mutation. We believe that this combination method will contribute to the identification of novel causative mutations, which are inaccessible by current pre-designed standard WES analysis, and to a greater understanding of mechanisms behind, and therapies for, disease.

1818F

Trans-ethnic analysis of complex traits: analysis approaches by the Population Architecture Using Genetics and Epidemiology (PAGE) II Study. *S. Rosse[1], N. Zubair[1], A. Martin[2], R. Walker[3], J. Kocarnik[1], G. Wocjik[2], C. Gignoux[2], M. Graff[4], L. Xia[5], N. Franceschini[4], C. Haiman[5], U. Peters[1], R. Loos[3], C. Kooperberg[1], E. Kenny[3], C. Carlson[1], Population Architecture using Genomics and Epidemiology Phase II (PAGE II).* 1) Public Health Genetics, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 3) Preventive Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC; 5) Keck School of Medicine, University of Southern California, Los Angeles, CA.

The Population Architecture using Genomics and Epidemiology, Phase II (PAGE II), Study was initiated in 2013 by the National Human Genome Research Institute to expand our understanding of well-replicated genetic disease associations in several large, ethnically diverse and well characterized population-based studies. PAGE II investigators harmonize phenotypes across studies and perform genetic analyses on a variety of outcomes. Approximately 1. 5 million variants will be genotyped in up to 51,750 DNA samples from African-American, Hispanic/Latino, Asian/Pacific Islander, and American Indian participants. Our objective was to design a genotyping platform that would allow genetic characterization of multiethnic samples in order to refine genetic associations with metabolic, cardiovascular, renal, inflammatory, anthropometric, and lifestyle traits. While PAGE II will perform 'agnostic' genomewide scans, we also customized the MultiEthnic Genotyping Array (MEGA Chip), for functional discovery by prioritizing known risk loci, clinically relevant mutations, putative risk alleles, and other functionally important variants. The custom content included on the MEGA Chip will allow researchers to prioritize likely functional variants for laboratory follow-up and identify new independent alleles positioned in known disease loci. MEGA Chip is comprised of a backbone containing Illumina's Humancore array with approximately 300,000 variants, the standard Illumina exome content of approximately 240,000 variants, the African Diaspora chip (700,000), a multiethnic GWAS scaffold (360,000), multiethnic exomic content (200,000), and over 50,000 custom content variants. Here we describe both the custom content of the MEGA Chip and the PAGE II consortium, including its goals, methods, and study design, highlighting how PAGE II can contribute to understanding the genetic architecture of confirmed associations across populations.

1819W

Use of whole genome sequence analysis following uninformative exome sequencing to uncover potential causal variants in siblings with interstitial lung disease. *B. A. Ozenberger[1,2], D. C. Koboldt[1], T. J. Nicholas[1], D. Wegner[3], J. Wambach[3], R. Fulton[1], R. K. Wilson[1,2], J. Coldwell[4], F. S. Cole[3].* 1) McDonnell Genome Institute, Washington University, Saint Louis, MO; 2) Department of Medicine, Washington University, Saint Louis, MO; 3) St Louis Children's Hospital, Department of Pediatrics, Washington University, Saint Louis, MO; 4) Department of Genetics, University of Oklahoma at Tulsa, Tulsa, OK.

The diagnosis of rare inherited pediatric disease often includes candidate gene sequencing assays, but these assays have limitations which often result in the absence of a definitive etiology. In genetically complex disorders, gene panel testing can be used to examine simultaneously many genes known to be potentially associated with a phenotype. More recently, the entire protein-coding territory (exome) may be interrogated in cases for which other diagnostic assays have failed to reveal a cause. The success rate of gene panel and exome assays depends on many factors, such as the phenotype, the specific assay, and the availability of family members for segregation analysis, but is disappointingly low, averaging <25% across diverse phenotypes. The absence of a diagnosis in some cases may be due to technical limitations of the sequencing method employed. Exome capture analysis can effectively reveal single-nucleotide variants but only in the genomic regions targeted. Furthermore, large structural variants (copy number alterations, insertions, deletions) are not readily measured during exome analysis. Whole genome sequencing (WGS) may be a more attractive option in cases of undiagnosed disease because the genome is investigated to maximal extent and WGS methods allow for effective interrogation for all variant types including structural alterations. Regardless, WGS analysis for clinical diagnosis remains uncommon at present. Full siblings presented with severe respiratory symptoms during infancy, underwent chest computed tomography (CT) scan and lung biopsy, and were diagnosed with interstitial lung disease. Parents were healthy without respiratory symptoms and family history was negative. The occurrence of affected siblings with healthy parents suggested a genetic etiology with possible recessive inheritance. Exome sequencing of both affected siblings was unrevealing. The family (siblings and parents) subsequently consented to participate in a WGS study. We present a comprehensive analysis for candidate causal variants using Illumina X WGS data and discuss the advantages of WGS over exome or gene panel strategies for the investigation of rare inherited diseases.

1820T

High speed, low cost processing of WGS data in Alzheimer's disease patients along with integrative analyses enables novel insights into risk mechanisms. *N. D. Beckmann[1], B. Tolhuis[2], J. Karten[2], T. Karten[2], Z. R. Giles[1], H. Shah[1], R. Tanzi[3], E. E. Schadt[1].* 1) Icahn School of Medicine at Mount Sinai, New York, NY; 2) GENALICE, Deventerweg 9d, 3843 GA Harderwijk, the Netherlands; 3) Massachusetts General Hospital-East Aging and Gen. Research Unit, Bldg. 149, 6th Fl. 149 13th Street Charlestown, MA 02129.

Alzheimer's Disease (AD) is a neurological disorder characterized by progressive memory loss. It is the leading cause of dementia, with an estimated incidence of 22 million cases that is projected to double in the next 20 years. Senile plaques in the brain consisting of accumulated Amyloid Beta and neurofibrillary tangles (hyperphosphorylated Tau protein) are the hallmarks of AD. AD is highly heritable, with several known strong genetic components of the disease already identified, but smaller-to-moderate sized effects remaining to be discovered. Here, we compose a database of over 3000 whole genome sequences (WGS) establishing the largest cohort of WGS for AD to date. Variant calling on large cohorts of WGS is impractical with standard variant calling methodologies. A novel proprietary pipeline, GENALICE MAP, is used to uniformly call variants across all samples. GENALICE is a single stream process that offers a significant improvement over standard pipelines employing GATK (>80 hours on one core per genome) in a more traditional high performance compute cluster, both in terms of computing time (6 minutes per genome), memory usage and disk space, while keeping high variant sensitivity (99. 5% on benchmark NA12878 genome). We were able to process all 3000 genomes in 3 days with a cost-reduction of more than an order of magnitude compared to GATK on Amazon EC2. An extensive comparison of variants called with GATK versus GENALICE demonstrated both pipelines performed similarly (93% concordance). Given the drastically reduced computing requirements, we can uniformly call variants in large cohorts of whole genomes, significantly reducing confounds and batch effects from multiple calling algorithms to increase the power of variant detection. With the rapid processing of large numbers of whole genomes, we identified broader sets of candidate AD variants than could be identified in single datasets processed independently. We used an integrative biology annotation strategy that involved calculation of CADD scores and eQTLs in disease relevant tissues from the AMP-AD consortium, assessing significance in single variant association studies for more common variants, gene-based and gene-set enrichment analyses, and causal network models of AD to identify novel genes and pathways implicated in AD. These uniformly called and annotated variants placed in an appropriate molecular context provide novel insights into the risk mechanisms of AD.

1821F

Utilization of PacBio long-read sequencing in comprehensive genomics. *H. Doddapaneni[1], A. English[1], M. Wang[1], C. R. Beck[2], W. J. Salerno[1], Q. Meng[1], Y. Han[1], E. A. Boerwinkle[1, 3], J. R. Lupski[1, 2], D. M. Muzny[1], R. A. Gibbs[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030, USA; 2) Dept. of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, 77030, USA; 3) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, 77030, USA.

Accurate identification of chromosomal alterations ranging from single nucleotide variants (SNVs) to large structural variants (SVs) is essential for comprehensive understanding of genomic architecture and disease mechanisms. At the Baylor College of Medicine Human Genome Sequencing Center (HGSC) we have developed applications for PacBio long-read sequencing technology to detect SVs. A novel method, Pacific Biosystems large Insert targeted capture-sequencing (PB-LITS) combines DNA target-capture enrichment technology with PacBio large-insert library preparation procedures for studying SVs at specific chromosomal regions (Wang et al. PMID: 25887218). This method provides deep sequence coverage with PacBio long reads at substantially reduced cost compared to whole genome PacBio sequencing. We have successfully analyzed target-capture reads through our analysis pipeline PBHoney (described below) to delineate the breakpoint junctions of low copy repeat (LCR)-associated complex SVs on chr17p11. 2 in patients diagnosed with Potocki–Lupski syndrome (PTLS; MIM#610883). PBHoney is a long-read SV detection software that leverages multiple detection and evaluation approaches to report SVs (English et al. PMID: 24915764). In addition to calling SVs with PacBio data from targeted sequencing, the method has also been applied to whole genome human haploid and diploid sequencing experiments. For the cell-line CHM1-tert, PBHoney made 9,496 SV calls between 50 bp and 100 kbp from 40x haploid PacBio coverage. When compared to the previously reported 12,047 SV calls from the same data (Chaisson et al. PMID: 25383537) and Illumina SVs generated from our multiple caller method, Parliament (English et al. PMID: 25886820), we found that PBHoney has a sensitivity of 66. 8%, and the highest positive predictive value at 70. 3%. When applied to a well characterized diploid human sample (HS1011), the software identified 9k deletions, 15k insertions, and 338 inversion calls with 20x coverage (i. e. , 10x per haploid genome) data input. By comparing PBHoney's results with SVs identified from Illumina short-reads and other methodologies, we illustrate the additive value of PBHoney and PacBio data. All of these newly accrued research and development experience has demonstrated that use of long-read sequencing alone or as a complementary data source for SV detection can provide a comprehensive assessment of genomic architecture and enhance our understanding of genetic disease mechanisms.

1822W

Integrative calling of short indels encompassing a wide spectrum of tandem repeats via fuzzy left-right alignment. *H. M. Kang, A. Tan.* Biostatistics Dept, Univ Michigan, Ann Arbor, Ann Arbor, MI.

Statistical models and computational approaches for detecting short insertions and deletions (indels) have rapidly evolved to accurately detect simple biallelic indels mostly located in >97% of non-repetitive regions of the genome. However, ~50% of indels reside in <3% of repeat-rich regions of the genome, mostly in the form of variable nucleotide tandem repeats (VNTRs). Due to increased sequencing errors, alignment errors, and PCR slippages in repeat-rich regions, calling indels in these regions of genome is extremely challenging and has not been the major focus of most indel calling algorithms. Here we propose novel methods to detect or integrate indels encompassing repeat-rich regions systematically. The key idea of our method is to use the difference between left-alignment and right-alignment of indel alleles to systematically classify VNTR based on repeat tract length and repeat units. To model VNTRs with inexact repeats (due to sequencing errors and/or true variants within repeat tract), we extend this method into "fuzzy left-right alignment" algorithm that allows mismatches in left- and right- alignment to when repeat tract length can be substantially increased. Repeat units can be determined by prefix tree implemented in a computationally efficient manner. Compared to the methods relying only on reference sequences, such as *lobSTR* or *RepeatSeq*, our method substantially improves the sensitivity to identify inexact VNTRs and VNTRs located outside of repeat-rich regions focused by lobSTR. We applied our algorithm to deeply sequenced 1000 Genomes samples. The concordance across three different calling algorithms – *HaplotypeCaller*, *Platypus*, and *samtools* - were 56% and 90% within and outside repeat-rich regions of genome (defined by *TandemRepeatFinder* and *mDUST*), respectively. Our VNTR-aware normalization on the same call sets increased the concordance to 75% and 95%, respectively. Recalling the variants with our VNTR-aware variant detection algorithm increased the sensitivity of VNTR detection by 11% compared to *HaplotypeCaller*, when evaluating based on the VNTRs overlapping with *lobSTR* calls. Our proposed VNTR-aware variant normalization and detection methods are implemented in our public sequence analysis software packages *GotCloud* and *vt*.

1823T

Mitigating Batch Effects in Whole Genome Sequencing. *J. Tom, W. Forrest, J. Reeder, J. Hunkapiller, T. Bhangale.* Bioinformatics and Computational Biology, Genentech, South San Francisco, CA.

With the arrival of the 1000 dollar whole genome sequence (WGS) for humans, large WGS datasets are now feasible. Unfortunately, not all sequencing is directly comparable – not even sequencing run solely using Illumina technology. Batch effects from chemistry changes, number of PCR cycles, and type of sequencer can all impact variant calling. Systematic biases in variant calling between batches complicate downstream analysis e. g. by producing false positive disease association. Often only genotype data is available for analysis (as opposed to the raw read or alignment data) and it's necessary to characterize batch effects based on genotypes alone. We have analyzed WGS data with target coverage of 30X for 1245 samples sequenced using Illumina technology. These samples are from five different vendors spanning at least four major differences in library prep and chemistry. We have computed various metrics using genotype data such as transition transversion ratio (Ti/Tv), percent confirmed in 1000 genomes, and ancestry coefficients to identify batch effects and have tested filters that can be applied to mitigate them. Ancestry coefficient estimation is necessary to disambiguate observed differences in metrics due to heritage from those due to batch effects. We have determined that removing calls with genotype quality less than 90 is the most effective filter for mitigating batch effects. In our samples we found, on average, improvements in Ti/Tv in coding regions from 2. 97 to 3. 09 and increases in percent confirmed in 1000 genomes from 80. 4% to 84%. Heterozygote concordance in technical replicates improves from 90. 88% to 99. 8%. Imposing this filter removes a modest percent of variants (16. 8% on average). We have implemented our method to compute these metrics in an R package genotypeeval that will be available from Bioconductor.

1824F

Challenges and approaches to genotyping large biobank projects. *T. Webster[1], J. Brodsky[1], D. Oliver[1], D. Petkova[2], C. Freeman[2], C. Bycrof[2], J. Gollub[1], A. Ollmann[1], A. Qiu[1], L. Reynolds[1], B. Eynon[1], J. Schmidt[1].* 1) Affymetrix, Santa Clara, CA; 2) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK.

As part of major national healthcare initiatives, recent biobank projects are designed to genotype large numbers of individuals. The UK Biobank is a prospective cohort study of over 500,000 individuals from across the United Kingdom. The custom UK Biobank Axiom® Array, with over 800,000 markers, was designed for the project. Due to the size of the study, genotyping was performed in batches of approximately 4,700 individuals who were genotyped on approximately 50 Axiom® 96-sample array plates, for a total of approximately 100 batches. This genotyping strategy, while beneficial for intra-batch consistency and computational tractability, increases the probability that a fraction of the markers will have inconsistent or missing genotypes in some of the dataset. However, the availability of multiple batches processed under the same strict guidelines, provides opportunities for new analytical workflows to improve the accuracy of all genotypes obtained as part of a large sample cohort. We have developed a workflow to increase accuracy and batch-to-batch consistency and reduce the amount of missing data in very large genotyping projects. The workflow consists of two rounds of genotyping. After the first round of standard genotyping, all batches are analyzed to select an exemplar batch for each probe set (the probes that interrogate a marker) as the source of SNP Specific Prior (SSP) information. These SSPs are used by the AxiomGT1 algorithm to improve consistency and accuracy in the other batches in the second round of genotyping. Other probeset specific modifications to the genotyping algorithm, such as modification of algorithmic parameters or advanced normalization to attenuate effects such as plate-to-plate variation, are applied to selected probesets, as well. A final round of analysis excludes a small percentage of probesets that produce sub-optimal clusters, and also probe sets with cluster patterns consistent with complex genetics at the marker site, resulting in more than three genotype clusters. We show that the new workflow decreases the missing information per batch from approximately 3% to less than approximately 2%, increases the ability to detect rare genotypes, and increases the consistency of allele frequencies across multiple batches.

1825W

Description Extractor: Automated HGVS-recommended sequence variant description. *P. E. M. Taschner[1], J. Vis[2,3], M. Vermaat[4], J. N. Kok[3], J. F. J. Laros[4,5].* 1) Generade Center of Expertise Genomics, University of Applied Sciences Leiden, Leiden, Netherlands; 2) Department of Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands; 3) Leiden Institute of Advanced Computer Science, Leiden University, Leiden, The Netherlands; 4) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 5) Leiden Genome Technology Center, Leiden University Medical Center, Leiden, The Netherlands.

The Mutalyzer sequence variation nomenclature checker1 was developed to promote error-free reporting of sequence variants. Mutalyzer applies the Human Genome Variation Society's (HGVS)2 standard nomenclature, which is recommended by the ACMG for variant descriptions. Nonetheless, construction of variant descriptions still requires basic knowledge of the standard nomenclature and manual comparison of reference and variant sequences. This leaves room for errors in reporting, which is highly undesired in relation to the diagnosis of genetic disease. The chance of finding a complex variant increases with the rise of long read sequencers. The newly developed Description Extractor3 helps users to describe complex variants. The algorithm closely follows the human approach for sequence comparison. It will first find the "region of change". Within this region it finds the largest overlap, producing two regions that need to be described in a similar way. This process is repeated until a compact description is found. The tool ensures that the same description will be generated for the same observed variant. The significant advantage: knowledge of the HGVS nomenclature is no longer required to generate correct descriptions. This not only helps clinicians to generate correct descriptions, but its implementation also allows automation of the description process. Database curators can use the tool to describe complete alleles or haplotypes. In addition, they can list the differences between reference sequence versions or even complete genome builds. The Description Extractor and Mutalyzer are freely accessible via the internet and available for download for local installation. 1) https://mutalyzer. nl/2) http://www. hgvs. org/mutnomen/3) https://mutalyzer. nl/description-extractorThis work was supported by the Dutch national program COMMIT.

1826T

Accuracy of Variant detection in Next-generation sequencing: a comparison between exome and RNA seq. *I. trotta, h. almusa, m. lepistö, a. palva, s. hannula, a. vaittinen, p. mattila, p. ellonen, c. heckman, j. saarela.* Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finl, Helsinki, Finland.

NGS approaches are currently widely used in clinical genetics for the rapid detection of rare deleterious variants in monogenic disorders. In order to assess the limitations of the methods for a reliable detection of variants in a diagnostic perspective we compared whole-exome (WES) and RNA sequencing in 27 samples. We examined only the coding variants included in the WES target (*Ensembl v75* genomic sequence) with sequencing depth >20X. We identified 14,772 positions (4,937 unique variant sites) in which the 2 methods produced a different call. We excluded the cases with a putative biological explanation (monoallelic expression), which left 1,340 discrepancies (527 unique) for further analysis, including the follow up by Sanger sequencing of 30 variants. The variant calling process was performed using GenomeAnalysisTK-3. 1-1 (GATK, BWA 0. 6. 2 aligner) in WES, and Tophat v2. 0. 6 and Samtools 0. 1. 18 (Bowtie2 2. 0. 2 aligner) in RNA sequencing. We showed that **15. 6%** of differences (n=209/1,340) could be explained by errors in the alignment or variant calling: 44 missed calls in WES and 165 in RNA seq. Indeed, we observed some variant caller specific challenges: the co-occurrence of heterozygous SNPs and short deletions in the same position prevented GATK to call the SNPs (n=4), while TopHat failed to call short indels located in proximity of the exon-intron boundaries (n=55). We further examined the putative non-unique mapping of the regions flanking the discordant variants (100bp seeds) and found 694/1,340 sites (**51. 8%**) to map in more than one location in the genome. The remaining **32. 6**% of the discrepancies (n=437, 248 unique) could not be explained by evident technical faults. Of them, 40 (4 unique) had been identified as RNA editing sites based on the *RADAR* database. In a total of 62 unexplained discordances we observed at least 3 reads with the alternative allele in WES with a frequency under the threshold required for the call, but in 329 cases no reads with the alternative allele seen in the transcripts sequence were observed in WES despite of >100x coverage of some of the sites. A total of 68 identified discordances referred to genes reported in the OMIM database (in association with 39 disorders), of which two were defined as pathogenetic according to the ClinVar classification. Altogether, these considerations suggest a thorough evaluation of the disease genes targets for NGS, considering the potential drop of causal variants for technical limitations.

1827F

Genotype Imputation Using Sparse Partial Least Squares Analysis. *H. Koh[1], D. V. Conti[2].* 1) New York University, New York, NY; 2) University of Southern California, Los Angeles, CA.

Genotype imputation techniques have been widely used for genetic researchers to obtain denser genetic data for the analysis of genome-wide association studies (GWAS) with much less financial and temporal burden than when they use higher coverage sequencing procedures. Here, we introduce a new model-based genotype imputation method using sparse partial least squares discriminant analysis (SPLS-DA) logistic regression models, termed as 'SPLS-IMPUTE'. The family of sparse partial least squares (SPLS) analysis tools has recently been developed by employing variable selection along with dimension reduction built in the traditional partial least squares (PLS) methods. We noticed that this additional variable selection property for SPLS analysis has facilitated much faster and more accurate analysis in the context of regression-based modeling using high-dimensional data. Taking such advantages of SPLS analysis, the SPLS-DA logistic regression model is used for the machinery of SPLS-IMPUTE by examining the association structure, linkage disequilibrium, among SNPs to tie underlying genotypes in a reference panel and a GWAS study sample. We emphasize that SPLS-IMPUTE does not require haplotype estimation, also known as phasing, for a reference panel or a GWAS study sample which is a major time component required from two benchmark methods, MACH and IMPUTE2. As a result, SPLS-IMPUTE can be used more computationally efficiently especially in a situation where a phased reference panel is not immediately available. We also observe a near perfect agreement of best guessed imputed genotype calls with true genotypes in our leave-one-out simulation experiment (Discordance rate: 3. 98% for SPLS-IMPUTE (K=10) and 3. 76% for SPLS-IMPUTE (K=15).

1828W

De novo and somatic indel variant analysis of whole genome and exome capture sequencing experiments with Scalpel. *H. Fang[1,3,4], E. Grabowska[2], K. Arora[2], V. Vacic[2], M. Zody[2], I. Iossifov[1], J. O'Rawe[3,4], G. Lyon[3,4], M. Wigler[1], M. Schatz[1], G. Narzisi[1,2].* 1) Simons Center for Quantitative Biology, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY; 2) New York Genome Center, New York, NY; 3) Stanley Institute for Cognitive Genomics, One Bungtown Road, Cold Spring Harbor Laboratory, NY; 4) Stony Brook University, 100 Nicolls Rd, Stony Brook, NY.

Insertions and deletions (indels) are the second most common of variation in human genomes and the most common structural variants. Despite their biological relevance, indels of more than five bases are still challenging to discover. Scalpel is an open-source algorithm (http://scalpel. sourceforge. net/), which combines read mapping and assembly for sensitive and specific discovery of indels in whole genome and exome data. A detailed repeat analysis coupled with a self-tuning *k*-mer strategy allows Scalpel to outperform other state-of-the-art approaches for indel discovery, particularly in regions containing near-perfect repeats. We showed that accuracy of indel detection with WGS is greater than WES even within the targeted regions. Short tandem repeats, especially homopolymer A/T runs are a major source of errors, and they are highly enriched in the WES data. PCR-free protocols help reduce such biases and improve the accuracy of indel calling. Further, common practice of sequencing WGS at 30X using the HiSeq platform is not sufficient for sensitive indel discovery, resulting in at least 25% false negative rates. Scalpel has been successfully used to identify novel candidate genes for autistic spectrum disorder. We analyzed 593 families for inherited and de novo indels from the Simons Simplex Collection, demonstrating Scalpel's power to detect de novo (PPV= 82%) and long transmitted events (PPV= 78%). Consistent with Iossifov *et al*. 2015, we confirm a significant two-fold enrichment for likely gene-disrupting (LGD) indel mutations: 35 frameshifts in autistic children versus 16 in unaffected siblings (P-value = 0. 01097). We found a notable overlap between the LGD target genes and the 842 FMRP-associated genes. We have recently extended Scalpel to detect somatic mutations from tumor and matched normal data and compared its performance to other widely used somatic callers such as SomaticIndelDetector, Strelka, Pindel, Platypus, using data from the DREAM challenge. We characterized false positive and false negative calls in the context of sequence composition and VAFs distributions. . Our results show that Scalpel has the best overall balance between sensitivity and specificity, and recovers somatic indels across the whole length spectrum while keeping a low FDR. Other recent improvements include improved support for highly variable coverage regions, such as targeted sequencing panels, as well as greatly improved runtime performance.

1829T

Platinum Genomes: A comprehensive package for assessing variant calling performance on GRCh37 and GRCh38. *E. Fritzilas[1], M. A. Bekritsky[1], P. Krusche[1], M. A. Eberle[2].* 1) Illumina Cambridge Ltd. , Chesterford Research Park, Saffron Walden, Essex, CB10 1XL, UK; 2) Illumina Inc. , 5200 Illumina Way, San Diego, CA 92122, USA.

Advances in high-throughput sequencing technology now make it possible to rapidly sequence individuals and call variants ranging from single base substitutions to large structural events. Since these sequencing technologies will likely be widely used for clinical applications in the near future, it is essential that we have the means to systematically assess the accuracy of the resulting variant calls. By sequencing a large, three generation pedigree, the Platinum Genomes exploits Mendelian inheritance to create a set of phased high-confidence variant calls, thereby enabling researchers worldwide to characterize the human genome more accurately. Using pedigree phasing in addition to having a replicate analysis allows us to cover a substantial portion of the genome with gold-standard calls, enabling researchers worldwide to characterize variant calling in the human genome accurately. In December 2013 the GRCh38 reference sequence was published, with several improvements over GRCh37. Here we present a new version of Platinum Genomes built using GRCh38, providing researchers with a much-needed resource to assess variant calling performance in the latest version of the human reference genome. Having truth sets for both GRCh37 and GRCh38 also enables us to perform a detailed comparison of the consequences of switching to the new reference, as it allows accurate benchmarking of individual variant calling pipelines on both reference versions. Moreover, analyzing the high-confidence calls unique to either assembly will reveal cases where the changes introduced in GRCh38 have resolved alignment artefacts and led to more accurate variant calling. Overall, we observe ~3. 6 million platinum SNP calls and ~590,000 indel calls per individual in the platinum truth set, comparable to the number of platinum variant calls from GRCh37. In addition to comparing the overall numbers, we lift over the variants from each truth set to quantify the causes and locations where variants fail to occur in both genome builds. VCF files with high-confidence Platinum Genomes variant calls are publicly available on www. platinumgenomes. org. Illumina also offers a companion open source tool that enables benchmarking of a query VCF against the Platinum Genomes truth data in terms of sensitivity and specificity: github. com/sequencing/hap. py.

1830F

Enforcing high quality variant calling in large scale whole genome sequencing. *Z. Huang[1], N. Rustagi[1], L. A. Cupples[2], R. Gibbs[1], E. Boerwinkle[1,3], F. Yu[1].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 3) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX.

Discovery of rare SNPs and novel genes is of critical importance in the disease associationstudy and whole genome sequencing (WGS) with thousands or larger sample size isindispensible. However, large scale WGSvariant calling has many challenges scientifically and computationally. In this work, we addressed the challenge in achieving high quality variant calling. The calling quality is affected by caller specific bias, population substructure, and the sequencing and mapping errors amplified by the sample size. These factors hamper the detection power of population specific rare variants. In order to call high quality rare variants, we developed 1) a novel ensemble calling approach, goSNAP, which employs multiple callers (each in joint calling mode) and features a consensus filter to ensure high quality site discovery, 2) a step-by-step quality control procedure following site discovery, genotype likelihood calculation and imputation, and 3) the optimal strategy to apply site level Hardy Weinberg exact test (HWE) which filters common sites affected by mismapping but keeps the population specific rare variants intact. We apply these methods tothe SNP calling of5297 WGS samples (36% African Americans (AfAm) and 64% European American (EuAm)) with 6-10x coverage in theCHARGEproject(theCohortsfor Heart and Aging Research in Genomic Epidemiology). The whole process only took 6 weeks and 5. 2million CPU hours. We called 72. 2 millionSNPs witha transition-to-transversionratio(Ti/Tv)of2. 14. With large sample size per population, we obtain high novel SNP rate, 49. 4% compared to 1000 Genomes Project (phase3) and 59. 9% compared to dbSNP (b141). Taking CHARGE WES SNPs as gold standard, our WGS call set has 63. 9% sensitivity, 99. 8% specificity and 2. 98% FDR, outperforming any single caller. Therediscovery rate of singleton, doubleton and common SNPs (AF>5%) are 43. 1%,72. 2% and 95. 0%, respectively. We obtain high genotype concordance, 98. 7%(Ref/Ref), 84. 1%(Ref/Alt) and 99. 3%(Alt/Alt), compared to CHARGE cSNP array. By choosing HWE p-value cutoff 10-15 in AfAm and 10-17 in EuAm, we keep all rare SNPs in each population with AC<10 and further filter 0. 8% of SNPs (Ti/Tv=1. 22) after imputation purification, most of which have AF>10%. 92% of the filtered sites have excess of HETs. We hope future large scale WGS projects may benefit from our high quality variant calling methods and practices.

1831W

A Bayesian framework for *de novo* mutations calling in nuclear families. *Q. Wei[1,2], R. Chen[1,2], X. Zhan[3], X. Zhong[4], W. Chen[5], B. Li[1,2].* 1) Molecular Physiology and Biophysics, Vanderbilt University, Nashville, TN; 2) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 3) Quantitative Biomedical Research Center, University of Texas Southwestern Medical Center, Dallas, TX; 4) Department of Biostatistics, Vanderbilt University, Nashville, TN; 5) Department of Pediatrics, University of Pittsburgh, Pittsburgh, PA.

Increasing evidence shows that de novo mutations play an important role in genetic etiology of both complex and Mendelian diseases. Although a common design is to sequence proband-parents trios, sequencing additional unaffected siblings provides natural controls to characterize disease-causing de novo mutations identified in probands. Current state-of-the-art methods were designed to handle offspring-parents trios and for nuclear families with multiple offspring such methods are inefficient in de novo mutation calling due to the ignorance of the family relatedness. In this study, we develop a Bayesian framework to jointly model the sequencing data in a family, leveraging the identity-by-descent (IBD) allele sharing among offspring, to optimize the de novo mutation calling accuracy. Through extensive simulations we showed that the new method is superior to the existing state-of-the-art method, especially for moderate sequencing coverage. This will be particularly beneficial for whole genome sequencing, as it is still prohibitively expensive to perform high-depth whole genome sequencing. Breaking down by IBD sharing categories, we observed that the greatest accuracy was achieved for de novo mutations in regions of IBD=2, and for IBD=0 and 1 the performance is comparable. Application of the new method to an exome sequencing dataset of quartet nuclear families showed increased sensitivity and specificity of de novo mutation calling over existing methods.

1832T

Resizing N-Base Gaps in the Human Reference Genome. *Z. Dzakula, A. Hastie, A. Pang, E. T. Lam, W. Andrews, Z. Zhu, X. Zhou, T. Liang, T. Anantharaman, H. Cao.* BioNano Genomics, Inc. , San Diego, CA.

With our Irys® genome mapping platform and NanoChannel technology, we have collected whole genome mapping data on an ensemble comprising over 30 euploid human genomes. We then used the resulting *de novo* assembled genome maps to estimate the sizes of the N-base gaps in the human reference. The ultra-long DNA molecules (ranging from 150 kb to 2. 5 Mb) enabled us to bridge a large fraction of N-base gaps present in the human reference and to accurately measure their widths. The corrected gap widths were evaluated by averaging the corresponding segment lengths observed in the euploid assemblies. The arbitrary sizes of N-base gaps currently present in the human reference significantly differ from our ensemble-based estimates, in sharp contrast with the high degree of consistency among the *de novo* assembled genome maps originating from different individuals. Our assemblies also suggest existence of an internal structure within a subset of N-base gaps. Using genome mapping data to this subset of gaps, we were able to accurately place sequence motifs inside a portion of the human genome previously considered “The Dark Matter. ” In addition, we analyzed reference regions surrounding N-base gaps that disagree with all our euploid assemblies. We are proposing an alternative structure for those regions, consistent with our observations both in terms of the number and the locations of the relevant sequence motifs. Finally, our genome maps reveal diversity among the individuals included in the ensemble, suggesting presence of haplotypes both inside and in the vicinity of the N-base gaps.

1833F

Characterization of the microDNome and its impact on the differential sensitivity to cancer treatments. *P. Mehanna[1], V. Gagné[1], M. Lajoie[1], D. Sinnett[1,2], I. Brukner[3], M. Krajinovic[1,2].* 1) Centre Hospitalier Universitaire Sainte-Justine Research Center, University of Montreal, Montreal, Canada; 2) Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, Canada; 3) Molecular Diagnostics Laboratory, Jewish General Hospital, McGill University, Montreal, Canada.

A novel form of extrachromosomal circular DNA, microDNAs, has recently been identified within normal and tumor tissue. MicroDNAs are characterized as small fragments (200-400bp) of extra-chromosomal circular DNA and are derived from unique and non-repetitive genomic sequences that are enriched in exonic regions, CpG islands and 5'UTRs. These entities seem to be generated during DNA replication leaving behind microdeletions at different genomic loci that could modify a cell's transcriptomic profile when a regulatory region is altered, or result in a gain/loss of function of a particular gene if the coding region is modified. Furthermore, in the event of selective pressures, such as those induced by treatment in a cancer setting, these alterations could be fixed in the cell if they present a selective advantage and therefore contribute to cancer development and drug resistance. In this study, we set out to investigate whether and how the formation of microDNAs and concomitant chromosomal microdeletions can shape resistance to chemotherapeutic drugs in cancer, more specifically childhood acute lymphoblastic leukemia (cALL). To identify microDNA signatures we used a panel of 20 human lymphoblastoid cell lines (LCLs): 10 that were previously shown to be resistant (higher IC50) and 10 that were sensitive (lower IC50) to two major chemotherapeutic drugs used to treat cALL patients (methotrexate and asparaginase). Following treatment of LCLs with respective IC50 drug concentrations, microDNAs were extracted using methods that are adapted for the extraction of circular DNA molecules followed by enrichment through rolling circle amplification and high throughput sequencing using Ion Torrent technology. Reads were then mapped to the reference genome using Torrent Mapping Alignment Program. The identity of microDNA clusters was confirmed by validating their circularity (control of nucleotide repeats between soft-clips), their location in the genome (outside of repetitive sequences) and their size using third-party softwares such as BedTools and custom in-house scripts. We defined the genomic distribution of microDNAs and assessed their eventual drug-resistance specific enrichments. Genomic counterparts of microDNA were then prioritised for sequencing based on their function and redundancy. Ultimately, this study will provide insight into possible novel mechanisms of genomic variability underlying resistance in cancer chemotherapy.

1834W

Updates in the human reference genome assembly (GRCh38). *T. Rezaie[1], V. A. Schneider[1], T. Graves-Lindsay[2], P. Flicek[3], R. Durbin[4] for the Genome Reference Consortium.* 1) National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD; 2) The Genome Institute at Washington University, St. Louis, MO; 3) European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge, UK; 4) The Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK.

Continued improvement and modernization of the human reference genome assembly, together with a growing catalog of human genomic diversity, is promoting the advancement of biological research. The Genome Reference Consortium (GRC), the group responsible for updating the reference genome assemblies for human, mouse and zebrafish, provides users with improvements even between major assembly releases. These improvements are comprised of fix and novel patches, which represent assembly corrections and alternate sequence representations, respectively. Patches are standalone scaffold sequences placed in chromosome context via alignment. Released quarterly, there have now been 4 patch releases associated with GRCh38, including 38 assembly corrections and 19 new alternate sequence representations. We will demonstrate how our recent analysis of assembly-assembly alignments as part of an effort to integrate missing sequences into the reference resulted in representation of several new structural variants. We will discuss how the availability of several individual genome assemblies derived from long-read WGS sequencing efforts has enabled us to further improve the reference assembly. We will also describe how use of the NCBI Remapping Service contributed to these analyses and can serve as a tool for making assembly updates. We will discuss how this GRC effort not only continues the ongoing curation of the reference, but is also driving development of new GRC tools and protocols for assembly improvement. The GRC strives to make its curation activities accessible to the public and we will present recent improvements to public GRC websites. These include an interactive page for finding and viewing issues under GRC review. Users now have access to more information about issues, and improvements in navigation make it easier for users to pinpoint issues relevant to their research interest. Moreover, the GRC also provides tracks that describe GRC curation efforts and details of assembly construction and quality. We will show how to access these tracks on the GRC's website, NCBI's 1000 Genomes or Variation Viewer browsers, as well as via the GRC track hub used by the genome browsers at UCSC and Ensembl. You may contact the first author (during and after the meeting) at Tayebeh. Rezaie@nih. gov.

1835T

A Single-Tube NGS Library Prep Workflow Integrating Enzymatic Fragmentation Results in High Yields and Low Sequencing Bias. *S. Liu-Cordero[1], B. Miller[2], M. Appel[1], V. Van Kets[2], B. van Rooyen[2], H. Whitehorn[2], M. Ranik[2], P. Jones[2], A. Geldart[1], R. Kasinskas[1], E. van der Walt[2].* 1) Kapa Biosystems, Wilmington, MA; 2) Kapa Biosystems, Cape Town, South Africa.

The use of whole exome sequencing (WES) through next-generation sequencing is accelerating as clinical application of the technique becomes more prevalent. WES is increasingly being applied for elucidating the genetic etiology of human disease in research areas such as inherited disease, personalized genomics, and drug discovery. These types of projects show the need for faster and more scalable NGS library preparation methods while providing a level of sequence quality and coverage that allows for rare genetic variants to be revealed. Accurate determination of a comprehensive set of variants in coding regions requires deeper and more uniform sequencing coverage and a reduction of bias across all steps in the sequencing workflow. Fragmentation of sample DNA for input to the NGS workflow is often an essential, but laborious step at the beginning of the library preparation process. Current techniques involve either enzymatic methods that exhibit a high degree of sequence bias and sensitivity to DNA input amounts, or mechanical shearing systems that are not scalable to meet growing throughput demands. We describe a novel enzymatic fragmentation formulation that provides a streamlined workflow and dramatically reduces sequencing bias relative to current methods, approaching the performance of mechanical shearing but with improved workflow and automation. We compare this enzymatic fragmentation method to current fragmentation solutions in WES workflows for high and low quality DNA input, including FFPE. We show that, when combined with a highly efficient library preparation, this method provides a powerful tool for optimizing these workflows and increasing throughput.

1836F

TruSeq Rapid Exome: A new improved exome enrichment strategy using a mutant transposase. *D. Schlingman[1], A. Kia[1], A. Khanna[1], J. Burgess[1], H. Grunenwald[1], V. Ruotti[1], A. Tan[1], M. He[1], T. Osothprarop[1], M. Costello[2], A. Cheney[2], J. Decker[1], R. Vaidyanathan[1].* 1) Illumina Inc, Madison, WI; 2) Broad Institute, Cambridge, MA.

Rapid expansion of studies and knowledge on the human exome has led to a myriad of scientific discoveries in recent years. Transposase-based Nextera facilitated many of these breakthroughs by utilizing an innovative methodology that combines an efficient and robust workflow with impressive performance. Nextera enables users to fragment and adaptor-tag an entire genome in minutes, dramatically improving workflow speed vs. the traditional ligation-based strategies. The adaption of Nextera for exome enrichment applications has furthered the impact of such technology. In continuing to improve and optimize transposase-based strategies, we have re-engineered a highly efficient bacterial transposase with a lowered AT/GC insertional bias, leading to more uniform coverage, therefore, reducing the amount of sequencing required to cover the genome at the same depth when compared to the 1st generation Nextera. In addition, this new transposase delivers consistent library insert size over a range of input DNA, allowing up to a tenfold variance from the 50 ng input recommendation. Such capability for input tolerance translates to workflow flexibility, robustness and is independent of quantification inaccuracies. For a fast, streamlined exome enrichment workflow, we incorporated significant optimizations to reduce the hybridization time, enabling a 1-day workflow. In addition, we have introduced a modified buffer to reduce sample mutation due to oxidation. These combined improvements of a new mutant transpose and hybridization conditions demonstrate an impressive > 80% read enrichment at 4 GB sequencing and > 85% exome coverage at 10x. Variant calling analysis against the Genome-in-a-Bottle reference data set (National Institute of Standards and Technology), demonstrated exome workflow precision of >99. 5% and recall of >91% at ~100x coverage. Overall, this new and improved 1-day exome workflow with a re-engineered transposase demonstrates better on-target performance and more accurate SNP/variant calling, which in turn holds promise to further applications with exome target enrichment.

1837W

Missing coverage of ACMG genes in clinical exome and genome sequencing. *R. L. Goldfeder[1,2], D. M. Waggott[2], E. A. Ashley[2].* 1) Biomedical Informatics Program, Stanford University, Stanford, CA; 2) Stanford Center for Inherited Cardiovascular Disease, Stanford University, Stanford, CA.

The human genome is complex; pseudogenes and other repetitive sequences, which comprise roughly half of the genome, create challenges in sequencing, alignment, and variant interpretation. Despite these challenges, clinical-grade genome sequencing and interpretation requires accurate and complete genotype calls across the entire genome. Of note, the American College of Medical Genetics and Genomics (ACMG) recommends pathogenic variant discovery, review, and reporting for 56 genes that are known be associated with disease. Since we consider this gene set to be a minimum set of genes where performance must meet medical standards, we sought to examine the read depth for each base in each ACMG gene for 12 clinical exomes (including Personalis ACE Exome, Agilent Clinical Research Exome, and Baylor Clinical Exome) and 39 clinical genomes (including several versions of the Illumina HiSeq platform). We observe that coverage of these important genes varies by sequencing approach. In particular, exome sequencing methods perform better than whole genome sequencing for coding bases. However, UTR bases are better covered by whole genome sequencing approaches. Adequate read depth is required to make accurate genotype calls; therefore, insufficient coverage of any base in a disease gene is problematic as DNA sequencing becomes incorporated into clinical decision-making. Further understanding of the parameters that contribute to variability in coverage will aid the development of new technologies and analysis approaches for clinical genome sequencing.

1838T

***De novo* assembly of a diploid Asian genome.** *Y. Guo, K. Wang.* Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA.

The arrival of second-generation sequencing enables us to rapidly re-sequence human genome in a matter of hours, at reasonable prices. However, re-sequencing-based studies are performing poorly in GC-rich areas, segmental duplication regions, repeat regions with length more than read length; furthermore, reference-guided alignment process is inherently biased towards known sequences. With increasing throughput and read length from third-generation sequencing, e. g. SMRT (single-molecule real-time) technology, *de novo* assembly has become fast, accurate and affordable. Here we present results from *de novo* assembly of an Asian subject. We sequenced the genome with an average coverage of 80X by third-generation sequencing technology. The N50 of read length is above 12kb. We were able to generate a *de novo* assembly of ~2. 8 Gbp using Falcon assembler. We compared the genome assembly with several other assemblies generated on subjects of European and Asian ancestry, and evaluated the completeness of the assembly by several bioinformatics approaches. Next we used Illumina HiSeq data and Illumina Omni SNP array for detecting CNVs, by comparing to GRCh37. We then used the de novo assembly to see whether these CNVs can be found or validated by de novo assembly. Initial analysis revealed that long reads were able to capture large-scale CNV events, for example, a common 14,199 bp deletion. We will present the latest results on the genome assembly and discuss lessons learned from this study.

1839F

Initial characterisation of 125 whole-genome sequenced trios from four ethnic groups in the Gambia. *Q. S Le[1], EM Leffler[1], KJ Kivinen[2], GBJ Busby[1], Muminato Jallow[3], Fatou Sisay-Joof[3], Jim Stalker[3], Kirk Rockett[1], G Band[1], DP Kwiatkowski[1,2], CCA Spencer[1], GGDP The Gambia Genome Diversity Pr[1,2,3].* 1) Wellcome Trust Centre for Human Genetics, Oxford University, Roosevelt Drive, Oxford. OX3 7BN; 2) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge. CB10 1SA; 3) MRC Unit, The Gambia, Atlantic Boulevard, Fajara, The Gambia.

We describe the initial construction of a haplotype reference panel based on 125 trios from four major ethnic groups in Gambia sequenced to 10X coverage, alongside read data obtained from the LWK and YRI sample collections sequenced by the 1000 Genomes project. Our pipeline is similar to one used by the 1000 Genomes project and uses GATK, SNPTOOLS and Beagle to detect, filter, and refine variant calls, and SHAPEIT 2. 0 to phase incorporating pedigree information. Overall we call 23. 8 million SNPs and 1. 7 million INDELs across the autosomal chromosomes, including 7. 1 million SNPs that are not present in dbSNP, and therefore represent putatively novel human variation. Using new methods we also identify parts of the genome where new haplotypes are found, using variation at the malaria associated sickle-cell locus as an example. Exploiting the new haplotype reference panel, we imputed SNPs in a large cohort of severe malaria cases and population controls from the Gambia, and demonstrate improved imputation quality and power to detect association at known risk loci. We also use the data to look at the structure of genetic diversity within and between the four Gambian groups. We highlight regions of the genome which appear to be shaped by natural selection and harbour new functional variation.

1840W

Comparison of exome and genome sequencing platforms for the complete capture of protein coding regions. *S. H. Lelieveld[1], M. Spielmann[2,3], S. Mundlos[2,3], J. A. Veltman[1,4], C. Gilissen[1].* 1) Department of Human Genetics, Radboud Institute for Molecular Life Sciences, Nijmegen, Netherlands; 2) Institute for Medical Genetics and Human Genetics, Charité Universitàtsmedizin Berlin, Berlin, Germany; 3) Max Planck Institute for Molecular Genetics, Berlin, Germany; 4) Department of Clinical Genetics, Maastricht University Medical Centre, Maastricht, Netherlands.

For next-generation sequencing technologies sufficient base-pair coverage is the foremost requirement for the reliable detection of genomic variants. We investigated whether whole genome sequencing (WGS) platforms offer superior coverage of coding regions compared to whole exome sequencing (WES) platforms, and compared single-base coverage for a large set of different exome and genome samples (24 Agilent V4 (at 78x and 160x coverage), 12 Agilent V5 (100x), 12 NimbleGen V3 (95x), 24 Complete Genomics (44x and 87x), 11 Illumina HiSeq (28x), 12 Illumina X-Ten (40x)). We find that WES platforms have improved considerably in the last years, but at comparable sequencing depth, WGS outperforms WES in terms of covered coding regions. At higher sequencing depth (95x-160x) WES successfully captures 95% of the coding regions with a minimal coverage of 20x, compared to 98% for WGS at 87 fold coverage. A comparison to published gene panel studies shows that these perform similar to WES and WGS in terms of coverage. Three different assessments of sequence coverage bias showed consistent biases for WES but not for WGS. We found no clear differences for the technologies concerning their ability to achieve complete coverage of 2,759 clinically relevant genes. We show that WES performs comparable to WGS in terms of covered bases if sequenced at 2-3 times higher coverage. This does, however, go at the cost of substantially more sequencing biases in WES approaches, which may impact applications such as the identification of copy-number variants and somatic variation. Our findings will guide laboratories to make an informed decision on which sequencing platform and coverage to choose.

1841T

The International Genome Sample Resource: Beyond the 1000 Genomes Project. *X. Zheng-Bradley, L. Clark, J. Khobova, A. Datta, I. Streeter, D. Richardson, P. Flicek.* The European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom.

The 1000 Genomes Project provides an essential reference catalog of human variation. It includes more than 80 million variant sites ranging from single nucleotide polymorphisms to structural variant events including inversions and duplications. Global allele frequencies and genotypes for 2504 individuals are also provided; these samples are from 26 different populations across Europe, Africa, East and South Asia and the Americas. This resource has enabled many other projects to better interpret their results. Primary uses for the 1000 Genomes data sets include serving as imputation panels to derive whole genome variant sets from exome or array-based genotypes; as filters of shared variation in rare disease or cancer sequencing projects; and for exploring demography and selection in human populations and evolution. Although the 1000 Genomes Project has reached its goal and finished successfully, the use of the resulting data sets and analysis results remains high. The International Sample Genome Resource (IGSR), launched in January 2015, will maintain, update and expand this valuable reference data set. The IGSR will maintain the FTP site (ftp://ftp. 1000genomes. ebi. ac. uk/vol1/ftp) and the project website (http://www. 1000genomes. org) to ensure the community can access both the raw data and the documentation about the 1000 Genomes Project. We will establish a stable version of the 1000 Genomes Browser and the tools it hosts (http://browser. 1000genomes. org) based on the project's final date release, which is based on the GRCh37 assembly of the human reference genome. IGSR will also extend and expand the 1000 Genomes dataset to ensure it remains useful to the community. These plans fall into three main categories. 1) Remap the existing data to the new human assembly GRCh38. The majority of the phase 3 data is now available mapped to GRCh38 from our FTP site. 2) Draw in other data created on the Coriell Cell lines such as the RNA-Seq data created by the Geuvadis project, whole genome sequencing at a deeper level or from new technologies. 3) Add new open population collections sequenced using the same strategy to expand the diversity of the public catalog. We aim to build new tools to improve data discoverability and allow people to view the data alongside other genomic annotation in Ensembl.

1842F

***De Novo* Assembly of the First Human Diploid Genome with Single-molecule Sequencing for Asian Genome Project.** *M. Sohn[1,2,4], A. Rhie[1,2,4], SJ. Lee[1,3], JS. Kim[1,3], SK. Yoo[1,2,5], DS. Lee[1,2,4], J. Baek[3], M. Roh[3], JY. Shin[1,3], KS. Yang[3], C. Kim[3], JS. Seo[1,2,3,4,5].* 1) Genomic Medicine Institute(GMI), Seoul National University; 2) Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul 110-799, Korea; 3) Macrogen Inc. Seoul 153-781; 4) Department of Biochemistry and Molecular Biology, Seoul National University College of Medicine, Seoul 110-799, Korea; 5) Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 110-799, Korea.

Through the Asian Genome Project, we wish to uncover the genetic variation within the Northern Asian population that is responsible foradaptationand Mendelian diseases. The Asian Genome Project is divided into two phases. In the first phase, we have sequenced 158 Mongolians, 67Kazakhstans, 150 Koreans, 62Chinese's and Japanese fromHapMapsamples. In the second phase, we wishto sequence another ten thousand Asian genomes. It has come to our attention that we require a reference genome tailored for the Asian population. To pursue this goal, we assembled the Altai Korean (AK1) diploid genome usingPacBioRSII with P6C4 chemistry (>100X) and 100,000 BAC clones. The assembled-contigshad a total sequence length of 2. 811 Gb and had a N50 of 7. 3 Mb. Ourcontigscovered 91% of the GRCh38 reference genome with gaps, and furthermore, we were able to close 250 gaps and extend 180 gaps out of 604 gaps existing in GRCh38. We were able to find 196 novel insertions and 260 novel deletions in Chr20, which may explain certain geneticsusceptibilityin Asians. We will also be utilizing the data fromBioNanoto further refine our assembly and use BAC clones to resolve many highly complex regions, with the goal of creating the most accurate Asian reference genome for clinical application and population genetics.

1843W

SHIELD: an integrative gene expression database for inner ear research. *J. Shen[1,2], D. I. Scheffer[3,4], K. Y. Kwan[5], D. P. Corey[2,3,4].* 1) Department of Pathology, Brigham and Women's Hospital; 2) Harvard Medical School Center for Hereditary Deafness; 3) Department of Neurobiology; 4) Howard Hughes Medical Institute, Harvard Medical School, Boston, MA; 5) Department of Cell Biology and Neuroscience, Rutgers University, Piscataway, NJ.

The inner ear is a highly specialized mechanosensitive organ responsible for hearing and balance. Its small size and difficulty in harvesting sufficient tissue has hindered the progress of molecular studies. The protein components of mechanotransduction, the molecular biology of inner ear development, and the genetic causes of many hereditary hearing and balance disorders remain largely unknown. Inner-ear gene expression data will help illuminate each of these areas. For over a decade, our laboratories and others have generated extensive sets of gene expression data for different cell types in the inner ear using various sample preparation methods and high-throughput genome-wide approaches. To facilitate the study of genes in the inner ear by efficient presentation of the accumulated data and to foster collaboration among investigators, we have developed the Shared Harvard Inner Ear Laboratory Database (SHIELD: https://shield. hms. harvard. edu/), an integrated resource that seeks to compile, organize and analyze the genomic, transcriptomic, and proteomic knowledge of the inner ear. Five datasets are currently available. These datasets are combined in a relational database that integrates experimental data and annotations relevant to the inner ear. The SHIELD has a searchable web interface with two data retrieval options: viewing the gene pages online or downloading individual datasets as data tables. Each retrieved gene page shows the gene expression data and detailed gene information with hyperlinks to other online databases with up-to-date annotations. Downloadable data tables, for more convenient offline data analysis, are derived from publications and are current as of the time of publication. The SHIELD has made published and some unpublished data freely available to the public with the hope and expectation of accelerating discovery in the molecular biology of balance, hearing and deafness.

1844T

aRrayLasso: a network-based approach to microarrayinterconversion. *A. S. Brown[1], C. J. Patel[2].* 1) Cancer Immunology and AIDS, Dana-Farber Cancer Institute, Boston, MA; 2) Center for Biomedical Informatics, Harvard Medical School, Boston MA.

Summary: Robust conversion between microarray platforms is needed to leverage the wide variety of microarray expression studies thathave been conducted to date. Currently available conversion meth-ods rely on manufacturer annotations, which are often incomplete, oron direct alignment of probes from different platforms, which often failto yield acceptable gene-wise correlation. Here, we describe aRray-Lasso, which uses the Lasso-penalized generalized linear model tomodel the relationships between individual probes in different probesets. We have implemented aRrayLasso in a set of five open-sourceR functions that allow the user to acquire data from public sourcessuch as GEO, train a set of Lasso models on that data, and directlymap one microarray platform to another. aRrayLasso significantly pre-dicts expression levels with higher fidelity than technical replicates ofthe same RNA pool, demonstrating its utility in the integration of datasets from different platforms. **Availability and Implementation**: All functions are availa-ble, along with descriptions, at https://github. com/adam-sam-brown/aRrayLasso.

1845F

Sequencing File Mover: A Tool for the Management of Sequencing Data. *K. A. Duerr, M. Z. Mawhinney, K. N. Hetrick, S. M. L. Griffith, A. M. Sanchez, A. B. Robinson, B. D. Tibbils, B. D. Craig, J. L. Goldstein, L. Watkins, Jr. , K. F. Doheny.* Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality sequencing and genotyping services and statistical genetics consultation to investigators working to discover genes that contribute to disease. The Sequencing File Mover is a Java desktop application created to aid lab management in filtering, moving, copying, and deleting large quantities of files of diverse types. The application features a custom set of file manipulation modes that have been designed around the needs of sequencing lab managers. To match the dynamic needs of lab data interpreters, the application's performance parameters can easily be altered. Originally operating solely to copy or move a specific set of files from a selected directory, the application evolved to accept additional metadata, to be used during file management. For example, using pedigree information, the application can sort files by family. The application also contains multiple options to clean up vestigial sequencing data files and directories. Each option has been customized to delete residual files and transfer desired files to a compressed archive based on the type of data generated within a specific time period. Before the files are manipulated, the application gives counts of file type or directory size, specific to the mode selected, giving the user a clear view of what will happen upon submission. Before submission, the application will first check if the selected destination already contains any of the chosen files and will display a warning if needed. Upon submission, to gauge wait time, the application displays a progress bar, estimating the time remaining. If the user wishes to cancel the submission, the user has the option to revert any changes that have been made. The application will then either replace moved files or delete copied files. Previously, the amount of time taken to move or remove terabytes worth of data manually on a monthly basis was prohibitive and error-laden. With the help of this application, lab management can submit their mass file manipulation request in a few minutes, resulting in significant time savings and fewer errors.

1846W

A flexible pipeline that extracts full ClinVar data set. *X. Feng, A. Jackson, S. Paithankar, A. Baker, S. Plon, A. Milosavljevic.* Baylor College of Medicine, Houston, TX.

The ClinVar database, maintained by the National Center for Biotechnology information (NCBI), provides information about medically important variants with associated phenotypes and effects of sequence changes. With the help from the NIH Clinical Genome Resource (ClinGen) project, ClinVar is becoming an authoritative resource for medical variants submission and interpretation.

ClinVar's content can be accessed in several ways: the website for live queries and release file download for integrated data analysis. Users can also access it via NCBI's Eutils application programming interface (API). However, ClinVar's full content is only available in its Extensible Markup Language (XML) release. The XML release is structured using deeply nested nodes that requires memory-efficient parsing algorithms. Consequently many existing pipelines that extract data from Clinvar use the alternative Variant Call Format (VCF) version that has incomplete data but easier to parse. Moreover, the data reported by ClinVar can be organized in many ways. Consequently, much redundant effort is spent by multiple groups performing the same data warehousing effort in order to extract the parts of data of interest from the complete ClinVar release.

To address these issues and enable local warehousing of ClinVar data, here we report our design and implementation of a memory-efficient flexible pipeline that extracts full ClinVar dataset on a regular desktop with typical hardware configuration. We invented a user-editable map file format that specifies the portion of the ClinVar's XML file targeted for extraction. By editing this map file, the users can extract various portions of ClinVar without modifying the pipeline code. The pipeline outputs ClinVar variants as javaScript Object Notation (JSON) files for direct consumption by document oriented databases. We implemented the map processor in Ruby and designed an example map file that models the official example release file. Using this map, our flexible pipeline automatically extracted about 150,000 (as of May 2015) ClinVar variants from the XML release into JSONs. We then validated the pipeline using a series of different maps. The pipeline has been published on github (https://github. com/clingendb/clinvar_xml_pipe). The full set of ClinVar JSON files generated from the pipeline is hosted at the GenboreeKB repository(http://genboree. org/genboreeKB/genboree_kbs?project_id=test_clingen&coll=clinvar_xml0. 8) that can be accessed via APIs.

1847T

An integrated RNA and DNA approach to unravel genetic regulation in cancer. *V. Hedatale, S. Ramachandrula, P. Tata, R. Gupta.* Strand Life Sciences, Bangalore, India.

Background: Clinically meaningful applications like cancer studies benefit from analyzing genomic and transcriptomic anomalies simultaneously in personalized oncology. Mutations in the genome result in diverse transcriptional aberrations that can be missed in a whole genome/exome analysis. The analysis of low allele fraction variants, RNA editing events and allele specific expression is possible only in an integrated analysis. This necessitates the development of biologist friendly tools like Strand NGS to provide an integrated platform for easy and accurate analysis of multiple next generation sequence (ngs) data and perform multi-omic pathway analysis. **Results:** We used a publicly available tumor-normal paired genome and transcriptome ngs dataset from a patient with aggressive prostate cancer (APCa) rapidly progressing towards castrate resistance. A comprehensive analysis of the primary and metastatic prostate cancer was carried out using Strand NGS for possible small and large structural variants, expression alterations, RNA editing events and allele specific expression. Large amplifications with a concomittantly higher expression of genes like MSI2, KLK3, AR was observed in both primary and metastatic prostate cancers and corroborated with a TCGA provisional dataset of 499 aggressive prostate cancer samples. The hybrid luminal-neuroendocrine tumor had a number of copy number aberrations and gene fusion events and expressed both luminal markers (*MSI2, NKX3-1, AR, KLK3*) and neuroendocrine markers (*CHGA, CHGB, NSE, CGA)*. We identified multiple A-to-I RNA editing events within the 3' UTR region of H6PD. Based on known interactions we generated a Natural Language Processing-based network of androgen-independent prostate cancer and overlayed genomic and transcriptomic data in a multi-omic pathway analysis. Androgen receptor interacting proteins EIF4E, ERG, MAPK, AKT in this network were highly upregulated in the tumors suggesting that with disease progression, these proteins are active even in androgen-depleted conditions. **Conclusion:** Overall, this case study highlights the importance of integrated analyses of genome and transcriptome for basic tumour biology, RNA-editing events and identification of potential biomarkers. We also demonstrated that Strand NGS is an excellent platform for the effortless analysis, integration, visualization and downstream analysis of different ngs data.

1848F

The PhenX Toolkit: A Genomic Resource for Collaborative and Translational Biomedical Research. *W. Huggins[1], H. Pan[1], D. S. Nettles[1], T. Hendershot[1], D. R. Maiese[1], D. C. Brown[1], J. Levy[1], M. Phillips[1], N. Gaddis[1], V. Bakalov[1], Y. Qin[1], R. P. Troiano[2], J. Himes[3], P. Byers[4], N. Butte[5], S. Smith[6], P. A. Harris[7], M. L. Marazita[8], C. A. McCarty[9], G. K. Farber[10], K. L. Wanke[11], B. Iglesias[12], E. M. Ramos[12], C. M. Hamilton[1].* 1) RTI International, Research Triangle Park, NC; 2) National Cancer Institute, Bethesda, MD; 3) University of Minnesota, Minneapolis, MN; 4) University of Washington, Seattle, WA; 5) Baylor College, Houston, TX; 6) Florida Hospital, Orlando FL; 7) Vanderbilt University, Nashville, TN; 8) University of Pittsburgh, Pittsburgh, PA; 9) Essentia Institute of Rural Health, Duluth, MN; 10) National Institute of Mental Health, Bethesda, MD; 11) Office of Disease Prevention, National Institutes of Health, Bethesda, MD; 12) National Human Genome Research Institute, Bethesda, MD.

To help investigators identify opportunities for collaborative biomedical research and improve the consistency of data-collection, the Web-based PhenX Toolkit (consensus measures for **Phen**otypes and e**X**posures, https://www. phenxtoolkit. org/) is a catalog of 395 standard measures with browse and search capabilities and bioinformatics support. PhenX phase I established a consensus process and a bioinformatics pipeline, and generated 349 measures relevant to common complex diseases with additional depth in substance abuse and addiction. PhenX measures and protocols are included in the Cancer Data Standards Repository (caDSR) Common Data Elements (CDE) Dictionary and the NIH Common Data Element (CDE) Resource Portal. PhenX phase II, now underway, is assembling Expert Review Panels (ERPs) to review and update Toolkit content. The ERPs are evaluating the protocols and supporting information for all PhenX Phase I measures. The first ERP reviewed 74 protocols from 4 related research domains. As a result of their review, the ERP recommended replacing 4 protocols and adding 3 new measures and 6 new protocols to the PhenX Toolkit. In Phase II, measures suitable for inclusion in the Toolkit extend beyond primary utility in genome wide association studies to other study designs, including clinical and translational research. The PhenX Steering Committee prioritized expansion of the Toolkit to include measures relevant to rare genetic conditions, including a crowdsourcing annotation effort, and to address obesity. To support investigators who want to collect data via the Web, PhenX protocols are being made available as REDCap Instrument Zip files. To help users find comparable data, a team of curators is combining programmatic and manual approaches to map the ~16,000 PhenX variables to all completed studies in dbGaP. Variable mappings for selected PhenX measures will be presented to demonstrate utility. In supplemental efforts new measures are being added to support mental health research, tobacco regulatory research, and sickle cell disease research. PhenX is managing change, expanding the scope, and increasing the utility of the Toolkit. These efforts will ensure that the PhenX Toolkit will continue to provide the biomedical research community with easy access to standard measures with the potential to increase the overall impact of individual studies by facilitating cross-study analysis. *Funding provided by NHGRI, co-funding from NIDA: U01 HG004597 and U41 HG007050.* .

1849W

Interpreting a migraine GWAS using gene expression in healthy human brain. *S. M. H. Huisman[1,2], E. Eising[3], A. Mahfouz[1,2], L. Vijfhuizen[3], B. de Vries[3], B. P. F. Lelieveldt[2], A. M. J. M. van den Maagdenberg[3,4], M. J. T. Reinders[1], International Headache Genetics Consortium.* 1) Department of Intelligent Systems, Delft University of Technology, Delft, Zuid Holland, The Netherlands; 2) Department of Radiology, Leiden University Medical Center, Leiden, Zuid Holland, The Netherlands; 3) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 4) Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands.

Genome Wide Association Studies (GWAS) are frequently used to find common variants with small effect sizes. However, they often provide researchers with short lists of single nucleotide polymorphisms (SNPs) with uncertain connections to biological functions. We present an analysis of GWAS data for migraine, where the full list of SNP statistics is used to find groups of functionally related migraine-associated genes. For this end we make use of gene co-expression in the healthy human brain. We performed genome wide clustering of genes, followed by enrichment analysis for migraine candidate genes. In addition, we constructed local co-expression networks around high-confidence genes. Both approaches converge on distinct biological functions and brain regions of interest. Materials & MethodsMigraine GWAS data was obtained from the International Headache Genetics Consortium, with 23,285 cases and 95,425 controls. Genes were scored by SNP load and divided into high-confidence genes, migraine candidate genes, and non-migraine genes. Spatial gene expression data in the healthy adult human brain was obtained from the Allen Brain Institute. It contains microarray expression values of 3702 samples from 6 donors. Robust gene co-expressions were used to cluster genes into 18 modules, which were then tested for enrichment of migraine candidate genes, and functionally characterized. In a second approach, local co-expression networks were built around the high-confidence migraine genes. These local networks were then compared to the modules of the first approach. ResultsThe genome wide analysis revealed several modules of genes enriched in migraine candidates. Two modules have preferential expression in the cerebral cortex and are enriched in synapse related annotations and neuron specific genes. A third module contains oligodendrocytes and genes preferentially expressed in subcortical regions. The local co-expression networks converge on the same pathways and expression patterns, even though the high confidence genes lie mostly outside of the modules of interest, which improves our confidence. DiscussionThe analyses confirm the previously observed link between migraine and cortical neurotransmission. They also point to the involvement of subcortical myelination, which is in line with recent tentative findings. These results show that more relevant information can be extracted from GWAS results, using (publicly available) tissue specific expression patterns.

1850T

Deploying a next-generation informatics infrastructure for genomic medicine: the Ohio State University and Ohio Supercomputer Center experience. *D. D. Kinnamon[1], D. Johnson[2], F. Oriyo[1], R. E. Hershberger[1].* 1) Division of Human Genetics, Department of Internal Medicine, The Ohio State University Wexner Medical Center, Columbus, OH; 2) Ohio Supercomputer Center, Columbus, OH.

Genomic medicine requires informatics infrastructure that facilitates integrated storage and high-throughput processing of phenotypic, laboratory, and genomic data. Such infrastructure presents a major implementation challenge; genomic data storage and processing require scalable high-performance computing resources not generally available within an academic medical center. Meeting this challenge by deploying these resources at the medical center is often fiscally and operationally untenable. While this challenge could also be met by moving all data management from medical center systems to an external cloud or shared high-performance computing resource, such systems may not fully comply with federal, state, and institutional regulations pertaining to confidential health information. To overcome these obstacles, the Division of Human Genetics at the Ohio State University Wexner Medical Center (OSUWMC) partnered with the Ohio Supercomputer Center (OSC), a shared high-performance computing resource funded by the State of Ohio, to design and deploy a comprehensive solution using the BC|Enterprise genomic data integration software platform from vendor BC Platforms. The platform is deployed on a dedicated server hosted at OSC and provides tools for managing and linking phenotypic, laboratory, and genomic data. These tools transparently leverage OSC's shared storage pool and compute clusters for high-volume storage and high-throughput processing of genomic data at low cost. In addition, the project team designing and implementing the platform worked closely with OSUWMC Information Security and the Ohio State University Office of Responsible Research Practices throughout all project phases to ensure full compliance with all applicable regulations, including for system components using shared OSC resources. The resulting platform provides end-to-end data management and analysis capabilities for a single patient, minimizes movement of genomic data for analysis, and provides scalable storage and processing to accommodate anticipated growth in genomic data. End-users at OSUWMC are able to access these features through web applications that do not require local installation of specialized software or knowledge of high-performance computing. Our platform provides a template for academic medical centers seeking to partner with shared high-performance computing resources to establish cost-effective software-as-a-service solutions that support genomic medicine.

1851F

Sequence and structural variability at the whole population scale for the Influenza Virus type A. *G. Mazzocco[1,3], P. Migdał[1], M. Łaś;niewski[1], J. Radomski[2], D. Plewczyński[1].* 1) Centre of New Technology, University of Warsaw, Warsaw, Poland; 2) Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Warsaw, Poland; 3) Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland.

Influenza type A viruses are commonly known for their extreme genetic variability, which represent a true virological challenge. The Hemagglutinin (HA) protein plays a crucial role in the mechanism of action of the virus, defining the likelihood of the interaction between the host cells (avian, human, swine, etc.) and the infecting viruses. The specific nature of the binding site within a given HA protein is in fact responsible for the differential recognition of some host-specific sialic acid derivatives. Hence the analysis of HA variability at molecular level is crucial for the correct understanding of the influenza problem at the population level. Although the analysis of HA genetic variability is classically applied to sequence and serological data, a comparative structural procedure can potentially provide additional insight. In order to perform such analysis, we initially selected a pool of 1198 HAs expressing the genetic variability of ~22k different HAs. We compared these proteins in terms of sequences, protein structures and electrostatics potentials (ESP), generating three layers of comparable information that were after used to generate three graph-based relational representations respectively. Metadata information (e. g. year, geographic location, virus host, etc.) was included, allowing the exploration of possibly interesting evolutionary patters.

1852W

Integrating coding variants and regulatory variation to improve the power of identifying disease-associated genes in complex human diseases. *A. Mezlini[1,2], A. Shlien[2], A. Goldenberg[1,2].* 1) Computer Science, University of Toronto, Toronto, Ontario, Canada; 2) SickKids Hospital. Toronto. Ontario. Canada.

Identifying genes associated with complex human diseasesis one of the main challenges of human genetics and computational medicine. Adisease causing abnormalityin a proteincan be due to protein-coding SNVsor due to regulatory changes, such asstructural DNA aberrationsor epigenetic variation,leading to theabnormalitiesin the quantity of the produced protein. The same disease in different patients may be due todifferent types of abnormalities. In complex diseases,the problem is compounded since these different types of abnormalities are likelyaffectingmultiple proteins making the discovery of the underlying disease mechanism very challenging.

To capture various types of abnormalities and to increase the power of the disease mechanism detection, we propose a novel method that integrates different types of data such asexomesequencing, gene expression and/or DNA methylation. Our approach utilizes a probabilistic graphical model to infer genes associated with complex diseases. Our method also incorporates external biological knowledge including gene networks and variant harmfulness predictions as Bayesian priors to improve the accuracy of gene detection.

Our extensive simulations show that our method has far superior sensitivity and precision compared to variant-aggregating (CAST, C-Alpha, SKAT, SKAT-O,dmGWAS) and differential expression methods. We show that integratingmultipledata typesand biological annotationshasagreaterpositive impact on the power of ourmethodthanthe sum of individualcontributions ofeach source of information, indicating the effectiveness of thesynergy of differenttypes of data in our model.

Our integrative approach was able to identify important genes in breast cancer, recovering the genes that were previously associated with different breast cancer subtypes and identifying novel candidate genes. Importantly, our analysis shows that many associated genes had coding aberrations in some patients and regulatory abnormalities in others, emphasizing the importance of data integration to explain the disease in a larger number of patients. For every patient in our cohort, we were able to analyze the posterior distribution overassociated genes, which enabledus tofurthercharacterize patientheterogeneity within the breast cancer subtypes. .

1853T

Integrated database and knowledge base for genomic prospective cohort study in Tohoku Medical Megabank toward personalized prevention and medicine. *S. Ogishima, T. Takai, K. Shimokawa, S. Nagaie, H. Tanaka, J. Nakaya.* Dept. of Bioclinical Information, Tohoku Medical Megabank Organization, Tohoku University, Sendai-shi, Miyagi, Japan.

To revitalize medical care and realize personalized prevention and medicine in the disaster area of Great East Japan Earthquake, the Tohoku Medical Megabank Organization has been established aiming at becoming a center for the reconstruction of the entire Tohoku region by conducting prospective genome-cohort study and developing the large-scale genome biobank toward personalized prevention and medicine. In our prospective cohort study, we will recruit 150,000 people at Tohoku university, satellites, health clinics, and Iwate medical university, and have collected biospecimen (blood, urine), blood and urine test, questionnaire, physical measurement as baseline and follow-up investigation. As for pathogenesis investigation follow-up, we will collect clinical data as electronic health records provided by Miyagi Medical and Welfare Information Network (MMWIN). MMWIN is expected to develop regional electronic network of medical records of most hospitals in Miyagi prefecture. Collected data will be de-identified and stored in the integrated database of Tohoku Medical Megabank. Researchers in universities, institutes, and companies can apply withdrawal of biospecimen and data. Integrated database is a database for integration of genomic (omics), specimen, baseline & follow-up and clinical data. On the other hand, knowledge base is a database for integration of existing and novel knowledge by large-scale correlation analysis on integrated database. It is a RDF store implemented by Virtuoso. Our database is is too huge to to grasp characteristic of data. Therefore, integrated database not only provides integration of data but also provides statistical significance and related knowledge in cooperation with knowledge base. Statistical significance of selected data is provided in integrated database to characterize selected cohort; e. g. , selected population has statistically significant bias in sex and alcohol intake. This characterization is expected to lead to hypothesis. Related knowledge of selected data is also provided in integrated database in cooperated with knowledge base; e. g. , breast cancer is known to have a highly linked and causative SNPs. This provision of related knowledge is expected to accelerate studies. We hope our integrated database and knowledge base which will be key database for realizing personalized prevention and medicine.

1854F

Phoenix Web: Presenting Relevant Lab Data and Receiving Feedback from Outside Sources. *A. M. Sanchez[1], K. A. Duerr[1], D. R. Leary[2], J. D. Newcomer[3], M. Adams[1], T. Shelford[1], C. M. Ongaco[1], I. A. McMullen[1], M. Z. Mawhinney[1], J. Romm[1], S. M. L. Griffith[1], J. L. Goldstein[1], L. Watkins, Jr. [1], K. F. Doheny[1].* 1) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 2) The New York Genome Center, New York, NY; 3) The Johns Hopkins Applied Physics Lab, Laurel, MD.

The Center for Inherited Disease Research (CIDR) provides high quality sequencing and genotyping services and statistical genetics consultation to investigators working to discover genes that contribute to disease. Phoenix Web is an application that allows investigators secure and indirect access to portions of CIDR's Phoenix Laboratory Information Management System (LIMS). The Phoenix software, from which Phoenix Web extends, was designed and built at CIDR to organize project data; and orchestrate and track the flow of a project's samples as they move through CIDR's labs. While Phoenix was originally designed solely for internal use within CIDR, two significant portions of Phoenix's workflow, Principal Investigator File Handling (PIFH) and Problem Handling, require direct input from investigators remote from CIDR. The PIFH workflow handles processing and storage of investigator-provided project files (typically spreadsheets). The Problem Handling workflow is composed of two main sections: the tracking of issues discovered at CIDR that pertain to an investigator's project, and the generation of problem reports to be presented to an investigator for resolution. Such problems include, for example, insufficient quantities of DNA for a given assay or contaminated materials. Because of the timepotentially spent waiting for investigator feedback, both PIFH and Problem Handling cancreate workflow bottlenecks. Additionally, because these workflows require direct interaction with the investigator, easy-to-use software wouldyieldmorepositive investigator relations. Previously, the information exchanged with investigators was done exclusively via email communication with spreadsheet attachments. As a result, the entire process was both susceptible to human error and time consuming. Phoenix Web improves on this process by automatically and securely presenting relevant and up-to-date data to the investigator directly from the Phoenix system as well as providing automatic feedback from the web interface. This eliminates the need for slow email exchanges while still providing the investigator with feedback on their uploaded project files and problem resolution responses in an easy to access way. Furthermore, Phoenix Web presents more data to CIDR's investigators than before, allowing for a more transparent experience. Phoenix Web is flexible enough to allow even further transparency to investigators in the future.

1855W

Integrating Big Data to Make Biological Sense of Statistical Models. *S. Verma[1], A. Frase[1], S. Pendergrass[2], M. Hall[1], D. Kim[1], R. Li[1], M. Ritchie[1,2].* 1) Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA; 2) Biomedical and Translational Informatics - Geisinger Health System, Danville, PA.

Researchers in human genetics have spent the last several decades searching for the genetic factors that influence susceptibility for complex disease, which has resulted in an outburst of biological data. Millions of research participants have been analyzed to identify thousands of single genetic variants associated with hundreds of common diseases. The search for complex gene-gene interactions and pathway/polygenic models has led to an even greater number of genetic variants in our pool of "putative disease variants". Making sense of all of these statistical associations in terms of underlying biology and disease etiology is the main objective for the transition from basic science in human genetics to translational science in precision medicine. Thousands of publicly available databases exist with biological knowledge, dozens of bioinformatics tools for functional predictions, and petabytes of functional genomics data from projects like ENCODE, GTeX, Roadmap Epigenome and others which can be integrated with these genetic variants to begin to put the pieces of the puzzle together. These databases can be used to prioritize results from association studies for functional validations in model organisms or they can also be used to prioritize variants/genes/omic regions that can be tested further to investigate interactions or regulatory networks associated with diseases. We have developed a unified framework, which is an integration of multiple public databases along with relevant annotation and visualization tools (Biofilter and iPhenoGram), allowing us to create a complete workflow to perform higher-level biological interpretation of association results. Recently we have extended this workflow by incorporating additional data on chromatin states from ENCODE. Using this workflow for specific genes (e. g. *PCSK9*), we can identify how variants associated with disease traits can be interpreted based on the activity and expression of each region. With this pipeline, we have improved the interpretation of results from novel and previously associated variants in glaucoma, type 2 diabetes, age-related macular degeneration and hypothyroidism by annotating them based on activity map of regions of genome (in most cases represented as repressed or transcribed regions). In conclusion, we present data visualization workflows to facilitate the integration of big data from biology with genetic association results to elevate ongoing human genetics studies to the next level.

1856T

RAPiD - An Agile and Dependable RNA-Seq Framework. *Y. Wang, R. Z. Castellanos, C. Pandya, H. Shah.* Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

RNA-Seq is a widely used technology for transcriptome analysis and rapid software development is underway for better analysis tools and algorithms. However, among numerous tools available in the literature, choosing tools of interest and integrating them as a seamless pipeline requires significant effort. There is an unmet need for a software suite that covers the breadth of available RNA-Seq algorithms. We present RNA-Seq Advanced Pipeline Deployment (RAPiD), an agile and dependable RNA-Seq framework that automates alignment (STAR and TopHat), quality control (FastQC, HTSeq and RNA-SeQC), expression analysis (featureCounts, Cufflinks, Kallisto and RSEM), variant calling (GATK), alternative splicing analysis (MISO) and fusion/chimera discovery (deFuse, Chimera, ChimeraScan, FusionCatcher, FusionMap, InFusion, OncoFuse and TopHat-Fusion). RAPiD is easily configurable with a recipe file that defines the set of tools/algorithms, corresponding parameters and dependency management. Integration of RAPiD with a high-performance computing cluster enables high-throughput computations with pipeline monitoring and manipulation capabilities. RAPiD runs are automatically tracked and visualized via a web-based dashboard through our in-house workflow management platform - APOLLO. RAPiD has been used for quality control, expression analysis, variant calling and fusion discovery. In total, RAPiD has processed tens of thousands of samples for a diverse set of clinical and research studies including The Cancer Genome Atlas (TCGA), Genotype-Tissue Expression (GTEX), CommonMind, PsychENCODE and LINCS. .

1857F

Piloting Methods for Integrating Neuroimaging and GWAS Data in Bipolar Disorder Studies. *H. Cao, MA. Brotman, N. Akula, W. Guo, L. Donahue, A. Oakes, B. Lehrman, FJ. McMahon, E. Leibenluft, Y. Yao.* National Institute of Mental Health, NIH, Bethesda, MD.

Background: Several functional magnetic resonance imaging (fMRI) studies have suggested that patients with bipolar disorder (BD) and youth at familial risk present amygdala dysfunction. Additionally, one genome-wide association study (Liu et al. , 2010) nominated a few genes influencing the magnitude of amygdala activation during face processing in BD patients. Our ultimate goal is to test whether there is a high level of correlation between BD-targeted neuroimaging phenotypes and genotypes. Toward that end, we compared the performance of two methods, a correlation method and a spares representation based variable selection (SRVS) method, in terms of the capability of finding the SNPs highly correlated with the fMRI phenotypes. **Methods:** Both task fMRI data (emotional/neutral faces processing task) and SNP data (712,077 SNPs after quality control) were acquired from 73 subjects, including 48 healthy controls, 15 adult bipolar patients and 10 children at risk for BD. Traditionally in case/control studies, the phenotype is quantitatively expressed as to represent disease status; however, here we used the averaged amygdala fMRI scores from both hemispheres for each subject as the quantitative phenotype in the form of . To detect significant association between the phenotype and the genotype data, we employed two methods including SRVS method and a SNP-wise -SNP-correlation analysis. The recently developed SRVS algorithm (Cao et al. , 2014) is proposed to solve sparse linear systems and select significant variable sets associated with the phenotype y. **Results:** In addition to the genes previously reported as BD candidate genes (e. g. AKAP13, DAOA, HTR2A, NRG1, SGK1), both statistical approaches nominated the MTR gene which has not been previously reported. Furthermore, the potentially significant SNP-fMRI phenotype association signals detected by both methods were different (overlap less than 25%). For example, CAMTA1 was only detected by the SRVS method; whereas, DISC1 and ODZ4 were identified by the correlation analysis. **Conclusions:** This is one of the first efforts to integrate fMRI and GWAS data with the aim of exploring the underlying mechanisms of BD using two statistical approaches. This preliminary analysis indicates the feasibility of using both methods for integrating GWAS and fMRI data. However, depending on the data structure, the performance of these two methods may be different. More studies are needed.

1858W

Accessing genomic evidence for clinical variants with new NCBI services. *S. Sherry, C. Xiao, D. Slotta, K. Rodarmer, M. Feolo, M. Kimelman, G. Godynskiy, C. O'Sullivan, E. Yaschenko.* National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD.

NCBI provides many resources for evaluating and declaring evidence of pathogenicity for an increasing number of human sequence variants. Primary details about a genetic test's analytical validity and clinical utility are reported in the Genetic Test Repository, and evidence for inferences of variant pathogenicity are summarized in ClinVar. NCBI has developed several tools that integrate these high level records with the more basic factual data describing them as population level variants (dbSNP/dbVar), and deeper still as individual-level observations with called genotypes (the NCBI Genotype Server), phenotypes (dbGaP), whole genome/exome sequences (SRA) and finally as positions on the reference genome (GRC human reference assemblies). Traversing the connections between variant-level records, e. g. GTR, ClinVar and dbSNP, and individual-level data (e. g. genotypes, sequnces, samples, and phenotypes) is a computationally intensive activity, and NCBI has developed several new services to pre-compute these relationships and permit users to quickly move from summary records to individual level data. This presentation will introduce several of these services including variation reporter, variation viewer, beacon search, and the genome browser with particular emphasis on how users can access and review individual level data for clinical variants or *ad hoc* genomic positions of particular interest. Examples of use include the review of ClinVar submissions by expert panels, research into the existence of specific sequence alleles, automatic notification when new data for specific potential alleles of interest are submitted to NCBI, confirmation of variant properties during manuscript review, and research in general questions of human genetic architecture.

1859T

A visual semantic similarity guided approach to variant prioritization and discovery in genome-wide diagnostics. *R. James[1], E. Chen[2], I. Campbell[2], P. Boone[2], M. Rao[2], M. Bainbridge[2,3], J. Posey[2], Y. Yang[2,4], C. Eng[2,4], C. Shaw[2,5].* 1) Structural and Computational Biology and Molecular Biophysics Program, Baylor College of Medicine, Houston, TX; 2) Molecular and Human Genetics Dept, Baylor College of Medicine, Houston, TX; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Baylor Miraca Genetics Laboratories, Houston, TX; 5) Statistics Dept, Rice University, Houston, TX.

Genome-wide data are increasingly important in the clinical evaluation of human disease. However, the large number of variants observed in individual patients challenges the efficiency and accuracy of diagnostic review. Recent work has shown that systematic integration of clinical phenotype data with genotype information can improve diagnostic workflows and prioritization of variants. We have developed visually interactive, analytically transparent analysis software that leverages existing disease catalogs, such as the Online Mendelian Inheritance in Man database (OMIM) and the Human Phenotype Ontology (HPO), to integrate patient phenotype and variant data into ranked diagnostic alternatives. Our tool, "OMIM Explorer" (http://www. omimexplorer. com), extends the biomedical application of semantic similarity methods beyond those reported in previous studies. The visual approach collapses high-dimensional phenotypic and genotypic feature data of an individual into a graphical format that contextualizes the patient within a low-dimensional disease map. This visual approach indicates differential diagnoses and algorithmically suggests potential alternatives for phenotype and gene queries. The tool also implements an adaptive approach for gene discovery based on patient phenotypes. We retrospectively analyzed pilot cohort data from the Whole Genome Laboratory at Baylor College of Medicine, demonstrating performance of the tool and workflow in novel visual display of clinical exomes, rapid matching of cohort cases to known diseases, and accurate prioritization of causative variants. Our tool assigned to clinically reported causal variants in disease genes a median rank in the top 1% of filtered candidates (4/463) across the 41 cohort cases with reported molecular diagnoses of exome variants in OMIM Morbid Map genes. Additionally, a novel semantic similarity method we developed asigned a median rank of 3/463, with 90% within the top 10,to the reported cohort variant host genes. This integrative paradigm can improve the quality of genomic medicine by more effectively utilizing available phenotype information, catalog data, and genomic knowledge.

1860F

Tracking data provenance at the ENCODE DCC. *E. T. Chan[1], J. M. Davidson[1], T. R. Dreszer[1], B. C. Hitz[1], M. Ho[1], B. J. Lee[2], V. S. Malladi[1], N. R. Podduturi[1], L. D. Rowe[1], M. Simison[1], C. A. Sloan[1], J. S. Strattan[1], F. Tanaka[1], E. L. Hong[1], W. J. Kent[2], J. M. Cherry[1].* 1) Department of Genetics, School of Medicine, Stanford University, Stanford, CA; 2) Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA.

The provenance of experimental reagents and transparency of computational analyses are essential to compare, reproduce, and interpret experimental data. The task of tracking this information consistently across diverse sequencing assays can be especially challenging in large projects like the ENCODE (ENcyclopedia Of DNA Elements) Consortium that perform 40+ genomic assays using 400+ cell and tissue types. The identification of a transcription factor binding site or the quantification of transcript's expression level is dependent on the software versions, and the parameters used when running the software, as well as the library preparation methods, and how the biological samples were obtained. To capture the provenance of experimental methods and computational results, the ENCODE DCC (Data Coordination Center) has created a rich data model that represents how experiments were performed, what software and pipelines were used, and which files were analyzed. These details of the experimental and computational methods, known as metadata, can then be used to identify related data for further analysis, interpret the results of the assays, and allow reproducibility of pipelines. All metadata and data generated by the ENCODE Consortium are freely available at at the ENCODE Portal(https://www. encodeproject. org/).

1861W

Facilitating Data Sharing through a Cloud-Based Data Platform. *M. Kaganovich, D. Gross, P. George, X. Xu, J. Cohen, L. Cherney, J. Hull, D. Caplan.* SolveBio, New York, NY.

High throughput generation of genomics data has led to an explosion in the breadth and quantity of reference data generated. Data sharing, theoretically, should be easy, rewarding for the data producer, and pervasive. Instead, the lack of systems and technologies to index, standardize, and share data, and the lack of interoperability and machine-readable standards means that critically important data remains in "silos", or walled off data repositories. Most individuals and institutions have little incentive to share data when sharing costs significant amounts of time and computing resources. Data sharing is crucial for genomics research and clinical genetics, as a source of reference data, for transparency, efficiency, and research reproducibility.

We have created a cloud-based data platform that indexes genomics-related reference data and a data integration pipeline that efficiently and systematically integrates different file types and sources. As of June 2015, our Data Library has integrated 25 publicly available and 4 commercially licensed data sources with 150+ independent data sets (consisting of over 1. 2 billion independent records and 28 billion data points) into one single platform. All data on our platform is versioned and dated, fully documented in terms of content and provenance, harmonized across the platform for common identifiers such as gene symbol and genomic position, and made accessible via a browser interface and an Application Programming Interface (API). The purpose of creating this data platform is to facilitate data sharing within and between organizations to advance genomics research and the practice of clinical genetics.

Here, we present our best practices, current models, and technical and logistical challenges involved in genomics data integration and data sharing. We present and contrast specific and real use cases of intra-institutional sharing, inter-institutional/intra-consortium sharing, variations on commercial licenses, and freely publicly available data sharing models. We also provide a calculation of the expenses (human hours and cloud-computing costs) of data integration and continued maintenance. We will show that our efforts improve the efficiency of data sharing by lowering the technical, logistical, and legal barriers associated with it.

1862T

A new publicly accessible resource for comprehensive recessive-disease variant curation: the GenePeeks Research Browser (GPRB). *R. M. Lim[1], M. J. Silver[1], C. Borroto[1], M. P. Hearing[1], A. J. Silver[1,2], J. L. Larson[2], L. M. Silver[1,2,3,4].* 1) GenePeeks, Inc. , New York, NY, USA; 2) GenePeeks, Inc. , Cambridge, MA, USA; 3) Department of Molecular Biology, Princeton University, Princeton, NJ, USA; 4) Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ, USA.

Purpose: The science and human genetics communities have developed publicly available tools such as the UCSC Genome Browser, ClinVar, OMIM, and GeneReviews to further the understanding of the human genome. GenePeeks introduces the first publicly accessible browser that introduces two new features in addition to incorporating links to these resources: 1) Variant-specific Gene Dysfunction (VGD) scoring and 2) a dynamic graph displaying variants based on VGD score and Exome Aggregation Consortium (ExAC) allele frequency.

Methods: Gene-specific information was gathered from OMIM and RefSeq. Analytic gene targets were principally defined by Illumina's TruSight Inherited Disease Gene List. Variant-specific clinical information was gathered from ClinVar and OMIM. Population-specific allele and genotype frequencies were derived from 60,706 exomes processed and made publicly available by ExAC. Every variant described in ClinVar or present in the ExAC dataset within the targeted intervals defined by Illumina's TruSight One Sequencing Panel is analyzed and displayed. VGD scoring is described elsewhere (Silver, M. et al. , 2015 ASHG Abstract Submission). Updates to the GPRB are made continuously as new releases from data sources and interface improvements arise.

Results: The GPRB employs a user-friendly search tool that can respond to gene- and variant-specific searches. Gene pages host a dynamic display of variants graphed by their VGD score (x-axis) and ExAC allele frequency (y-axis, logarithmic scale). All variants are listed below the graph with brief variant information. Filters can be added to a query, thus applied to both the graph and the list. Each displayed variant links to a comprehensive variant page. Variant pages include general information such as genomic locus, ClinVar accession ID, OMIM allele ID, HGVSp and HGVSc; VGD information such as final VGD score and pathogenic predictor inputs; frequency information such as ExAC regional allele and genotype frequency; and clinical information such as ClinVar clinical significance and OMIM disease ID association. Sourced IDs such as ClinVar accession ID are hyperlinked to their respective pages on external sites; locus information links to the corresponding region in the UCSC Genome Browser. The GPRB is available at research. genepeeks. com.

Conclusion: The GPRB delivers novel functionality and content in a publicly available resource that is designed to adapt and improve based on the needs and feedback of users.

1863F

Bitmap indexing of genotyes improves functionality of an in-house distributed data query system. *H. Qiu[1], H. HAKONARSON[1,2].* 1) Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Bitmap indexing has been proven very efficient for query of static, low cardinality columns in data warehouses. Recent studies by other research groups applied bitmap index to large scale genomic variant datasets. Analyst Portal is our in-house distributed and integrative data query system where users, through web interface, perform complex query on Electronic Medical Records (EMR), laboratory workflow information, genotyping and next generation sequencing data. In order to support query into actual genotype calls of high-density SNP array, we first build bitmap indexes from existing source genotype files in plink PED format. Then the bitmap indexes are compressed using CPU friendly word-aligned hybrid (WAH) encoding scheme. Query of genotypes are performed using bitwise operations on the compressed bitmap indexes. The bitmap indexing package is deployed as a web service and made available to the relational database management system (RDBMS) on which Analyst Portal runs. Analyst Portal composes a query string based on user's search terms and makes API calls through the bitmap index web services. A user search term like "only return subjects with a list of known SNP mutations associated to a disease" can be effectively accommodated as a part of a more complex query. We conclude that WAH compressed bitmap indexing adds important genotype query functionality to our existing data query system.

1864W

Extending data access at the EGA. *D. Spalding[1], J. Almeida-King[1], C. Y. Gonzalez[1], J. Kandasamy[1], G. Saunders[1], A. Senf[1], I. Lappalainen[1], S. Ur-Rehman[1], M. Alberich[2], A. Carreno-Torres[2], J. Rambla[2], F. J. Lopez[1], J. M. Mut[1], R. Easty[1], O. M. Llobet[2], M. Sitges[2], S. de la Torre[2], A. Navarro[2,3,4], P. Flicek[1], J. Paschall[1].* 1) European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK; 2) Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain and Universitat Pompeu Fabra (UPF), Barcelona, Spain; 3) Institute of Evolutionary Biology, Barcelona, Spain; 4) Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain.

The European Genome-phenome Archive (EGA) is a controlled-access collection of human genomic and phenotypic data, for the purpose of biomedical research. Our collection includes major reference datasets for rare and common diseases, for example UK10K, Blueprint and the International Cancer Genome Consortium (ICGC), as well as control datasets, such as the data generated by the Wellcome Trust Case-Control Consortium (WTCCC). These data include over 1000 studies, 270,000 samples and comprise over 2PB of genomic data. In order to improve access and discoverability of the data within our archive, we have implemented a three tier Beacon that is fully compliant with the GA4GH specifications. At tier one, the public EGA beacon confirms allele existence within a subset of EGA datasets; the tier two registered access beacon increases the number of data sets and adds extra information where consent exists, such as allele frequency, whereas the tier three controlled access beacon allows full access to all authorised datasets. EGA is working with ELIXIR to share this technology across ELIXIR nodes. For registered users the facility to automate the application for full access to datasets with allele of interest is planned. The full GA4GH variant API is being developed for authorised users as EGA moves towards a secure cloud compute environment. This utilises the EGA's new data streaming download service that streams slices of encrypted data from the archive. Additional new services also enable the user to query all EGA metadata through a REST API, with the associated links to ENA, EVA, and BioSamples, both placing these data in the correct context and integrating these data with associated public data, for example UK10K and GoNL aggregated public data at EVA. These changes demonstrate how we at the EGA are implementing fundamental changes to facilitate data discovery and improve data dissemination to enhance the experience of EGA users. The EGA is jointly maintained by European Bioinformatics Institute (EMBL-EBI) and the Center for Genomic regulation (CRG), and this collaboration is motivating further developments towards a geographically distributed and federated archive, facilitating access across differing legal boundaries world-wide. The EGA is available at both www. ebi. ac. uk/ega/ and https://ega. crg. eu/.

1865T

An empirical evaluation of redundant annotations in common reference sources for tertiary analysis. *J. Warren, J. Li, A. Chhibber, E. Colak, A. Kiani, N. Bani Asadi, S. Barr, H. Y. K. Lam.* Bina Technologies, Roche Sequencing, Redwood City, CA.

After obtaining the genetic variants from next generation sequencing data, a precursory step in tertiary analysis is to annotate each variant with available relevant information. There is no standardized compendium for this purpose; researchers are instead required to compile data from a motley of annotation tools and public datasets. These annotation sources are quite diverse, with separate references providing information about different domains such as predicted functional impacts, known disease associations and population allele frequencies. Once assembled, the resulting data is subsequently analyzed to identify potential causative variants for the disease in question.

The choice of annotations sources accordingly has a direct impact on the results of the analysis, but the selection process is fraught with potentially overlooked complications. Many datasets provide the same information, but there is no guarantee of consistency in the reported data. Moreover, many public datasets are the result of studies of specific diseases or conditions, and the study selection criteria can bias the dataset contents. Predictions from functional impact tools also vary due to differences in assumptions and underlying data models. In sum, variants identified as potentially causative in one analysis could be omitted in another using a different set of annotation sources.

We have performed an empirical study of the quality and consistency of redundantly reported information to understand the impacts of selecting different annotation sources. We specifically examined differences in degree of agreement, predicted functional impacts, reported prevalence in populations, identified exonic regions and error rate. As anticipated, concordance of redundantly reported information is high, but data inconsistencies and errors are observed in effectively all aspects.

To illustrate the characteristics of using different datasets beyond reporting statistics, we also provide a funnel analysis of the NA12878 trio using different annotation sources. In this analysis, we filter the genomic variants using multiple criteria (e. g. rare variant, predicted damaging, associated with disease) using different data sources, reporting cardinality and overlap and each stage. This study illustrates the potential issues researchers may encounter during tertiary analysis.

1866F

Integrated analysis of germline, omic and disease data. *Z. Yang[1], D. C. Thomas[1], D. V. Conti[1,2].* 1) Department of Preventive Medicine, University of Southern California, Los Angeles, CA, USA; 2) Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA, USA.

The availability of various omics data, such as metabolites, expression, and somatic profiles, facilitates potentially new insights into the underlying etiologic mechanism of disease. However, such data presents many analytic challenges including effect heterogeneity and high dimensionality. Recently proposed methods for omic data often ignore the underlying causal relationships of the various data types and focus mostly on data reduction by estimating underlying clusters. Here, we present a novel approach for the integrated analysis of germline, omic and disease data. Via a specific directed acyclic graph (DAG), we use a latent variable to relate information from germline genetic data to either a continuous or binary disease outcome. Within a measurement error framework, the omic data is viewed as a flawed measure of underlying latent clusters, categorized to simplify interpretation. We use an expectation-maximization (EM) algorithm to simultaneously estimate the unobserved latent clusters and model parameters, including genetic effects on the latent cluster and the impact of the cluster on omic patterns and on the disease outcome. Additionally, we incorporate penalized methods for variable selection in a high dimensional setting for both the genetic data and the omic data. Using simulations, we demonstrate the ability of our approach to accurately estimate underlying clusters and their corresponding genetic, omic and disease effects. Moreover, we demonstrate the feasibility of the variable selection to identify genetic and omic factors as both the means and correlational structures are varied. We discuss extensions to accommodate ascertainment and missing data.

1867W

Estimating Components of Heritability Explained by Gene Expression. *L. J. O'Connor, A. L. Price.* Harvard University, Cambridge, MA.

GWAS have identified many noncoding variants that contribute to disease risk, and it is increasingly clear that the genetic component of disease risk is largely driven by regulatory variation. It remains an open problem to quantify the proportion of SNP-heritability (*hg2*) that is explained by gene expression in a given tissue. We introduce a new method, called χ^2 regression, to estimate this proportion. Instead of using gene expression and phenotype measurements in a single cohort, we use disease χ^2 statistics and eQTL χ^2 statistics from separate cohorts; this SNP-centric approach eliminates environmental confounders, and it allows a gene-expression cohort to be analyzed in conjunction with GWAS summary statistics at much larger sample size. In order to integrate the two datasets, we use a simple idea: if gene expression impacts disease risk, then the effect size of a SNP on gene expression will impact its effect size on disease risk accordingly. More precisely, we regress disease χ^2 statistics on eQTL χ^2 statistics, summed across genes, to obtain the phenotypic variance explained by gene expression. The eQTL χ^2 statistics (which are proportional to squared effect sizes) are shrunk, so that they equal the posterior expected value of the true effect size; this shrinkage not only eliminates downward bias in the estimator but also reduces its variance, exploiting the fact that *cis* variants have much larger expected effect sizes. The method distinguishes genetic effects that are not mediated by gene expression, and it can partition explained variance by gene set or tissue type. We show that the proposed method is accurate and unbiased in simulation studies, and we apply it to real datasets.

1868T

Standardized analysis and sharing of genome-phenome data for rare disease research through RD-Connect's platform. *S. Beltran[1], D. Piscia[1], S. Laurie[1], A. Cañada[2,13], J. M. Fernández[2,13], J. P. Desvignes[3,4], M. Thompson[5], R. Kaliyaperumal[5], E. van der Horst[5], S. Lair[6], P. Sernaleda[7], C. Kingswood[1], M. Girdea[8], M. Brudno[8], A. Blavier[6], R. Thompson[9], H. Lochmüller[9], M. Bellgard[10], J. Paschall[11], P. Lopes[7], M. Roos[5], P. A. C. 't Hoen[5], V. de la Torre[2,13], A. Valencia[2,13], D. Salgado[3,4], C. Béroud[3,4,12], I. Gut[1], RD-Connect Consortium.* 1) Centro Nacional de Análisis Genómico, Barcelona, Catalonia, Spain; 2) Centro Nacional de Investigaciones Oncológicas (CNIO) , Madrid, Spain; 3) Aix-Marseille Université, Marseille, France; 4) Inserm, UMR_S 910, Marseille, France; 5) Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands; 6) Interactive Biosoftware, Rouen, France; 7) DETI/IEETA, University of Aveiro, Portugal; 8) Centre for Computational Medicine, Hospital for Sick Children and University of Toronto, Toronto, ON, Canada; 9) Institute of Genetic Medicine, MRC Centre for Neuromuscular Diseases, Newcastle University, UK; 10) Centre for Comparative Genomics, Murdoch University, Perth, Western Australia; 11) European Molecular Biology Laboratory , European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Cambridge, United Kingdom; 12) APHM, Hôpital TIMONE Enfants , Laboratoire de Génétique Moléculaire, Marseille, France; 13) Instituto Nacional de Bioinformática (INB), Spain.

RD-Connect (rd-connect. eu) is an EU FP7 funded project building an integrated platform to narrow the gaps in rare disease research, where patient populations, clinical expertise and research communities are small in number and highly fragmented. RD-Connect's system securely integrates -omics data with biosample and clinical information using standardized analysis pipelines and terminologies such as the Human Phenotype Ontology, providing not only a centralized data repository but also a sophisticated and user-friendly online analysis platform. Although the system is being developed in a close collaboration with two EU flagship omics research projects, NeurOmics and EURenOmics, several other projects have already committed to providing data and RD-Connect gladly accepts submission proposals. Raw genomic data (from Whole Genome, Exome or gene panel NGS) is deposited at the European Genome-phenome Archive (EGA) and the corresponding clinical information is entered into a PhenoTips instance. The data is then processed by RD-Connect's standardized mapping, variant calling and annotation pipeline, which has around 99% precision and sensitivity when compared to the NA12878 reference set of calls from the NIST. The results are made available to authorized users through the platform, which runs on a Hadoop cluster and uses technologies such as ElasticSearch, Postgres, Scala and Angular. js, making it highly configurable and efficient. The user-friendly interface enables filtering and prioritization of variants using the most common quality, genomic location, effect, pathogenicity and population frequency annotations, including CADD and ExAC. Additional tools can be integrated at several levels: DiseaseCard, Alamut Functional Annotation (ALFA), UMD Predictor and gene-disease relationships in nanopublication format are already available. Current focus is on the integration of Exomiser (including PhenIX) to prioritize variants through genotype-phenotype queries, the provision of reliable allele frequencies, the lighting of a GA4GH beacon and patient matchmaking. A pre-release of 367 exomes was made available to selected users for beta-testing at platform. rd-connect. eu. A first official release including data from ~1000 exomes and access by any authorized researcher is expected by 2015 Q4.

1869F

Integrative longitudinal analysis of ribosome occupancy and protein synthesis during chemotherapeutic response reveals complex translational dynamics. *T. -Y. Liu[1], H. H. Huang[2], D. D. Wheeler[2,3], J. A. Wells[3], Y. S. Song[1,4,5], A. P. Wiita[2].* 1) Computer Science Division, University of California, Berkeley; 2) Dept. of Laboratory Medicine, University of California, San Francisco; 3) Dept. of Pharmaceutical Chemistry, University of California, San Francisco; 4) Dept. of Statistics, University of California, Berkeley; 5) Dept. of Integrative Biology, University of California, Berkeley.

Dynamic changes in the cancer proteome control tumor growth, proliferation, metastasis, and response to the therapy. Hence, it is important to understand the translational regulation in cancer cells during chemotherapeutic response. We studied multiple myeloma cells exposed to a low dose of bortezomib, designed to elicit a drug-induced stress response but not lead to widespread translational shutdown and cell death. mRNA-seq, ribosome profiling, and pulse-chase isotopic labeling mass spectrometry-based proteomics were integrated to directly monitor the synthesis of new proteins and degradation of existing proteins across a time course. The longitudinal study enabled us to describe translation rate as a function of transcript abundance and footprint density. We modeled the dynamic changes with a system of differential equations and solved the equation using functional data analysis. The pulse-chase mass spectrometry technique enables us to extract the degradation rate constant to disentangle the degradation process from the synthesis process. And the translational rate parameter is modeled as a time varying function. While it is widely assumed that ribosome footprint occupancy directly reflects protein synthesis, to our knowledge this assumption has not been previously tested in the same experimental system across a time course. We directly compared measurements of translational efficiency from ribosome profiling and the estimates of translational rate parameter using our proposed quantitative model. Results showed that they are imperfectly correlated. However, by incorporating a gene-specific modulation factor, a function of active ribosomes bound to transcript were incorporated into the quantitative model, we were able to correlate the translational efficiency and translational rate parameter. Our model implies gene-specific modulation factors that govern the relationship between footprint density and protein synthesis. These factors are found universal during chemotherapeutic response in our experiments. Our work offers a novel quantitative framework to understand translation using a combination of emerging technologies. Taking advantage of this model with concurrent biochemical and genetic experimentation may allow us to identify these factors that govern translational regulation in cancer and potentially eukaryotes more broadly, and shed light on targeted therapies.

1870W

Structural variation discovery in dogs using whole-genome sequencing. *M. Arumilli[1], MK. Hytönen[1], E. Salmela[1], J. Salojärvi[2], H. Lohi[1].* 1) Department of Veterinary Biosciences and Research Programs Unit, Molecular Neurology, University of Helsinki, Finland; 2) Department of Biosciences, University of Helsinki, Helsinki, Finland.

Background: Copy number variation (CNV) contributes to genetic and phenotypic diversity within and between species. CNVs in domestic dog have been correlated with many phenotypes, which are of comparative interest in humans. A reduced genetic variation in dog breeds than humans makes it feasible to study complex genetic traits in dogs with smaller sample sizes. In dogs, only four CNV studies using comparative genomic hybridization arrays (aCGH) and one CNV screen using CanineHD 170K genotyping array have been previously reported presenting a catalogue of CNVs. However, these approaches suffer from hybridization noise, limited coverage of genome, low resolution and inaccurate detection of breakpoints. Here, we expand from the earlier studies by carrying out the first Structural Variation (SV) study in dogs based on the whole-genome sequencing technology to a) catalogue the SV in dogs across breeds at ultra-high resolution b) compare the accuracy of CNV detection with genotyping array methods. In addition, we aim to perform an association study to identify CNVs associated with osteochondrosis phenotype in Border Collie breed. **Results:** SVs were detected using the Lumpy and GenomeSTRiP from whole-genome sequences of 159 dogs from 53 breeds. Many of the CNVs appeared to be highly breed-specific and previously known CNVs were generally replicated by our analysis. Furthermore, the comparison of CNVs from SNP genotype data of 22 individuals from IlluminaHD array with the same set of sequenced individuals showed that sequencing-based approach detected CNVs with more precise breakpoint detection than array based methods. In addition, comparing 13 affected individuals with osteochondrosis and 9 healthy individuals from Border Collie suggested CNV regions that can possibly contribute to the phenotype. **Conclusions:** This study provides a comprehensive analysis of canine CNVs cataloguing the CNV regions in different breeds to identify CNVs that contribute to breed specific phenotypes. The CNV regions identified for osteochondrosis may provide novel insights to study the complex genetics of the disorder in humans.

1871T

The Human-Mouse Disease Connection (HMDC) Portal: recent updates enabling discovery. *J. T. Eppig, S. M. Bello, J. A. Kadin, J. E. Richardson, C. L. Smith, MGI Staff.* Jackson Laboratory, Bar Harbor, ME.

The laboratory mouse is used extensively as a model system to investigate the eitopathogenesis of human disease and to test new clinical therapeutics. The Mouse Genome Informatics (MGI, www. informatics. jax. org) resource has recently added new features to its human-mouse translational tool, the Human-Mouse: Disease Connection (HMDC, www. diseasemodel. org). This tool, which provides simultaneous search and access to human and mouse genomic, phenotypic, and genetic disease information, assists researchers in identifying and evaluating candidate genes and mouse genotypes modeling a spectrum of clinically relevant phenotypes. Users can search the HMDC portal from a human or mouse standpoint, using (1) genes or gene IDs for either species, (2) genome locations from either species, and (3) mouse phenotype or human disease (OMIM) terms. Alternatively, data can be uploaded from human or mouse VCF files or text files of gene symbols or IDs. The HMDC phenotype and disease search has been updated to allow Boolean search expressions and includes an optional autocomplete feature. Initial search results display in an interactive grid for visual comparison of phenotypes and diseases across multiple genes, phenotypes, and diseases. A hybrid of HomoloGene and HCOP homology relationships is used to relate human and mouse genes in the grid. This hybrid maximize the number of homology groups including both human and mouse genes. The grid employs color cues that reflect the depth of human and mouse annotations. Grid cells are dynamically linked to underlying MGI data, including phenotypic detail, model-relevant publications, and global availability of mouse resources via IMSR (www. findmice. org). Alternate HMDC views include gene- and disease-centric information presented in tabular format. We will present examples of current content and capabilities of the HMGC Portal, including (1) display of known mouse model data for human disease, (2) potential human candidate genes for disease based on mouse phenotypes, (3) and potential target genes for developing new human disease models in mice. Supported by NIH grant HG000330.

1872F

Homology curation at SGD: Yeast and yeast research inform genetic medicine. *S. R. Engel, M. C. Costanzo, R. S. Nash, E. D. Wong, J. M. Cherry, The SGD Project.* Genetics, Stanford University, Palo Alto, CA.

The foundation for much of our understanding of basic cellular biology has been learned from the budding yeast*Saccharomyces cerevisiae*,and studies with yeast have provided powerful insights into human genetic diseases and the cellular pathways in which they are involved. This utility of yeast as a model for human disease arises from the biochemical unity that underlies all forms of life. Yeast has become extremely useful in the study of various diseases that afflict humans, such as cystic fibrosis, kidney disease, mitochondrial diseases, and neurodegenerative diseases such as Parkinson's. Recent work with humanized yeast (in which yeast genes have been replaced with human orthologs) and humanized yeast proteins (in which key residues have been altered to match the human sequence) has demonstrated extensive conservation of ancestral functions through time and across taxa. We will present an update on new developments at the*Saccharomyces*Genome Database (SGD;www. yeastgenome. org), the premier community resource for budding yeast. In order to promote and support the ways in which yeast and yeast research can inform genetic medicine, we are providing comprehensive curation for human disease-related genes and their yeast orthologs, including high quality manually curated information regarding functional complementation and conserved function. We also associate sequence changes with variations in yeast phenotypes and corresponding human disease manifestations. Curated information for yeast genes will be displayed on new Homology pages at SGD. Curated information for human genes will be available from the new "Yeast to Human Portal" knowledge center athumanportal. org. This new information is provided in ways that allow data mining and encourage innovation, for researchers studying both yeast and other organisms. These expanded efforts are part of our continuing mission to educate students, enable bench researchers, and facilitate scientific discovery. This work is supported by a grant from the NHGRI (U41 HG001315).

1873W

Applications of long read sequencing: Human cDNA sequencing on the Oxford Nanopore MinIon. *S. Goodwin, A. Dobin, L. See, TR. Gingeras, WR. McCombie.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Next-Generation sequencing has revolutionized our understanding of human genetics and the genetics of human disease. Short read, high-throughput genetic sequencing has become the work horse of sequencing science and has proven to be an invaluable tool in our understanding of genetic variability. However, the detection of large rearrangements, identification variations in highly repetitive regions, phasing genomes to track inheritance blocks, and elucidating exon connectivity in long transcripts all remain exceedingly difficult in short read sequencing routines. Long read sequencing routines, such as those employed by Pacific Biosciences and Oxford Nanopore may prove to be essential to fully characterize structural variations, and to identify precise exon connectivity within the transcriptome. Because these technologies generate long, contiguous reads in excess of 10kb they are able to span very large structural elements and full length isoforms that would otherwise be unresolvable in short read sequencing. Over the last few years, significant advances have been made in the field of long read sequencing bringing their utility to the strength needed to address research into human variation. To highlight the utility of long reads and to investigate the handheld Oxford Nanopore MinIon device, we attempted cDNA sequencing of the human GM12878 cell line. From a single flowcell we were able to generate >40k consensus reads, of which > 10k were able to be annotated. From these 10k annotated transcripts we were able to identify more than 1000 novel transcripts including both novel extensions and splice isoforms. We were also able to show that the Oxford Nanopore reads are able to accurately identify transcription start and stop sites. We are currently undertaking additional experiments involving barcoding and other optimizations to increase the yield of annotated full length transcripts. These results can lead to greater insight into the underlying interaction of RNA and phenotype and the rapidly evolving Oxford Nanopore MinIon (predicted to generate more than 20Gb of data in a single run by 2016) is a potential tool for facilitating research of this nature.

1874T

Fast and computationally efficient joint genome calling using Spiral Encrypted Compression (SpEC). *J. Bruestle, B. Drees, A. Mangubat, S. N. Shekar.* Spiral Genetics, Seattle, WA.

The recent reduction in sequencing costs and increase in sequencing capacity indicates that the future of NGS analysis is moving rapidly toward large population-based studies. There are two important bottlenecks to the effective multi-sample analysis needed for these studies: ability to store the data and sufficient accuracy of variant calls to be able to compare across samples. Here we show that Spiral Encrypted Compression (SpEC) addresses the storage and accuracy issues in population-scale sequencing studies by storing read and alignment data with enhanced compression and allowing for information to be withdrawn at any genomic location on demand rapidly. SpEC is built to be lossless and compresses both the reads, as well as other fields, optimizing the compression of the data that are stored.

We compare the SpEC to CRAM by compressing a standard 50x NA12878 Illumina HiSeq data set (114GB) . SpEC compresses the file to 61GB (46. 5% reduction), compared to 76GB (33% reduction) with CRAM in lossless mode. When decompressing the SpEC file, the original BAM file is recreated bit for bit. When the CRAM is decompressed, all the data are restored. However, the fields are in a different order compared to the original BAM. Further, the compression time on a standard machine was approximately 2 hours for SpEC but 7 hours for CRAM.

SpEC allows for high accuracy variant calls. If a variant is not called in the VCF, but it is important to confirm whether it exists, SpEC can be used to quickly search for unique subsequences associated with the variant. The data can be searched even while compressed. This allows for backfilling of the variant calls in a sample with a greater level of certainty. This approach can also be used to quickly identify structural variants. We expect SpEC to be useful for joint variant calling offor SNPs, indels and structural variants in population-scale studies.

Overall, SpEC allows for a reduction in hardware costs, and the ability to accurately identify whether particular variants exist in the data. Both of these benefits are expected to be of considerable benefit to rigorous population based studies.

1875F

NGS Library Prep Methods to Achieve Comprehensive Coverage for WGS and WGBS from Low DNA Input. *L. Kurihara, J. Laliberte, C. Couture, C. Schumacher, S. Sandhu, R. Spurbeck, J. Irish, S. Chupreta, V. Makarov.* Swift Biosciences, Inc. , Ann Arbor, MI.

When performing whole genome sequencing (WGS) or whole genome bisulfite sequencing (WGBS), unbiased, even coverage of the genome is required in order to conduct comprehensive analysis from the lowest possible sequence read depth. Highly efficient conversion of DNA fragments into library molecules is also necessary when DNA input quantity or quality is limited. We have developed two library preparation methods to achieve these results for WGS, WGBS, or when targeted hybridization capture is performed. The WGS method uniquely repairs damage on both the 3' and 5' termini to enhance ligation efficiency to sheared DNA fragments. Combined with sequential ligation steps that optimize attachment to each terminus, this single tube 'with bead' method supports PCR-free sequencing from inputs as low as 10 ng circulating, cell-free DNA or 100 ng physically sheared DNA. Input quantities down to 10 pg can be used with PCR amplification. Efficiency of library conversion is ~50% for physically sheared DNA and up to 90% for circulating, cell-free DNA. Adapter titration is not required with reduced input quantities, which ensures efficient adapter ligation at all supported inputs and simplifies the workflow. Human WGS using this method demonstrates high complexity with exceptional coverage of GC-rich promoter regions compared to other methods. At inputs as low as 1 ng human DNA, PCR duplicates were present, however at 18X coverage, the genome was fully represented without loss of data or reduction in relative coverage of GC-rich promoter regions. For WGBS, our library preparation is performed post-conversion on denatured, bisulfite-converted fragments. This improves library recovery up to 100X compared to traditional library prep methods that ligate methylated adapters to double-stranded DNA prior to bisulfite conversion. Our efficient adapter attachment to single-stranded DNA supports inputs from 100 pg to 100 ng. Libraries made using this method require less PCR amplification than other methods, thus reducing the number of PCR duplicates. Human WGBS demonstrated comprehensive coverage of CpG islands when 10 ng input was used. Even at lower depth of sequencing to an average 11. 4X coverage from the 10 ng input library, 99. 7% of CpG islands were covered and showed an average depth of coverage of 7. 8X. This library preparation method enables single base resolution of methylation status throughout the genome, even from limiting DNA input quantities.

1876W

Performance comparison of four commercial human whole-exome capture platforms. *D. Shigemizu, Y. Momozawa, T. Abe, T. Morizono, KA. Boroevich, S. Takata, K. Ashikawa, M. Kubo, T. Tsunoda.* RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.

Whole exome sequencing (WXS) is widely used to identify causative genetic mutations of diseases. However, not only have several commercial human exome capture platforms been developed, but substantial updates have been released in the past few years. We report a performance comparison for the latest release of four commercial platforms, Roche/NimbleGen's SeqCap EZ Human Exome Library v3. 0, Illumina's Nextera Rapid Capture Exome (v1. 2), Agilent's SureSelect XT Human All Exon v5 and Agilent's SureSelect QXT, using the same DNA samples. Agilent XT showed the highest target enrichment efficiency and the best SNV and short indel detection sensitivity in coding regions with the least amount of sequencing. Agilent QXT had slightly inferior target enrichment than Agilent XT. Illumina, with additional sequencing, detected SNVs and short indels at the same quality as Agilent XT, and showed the best performance in coverage of medically interesting mutations. NimbleGen detected more SNVs and indels in untranslated regions than the others. We also found that the platforms, which enzymatically fragment the genomic DNA (gDNA), detected more homozygous SNVs than those using sonicated gDNA. We believe that our analysis will help investigators when selecting a suitable exome capture platform for their particular research.

1877T

Evaluation of HiSeq X Ten Performance: Towards Clinical Applications. *K. Walker[1], C. Buhay[1], R. Sanghvi[1], Q. Wang[1], H. Doddapaneni[1], J. Hu[1], M. Wang[1], Y. Han[1], H. Dinh[1], E. Boerwinkle[1,2], R. A. Gibbs[1], D. M. Muzny[1].* 1) Baylor College of Medicine HGSC, Houston, TX; 2) University of Texas Health Science Center at Houston, Houston, TX.

High-throughput parallel nucleotide sequencing has revolutionized genomic research and reshaped applications in clinical health care. The HiSeq X Ten platform further expands these opportunities with unprecedented capacity. The Human Genome Sequencing Center (HGSC) at Baylor College of Medicine adopted the HiSeq X Ten system in the fall of 2014, with a view to eventual deployment in a CAP/CLIA environment. To evaluate the instruments we have analyzed more than 145 flowcells, representing 1,040 30X human genomes. These studies have included common disease cohorts, inherited cancers, mendelian disease cases as well as DNA from cell lines of lung and endometrial cancer. We assessed several metrics of performance, including genome coverage, coding region coverage and error rate relative to standards. We also measured the number of read cluster duplicates and examined variants (in duplicate samples) in order to assess the sensitivity and specificity of the platform. These evaluation efforts have provided valuable insight as to how sequencing depth and coverage uniformity impact the ability to accurately detect variants. Overall the platform has been consistent with its lane-to-lane performance. We have, however, observed unique modes of error due to the use of patterned flowcells including an increase in clustered optical duplicates, correlating with library loading concentration. We also have observed a periodic decrease in Q30 values in the last 50bp of the Read 2 sequence. Each of these issues are being addressed with local and vendor-driven solutions, and progressive improvements show that it is well on the way to being as robust as other Illumina platforms. Finally, we have recently evaluated DNA libraries prepared with Swift Biosciences Accel-NGS 2S DNA Library kits. These have exhibited excellent genome coverage in GC-rich regions for both PCR and PCR-free preparations even with FFPE sample inputs as low as 10 ng. These and other innovations illustrate that there remain many avenues for process improvement.

1878F

Accurate quantification and qualification of FFPE samples increases success rate of NGS library prep and sequencing. *B. Arezi, B. Rogers, K. Chen, B. Hsue, H. Hogrefe.* Agilent Technologies, La Jolla, CA.

FFPE tissue archiving is the most widely used method for clinical sample preservation, and provides a valuable source of diverse genetic information for cancer biomarker discovery. DNA recovered from FFPE tissues exhibits varying degrees of fragmentation, cross-linking, deamination, depurination and other lesions due to formalin fixation, paraffinization, and storage conditions; and as a result, NGS library preparation is often challenging. To increase success with FFPE-derived DNA, we developed a qPCR-based method to determine quantity of amplifiable DNA and extent of degradation. SureSelectXT libraries were prepared from multiple FFPE samples of varying integrity, and enriched using the ClearSeq Comprehensive Cancer Panel which targets 151 genes frequently mutated in solid and hematological cancers. As we will show, qPCR integrity scores are highly correlated with pre-capture PCR yield and sequencing metrics (library complexity and coverage), and can be used as a guide to determine appropriate sequencing depth for optimal coverage.

1879W

Repair of FFPE Extracted DNA Increases Next Generation Sequencing Library Yields and Quality. *F. Stewart[1], L. Chen[1], P. Liu[1], L. Ettwiller[1], R. Corbett[2], H. McDonald[2], P. Pandoh[2], C. Sumner[1], E. Dimalanta[1], T. Davis[1], Y. Zhao[2], M. Marra[2], T. Evans[1].* 1) New England Biolabs, Inc. , Ipswich, MA; 2) Genome Sciences Centre, BC Cancer Agency, Vancouver, Canada.

Clinical materials are commonly archived as Formalin-Fixed, Paraffin-Embedded (FFPE) samples. While these samples are an invaluable source of information about genetic alterations in human disease, especially cancer, unfortunately the methods used for fixation and storage significantly damage the nucleic acids contained within. As a result, it can be challenging to obtain useful information, including high-quality sequence data, especially when sample amounts are limited, as is frequently the case. In this work, we investigated the effects of DNA repair on library preparation and sequencing of FFPE samples. We evaluated a mix of repair enzymes that is designed to work on a broad range of DNA damage including modified bases, nicks, gaps, and blocked 3' ends. Results from FFPE DNA samples of varying quality show that DNA repair generally increases library yields, and improvements from 10% to 458% have been observed. Sequencing data analysis demonstrates that base calling qualities for all 4 bases are improved after DNA repair. Additionally, the G:C to A:T mutations that result from deamination of cytosine to uracil (a signature of FFPE DNA damage) are repaired by treatment with the DNA enzyme repair mix. Interestingly, sequencing miscalls for base pair changes not typically associated with fixation were also reduced. Importantly after DNA repair, a positive impact on read mapping and read pairing was observed, resulting in the generation of more useable data. Treatment of FFPE DNA by the repair enzyme mix is easily implemented upstream of library construction, enabling fast protocol times and easy automation. We anticipate that these improvements to the FFPE DNA sequencing workflow will enable the analysis of many FFPE samples that otherwisewould not be accessible.

1880T

iDASH: a secure cloud environment for biomedical data analysis, storage and sharing. *O. Harismendy[1,3], J. Kim[3], C. Farcas[3], A. Koures[2,3], L. Ohno-Machado[3].* 1) UC San Diego Moores Cancer Ctr, La Jolla, CA; 2) UC San Diego Clinical And Translational Research Institute, La Jolla CA; 3) Department of Biomedical Informatics, UC San Diego, La Jolla, CA.

Current analysis of biomedical data for research is confronted with a number of challenges: 1) a growing volume of data, 2) an increasing complexity and diversity of data and workflows, 3) a highly collaborative research environment and 4) growing concerns for data safety and privacy. Cloud computing and cloud storage have become major information technology solutions. While they are also becoming popular for the analysis of biomedical data, it is not always clear if and how they address these specific challenges. In particular the genome sequence of the participants in translational research studies is considered Protected Health Information and its storage and analysis must be performed in compliance with HIPAA regulation. Here, we present the iDASH private cloud, a HIPAA computing environment. Access to the resource uses RSA-based two-factor authentication behind redundant firewalls. Located in a Virtual Private Network within UC San Diego Supercomputer Center, iDASH benefits from point to point connection to major US academic centers, a security advantage over standard internet connections. The resource uses VMware virtualization technology to manage and instantiate virtual machines with up to 32 vCores, 256Gb of RAM and 1Tb disk under a 10G network, accessing both high capacity and high-speed storage. We have customized this environment for the analysis of human genetic data. Specifically, we provide a local copy of public resources such as 1000 Genomes, dbSNP or the human reference genome. We also have prepared and tested virtual machine images dedicated to genomic analysis that include popular analysis tools (BWA, BOWTIE, samtools, vcftools) and frameworks (bcbio-nextgen). Finally, we built FlightDeck, a web-based portal to prepare analysis manifests based on popular workflows and recipes, such as variant calling, and start the analysis with a seamless integration to the cloud in the background. We will present the results of testing standard workflows including sequencing read alignment and variants calling. The iDASH environment is a highly secure, academic computing resource geared towards the analysis of large biomedical data. IDASH is a member of the NIH/NHLBI National Centers for Biomedical Computing (U54HL108460).

1881F

Mutations on a DNase I Hypersensitive Sites in Human lead to Tumorgenesis. *W. Jin, K. Zhao.* Systems Biology Center, NHLBI,NIH, Bethesda, MD.

Cellar differentiation is dependent on gene transcription that is controlled by a large number of proximal and distal regulatory elements in the genome. These elements are focal points of various histone modifications and are located in accessible chromatin regions as indicated by the hypersensitivity to DNase I digestion. Recently, the combination of DNase I digestion with next generation sequencing (DNase-Seq) is becoming a very popular approach for accurate identification of DNase I hypersensitive sites (DHS) across the entire genome, which has significantly advanced our understanding of the dynamics of chromatin accessibilities and gene regulations during biological processes such as cell differentiation, proliferation, and response to the environment stimuli. However, these conventional DNase-Seq methods for genome-wide DHS profiling require millions of cells, which limits their utility. Here we report, an ultrasensitive and robust strategy, called Pico-Seq, for detection of genome-wide DHSs using only a small number of cells. Our analysis showed that Pico-Seq libraries made by 1 thousand cells could perform the same as that of conventional DNase-seq libraries using millions of cells. We apply Pico-Seq to the very small number of tumor and adjacent normal cells, dissected from formalin fixed paraffin embedded tissue slides from thyroid cancer patients, and detect thousands of tumor-specific DHSs, some of which seemed to have clinical relevance not previously appreciated. Analysis of the DHS sequences uncovers tens of mutations in tumor cells, one of them seems to be a drive mutation in the tumorgensis. The mutation disturbs a p53 binding motif and lead to significantly changes in the expression of genes critically involved in cancer development. In conclusion, Pico-seq can reliably detect DHSs in libraries using a few cells, which greatly extends the range of applications of DHS analysis for both basic and translational research and may provide critical information for personalized medicine.

1882W

Development of a Low Input FFPE workflow for Whole Exome Sequencing. *B. Marosy, B. Craig, K. Hetrick, P. D. Whitmer, H. Ling, S. Griffith, B. Myers, K. F. Doheny.* Center for Inherited Disease Research (CIDR), Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. CIDR is continually seeking new ways to improve its workflow and generate high quality data. Historically, sample requirements for a variety of targeted selection and NGS platforms have been a limiting factor for investigators. Regardless of input, not all samples are of high quality. Formalin Fixed Paraffin Embedded (FFPE) samples are a useful source of DNA for NGS studies but present additional challenges in the lab due to age of the sample, damage to the DNA from fixation resulting in changes to the nucleotide sequence and fragmentation of the DNA. Here we investigate different library prep kits (KAPA Hyper Prep (KHP) and Swift Accel-NGSTM 2S) designed for low input (50ng); DNA repair methods (USER, NEB PreCR, NEBNext FFPE DNA Repair); optimization for small fragment size (KHP and Swift Accel-NGSTM 1S); and quantitation methods (KAPA hgDNA Quantitation and QC kit) to develop a best practices workflow to generate libraries for whole exome sequencing from low input and FFPE derived samples. Both HapMap samples and FFPE samples of varying age and quality were used to generate libraries using these methods. Libraries were then enriched for whole exome sequencing using Agilent SureSelectTM XT and were sequenced on the Illumina HiSeq 2500 platform. Evaluation and optimization of a variety of conditions and commercial kits has enabled us to establish a workflow for low input and compromised samples for library prep using 50ng of DNA to generate high quality exome enriched sequencing data.

1883T

Massively parallel identification and annotation of causal 3'UTR regulatory variants. *A. Biton, O. Le Tonquèze, D. Torgerson, W. Zhao, N. Zaitlen, D. Erle, The CAAPA Consortium.* Department of Medicine, Lung Biology Center, University of California, San Francisco, San Francisco, California, USA.

Most genetic variants associated with diseases are found in non-coding regions, but our understanding of how these variants affect gene function remains very limited, as does our ability to determine which of the multiple linked associated variants in a region is causal. To address these issues, we recently developed a powerful massively parallel experimental method for functional analysis of sequences from 3'UTRs (fast-UTR, Zhao *et al. , Nat Biotechnol* 32:387, 2014). The 3' untranslated region (3'UTR) of genes contains binding sites for regulatory proteins and non-coding RNAs that regulate mRNA expression, stability, and translation. In our previous work, we demonstrated that fast-UTR accurately models difference in mRNA stability within a set of ~2000 selected 3'UTR sequences, identified binding motifs of known and novel 3' UTR cis-regulatory elements, and tested a small set of SNPs to demonstrate that fast-UTR could be used to detect effects of genetic variants. In this work, we are leveraging fast-UTR to survey the complete 3'UTR space of the human genome. We constructed a fast-UTR oligonucleotide library containing >200,000 3' UTR haplotypes encoded in 160-nt segments. These haplotypes include selected rare and common 3' UTR variants derived from whole genome sequences from the 1000 Genomes Project (>90,000 variants) and the Consortium on Asthma among African-ancestry Populations in the Americas (CAAPA) study (>12,000 variants). Rare variants were chosen based on their overlap with 1) regulatory motifs including AU-rich elements (ARE), constitutive decay elements (CDE), Pumilio, and miRNA binding sites, and 2) 3'UTRs of 400 asthma candidate genes identified from the Genetic Association Database (GAD) and the NHGRI GWAS catalog. We will report the activity of these sequences in three different cell types that are important for asthma pathogenesis: airway epithelial cells, CD4+ T cells, and airway smooth muscle cells. These data will provide an unprecedented framework to functionally annotate 3'UTR sequences as well as identify and predict functional consequences of individual genetic variants. These data will also help assess and interpret how SNPs discovered by eQTL and GWAS studies lead to variation in mRNA expression and thereby help understand the contributions of genetic variation to asthma and other diseases.

1884F

Towards a Standardized and Automated Workflow for Extracellular Vesicle Isolation and Characterization. *C. T. Schwartz, Z. Smith, B. Lee.* Beckman Coulter, Indianapolis, IN.

Introduction:Research involving extracellular vesicles (EVs) is rapidly expanding due to evidence suggesting a role in cancer metastasis and cell-to-cell communication. Improved and more efficient isolation and characterization processes for EVs are critical to advancing this exciting field and experts have recently called for the establishment of standardized methods. EV isolation is particularly difficult, requiring a labor-intensive density gradient centrifugation step to obtain highly pure vesicles. Downstream challenges involve a standardized sizing method and genetic profiling of encapsulated miRNA. Here, we describe a complete workflow from EV isolation to small RNA next-generation sequencing using Beckman's Vi-CELL Cell Viability Analyzer, DelsaMax Core, Biomek Automated Workstations, High-Performance and Ultracentrifuges, and SPRI-based reagents that improves the process and minimizes challenges. **Methods**:The Vi-CELL measured HCT 116 cell concentration and viability. Crude exosomes from the cell culture media were isolated by differential centrifugation. Gradient layering and fractionation for a discontinuous density gradient was achieved through a Biomek automated process, and following a centrifugation step, aliquoted fractions. Phenotypically-enriched populations of highly pure vesicles, assayed by Beckman's DelsaMax CORE, were separated, and total RNA was extracted utilizing Beckman's RNAdvance Cell v2 reagent kit with Biomek automation. Total RNA was analyzed using an Agilent BioAnalyzer 2100 for size and purity and subjected to a TaqMan® MicroRNA assay for quantification of relevant EV biomarkers. Finally, small RNA sequencing libraries were generated using the New England Biolabs NEBNext® Small RNA library kit, also automated on the Biomek platform. The libraries were subsequently sequenced on the Illumina MiSeq® platform and variant expression quantified for various EV fractions. **Summary**:The automated workflow provides a robust, high-throughput solution that can be translated to both basic and clinical research. The usual benefits of automation such as reduced processing time and user-to-user variability were achieved in the density gradient layering and fractionation, RNA extraction, and small RNA library prep processes. Lastly, next generation sequencing identified differentially-expressed genes among EV fractions with potentially clinical implications.

1885W

A Multi-Ethnic Genotyping Array for the Next Generation of Association Studies. *C. R. Gignoux[1], G. L. Wojcik[1], H. R. Johnston[2], C. Fuchsberger[3], S. Shringarpure[1], A. R. Martin[1], S. Rosse[4], N. Zubair[4], D. Taliun[3], R. Welch[3], C. Pethigoya[5], J. O'Connell[5], L. McAuliffe[5], C. Rosenow[5], N. S. Abdul-husn[6], G. Belbin[6], H. M. Kang[3], G. Abecasis[3], M. Boehnke[3], Z. S. Qin[2], J. M. Romm[7], C. A. Haiman[8], C. Kooperberg[4], R. J. Loos[6], T. C. Matise[9], K. E. North[10], C. Carlson[4], K. C. Barnes[11], C. D. Bustamante[1], E. E. Kenny[6], Population Architecture using Genomics and Epidemiology (PAGE) Study.* 1) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 2) Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Atlanta, GA; 3) Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI; 4) Fred Hutchinson Cancer Research Center, Seattle, WA; 5) Illumina Inc. , San Diego, CA; 6) Department of Genetic and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 7) Center for Inherited Disease Research, Johns Hopkins University, Baltimore, MD; 8) Department of Preventive Medicine, Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA; 9) Department of Genetics, Rutgers University, Piscataway, NJ; 10) Department of Epidemiology, School of Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC; 11) Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD. In the past decade, genome-wide association studies (GWAS) have been extraordinarily successful in uncovering genetic factors underlying complex traits. However much of this work has only been performed in populations of European descent. To address this disparity, a collaboration between Illumina, PAGE, CAAPA (Consortium on Asthma among African-ancestry Populations in the Americas), and T2D Genes Consortium developed the 1. 9M SNP Multi-Ethnic Genotyping Array (MEGA). This array is designed to be a single platform to interrogate diverse variation from across the frequency spectrum and screen for prior genetic discoveries. The GWAS backbone is informed by whole genome sequences from the 1000 Genomes Project (TGP) and CAAPA, with additional compatibility from the Illumina HumanCore array. We developed a novel cross-population tag SNP strategy to maximize imputation accuracy for low frequency variants across six continental populations, with improved performance from previous generations of arrays. Importantly the performance of the array is high across all continental TGP superpopulations (>90% accuracy for MAF >=1%). We chose rare, functional candidates from >36,000 multi-ethnic exomes, prioritizing loss-of-function and predicted damaging sites. We also curated variants with domain experts, including boosting coverage in regions of interest (e. g. MHC), over 5,000 ancestry informative markers, uniparental markers, and over 25,000 variants of clinical, prior GWAS, pharmacogenetic, and eQTL importance. A reference panel of several thousand individuals, including HGDP and a large panel of Native Americans, is available on MEGA to aid in rare variant calling, ancestry characterization, and admixture analyses. We will present data on MEGA genotyping of >60,000 African-American, Hispanic/Latino, East Asian and Native Hawaiian individuals from PAGE and CAAPA. We will describe imputation performance in these populations and others, including the improvements to imputation accuracy derived from larger whole genome reference panels. Further details, summary statistics and other relevant information generated from the platform will be placed in a centralized community repository at *www. pagestudy. org/mega*. We intend MEGA to be a platform and an analytical resource for researchers interested in large-scale studies of diverse populations from across the globe.

1886T

Simplified adapter design for sequencing the Genome-scale CRISPR-Cas9 Knockout (GeCKO) libraries on the Illumina platform. *W. Wang[1], Q. Ding[2], L. Parsons[1], J. Wiggins[1], L. Guo[1], J. Miller[1], A. Ploss[2].* 1) Institute for Integrative Genomics, Princeton University, Princeton, NJ; 2) Dept of Molecular Biology, Princeton University, Princeton, NJ.

High throughput loss-of-function screening is a powerful strategy to functionally characterize large numbers of genes. The genome-scale CRISPR (clustered regularly interspaced short palindrome repeats)–associated nuclease Cas9 Knockout (GeCKO) system can effectively generate loss-of-function mutations in most human or mouse genes simultaneously for positive or negative screening (Sanjana, N. E. et al. Nature Methods 11, 783–784, 2014). However, the indexed staggered PCR primers for the final amplification are quite long, in the range of 91nt to 100nt. Recent improvements to the Illumina sequencing platform, both chemistry and software, lead to much better read data quality for low diversity samples. Therefore, we tried to shorten the Illumina adapter regions of GeCKO guide RNA libraries to increase the efficiency and specificity of the final PCR. In the forward PCR primer containing the Illumina P5 sequence, the stagger sequences of 1nt to 10nt were removed and the barcodes were shortened from 8nt to 6nt. For low levels of sample multiplexing, the reverse PCR primer was reduced to contain only the Illumina P7 sequence followed by the 24nt PCR priming site, while the reverse sequencing primer, indexes, and the stagger sequences were all eliminated. Using the simplified PCR primer design with up-front barcodes in Read-1, the original and E coli amplified GeCKOv2 human sgRNA libraries were successfully amplified by PCR and sequenced on Illumina HiSeq 2500. The read quality Q30 scores were over 95%, and the abundance of all sgRNA constructs was characterized. Further improvement is underway to develop custom sequencing primers to allow dual index reads and increase nucleotide diversity in the sequencing reaction.

1887F

Enabling population-scale PCR-free whole genome sequencing. *J. Abreu, R. Hegarty, M. Coole, D. Dionne, L. Holmes, M. Lee, S. McDonough, A. Tam, T. Desmet, B. Haas, Y. Farjoun, D. Perrin, N. Lennon, S. Fisher, S. Gabriel.* Genomics Platform, Broad Institute of MIT and Harvard, Cambridge, MA.

PCR-free human whole genome libraries sequenced on the Illumina HiSeq X are the current NGS gold standard for population-scale genomics, mendelian disease gene discovery, and understanding the genetic components of common diseases. However, input amount requirements and library production at scale represent challenges to implementation for this method. Here we present a scalable whole genome sequencing process which we have implemented to enable production of >100 whole genomes per day. By streamlining this workflow and optimizing library construction steps such as shearing and reaction cleanups, we have been able to reduce the amount of input gDNA while minimizing variability in insert size. Using this process, PCR-free libraries can be created using as little as 250ng of gDNA. Moreover, by applying only a few cycles of PCR to this process, PCR-plus libraries may also be created with only 25ng of gDNA. Completed libraries are combined to create pools of 24 samples or greater prior to sequencing. These pools are then assessed for evenness of sample representation using the Illumina MiSeq. Sample pools are spread across the number of lanes required to meet target coverage on the HiSeq X. This workflow, combined with enhancements made to our automated HiSeq X setup process, has minimized library seeding density variability on the Illumina HiSeq X flowcell while maximizing PF yield.

1888W

Sequencing the needle in the haystack: targeted capture for forensic DNA sequencing. *M. Carpenter[1], K. Elmer[1], J. Degenhardt[1], C. D. Bustamante[1,2], T. Dickinson[1].* 1) IdentifyGenomics, LLC, Menlo Park, CA; 2) Department of Genetics, Stanford University, Stanford, CA.

Many forensic DNA samples are heavily degraded and contaminated, making traditional PCR- and STR-based approaches unfeasible. Furthermore, conventional STR profiling can only be used for identification and does not provide phenotype information. More recently developed multiplex PCR-based resequencing approaches do provide targeted SNP coverage, but are limited by the number of targets that can be sequenced simultaneously and perform poorly on highly degraded DNA. We have developed a method for the targeted hybridization-based capture of DNA fragments containing SNPs relevant to numerous aspects of ancestry and phenotype. Our method, which we term PhenoCap, allows the prediction of phenotypic traits ranging from autosomal ancestry to facial morphology in a single workflow. PhenoCap targets over 3500 SNPs, as well as the full mitochondrial DNA sequence, Y chromosomal SNPs, and STR loci. In addition, we have extensively optimized our DNA extraction and PhenoCap protocols to enable analysis of even the most degraded, contaminated samples, including mixtures derived from multiple individuals. We aim to make PhenoCap analogous to a "forensic exome" that will provide a comprehensive individual profile for even the most challenging samples.

1889T

Producing 768 whole exomes per day with a highly optimized transposase-based library preparation protocol. *A. Cheney[1], C. Walsh[1], M. Costello[1], A. Kia[2], M. He[2], J. Walsh[1], J. Thompson[1], T. Desmet[1], C. Friedrich[1], N. Lennon[1], D. Perrin[1], J. Abreu[1], S. Dodge[1], S. Gabriel[1].* 1) Technology Development, Genomics Platform, Broad Institute, Cambridge, MA. , USA; 2) Illumina Inc, 5200 Illumina Way San Diego, CA 92122.

With over 50,000 samples to prepare for whole exome sequencing in 2015, the Genomics Platform at the Broad Institute needed a more efficient sample preparation process than current off-the-shelf products, featuring shearing followed by multiple enzymatic steps, enabled. In response, we have scaled up our process and achieved faster turnaround times by implementing a modified workflow using transposase-mediated library construction. Through a series of chemistry optimizations in collaboration with Illumina, and process optimizations within the Genomics Platform, we have incorporated an enhanced version of Illumina's Nextera Rapid Capture Exome into our fully automated workflow for efficient sample preparation that produces diverse libraries, enabling superior germline exome coverage equivalent to traditional sheared library construction. Chemistry optimizations include development and selection of a novel tagmentation transposase mutant with improved performance and implementation of a modified buffer that reduces oxidative DNA damage during capture, both of which will be included in Illumina's upcoming TruSeq Rapid Exome Library Prep Kit. Process optimizations include full, streamlined automation, design of custom indexed library adapters that enable higher plexing and better contamination detection, and 96-library pooling prior to hybridization. .

1890F

Evaluation of DNA sample fragmentation methods for input into library construction protocols. *B. D. Craig, B. A. Marosy, K. F. Doheny.* CIDR, Johns Hopkins University, Baltimore, MD.

The Center for Inherited Disease Research (CIDR) provides high quality next-generation sequencing (NGS), genotyping and statistical genetics consultation to investigators working to discover genes that contribute to disease. CIDR currently uses a mechanical method to fragment DNA with the Covaris E210 for input into library construction workflows. Samples undergo independent customizable fragmentation events in specialized sample tubes. When processing large sample numbers, serial fragmentation adds considerable time to the overall workflow (up to 10 hours for a 96 sample plate). To reduce workflow time and cost in the lab for library construction, we evaluated the KAPA enzymatic fragmentation kit that is controlled by varying time and temperatures within the reaction. This product does not require specialized equipment or consumables and all samples in a plate will be processed simultaneously with the same reaction conditions. We evaluated reproducibility of the fragmentation events, the impact of DNA amount on the profile and compared these profiles to current Covaris programs. Additionally, CIDR receives samples in a variety of EDTA concentrations which is an inhibitor of the KAPA reaction. In order to incorporate this product as a drop-in replacement to current CIDR best practice workflows, a bead cleanup must be performed to remove all EDTA. This extra cleanup resulted in an average of 45% sample loss prior to fragmentation. Depending DNA sample number and input amount required, the additional cleanup step could require more starting DNA and offset time saved over the Covaris protocol. Samples were run on the Agilent Bioanalyzer to evaluate the size profile and optimize the reaction time to fit the current Covaris protocols. Size distributions were not affected by DNA input amounts for experiments with the same reaction time and were somewhat consistent for the replicated time and input experiments. Our results show fragmentation profiles larger than the expected mode outlined in the default protocol suggesting more optimization to the reaction time will be required to better match the profiles of our current Covaris protocols and integrate with existing methods. Future experiments involve additional evaluation of the reproducibility of fragment size profiles generated under the same reaction conditions and the evaluation of sequencing data (insert size, library complexity, and target coverage) from libraries created with the KAPA fragmented samples.

1891W

Sequencing-based, Megabase-scale haplotypes resolve the complex genomic structure of germline and primary cancer genomes. *H. P. Ji[1], B. T. Lau[2], J. M. Bell[2], S. Greer[1], S. M. Grimes[2], E. Hopmans[2], C. Wood[2], R. W. Davis[2].* 1) Division of Oncology / Department of Medicine, Stanford University School of Medicine, Stanford, CA; 2) Stanford Genome Technology Center, Stanford University School of Medicine, Palo Alto CA.

Cancer genomes contain multiple types of genetic aberrations such as point mutations, genomic deletions, copy number variants and chromosomal rearrangements. Despite advances in next generation sequencing for cancer genome analysis, it remains a significant challenge to delineate the structure of cancer genomes given the intrinsic complexity and multiplicity of somatic genomic alterations. Contiguous phasing and haplotypes involves the assignment of genetic variations such as mutations and structural variants to specific haplotype blocks and homologous chromosomes. Thus, one can resolve complex structural variants such as somatic rearrangements, identify phased mutations and quantify aneuploidy composition. However, it has been impractical to generate contiguous phasing on a Megabase (Mb) scale with current whole genome sequencing approaches. For this study, we provide our results from a first time demonstration of Mb-scale whole genome analysis applied to primary colorectal cancer genomes directly from clinical samples. This approach is rapid, efficient and potentially scalable to large population studies. We used a novel and highly efficient DNA sequencing technology that enables phasing of whole human genomes. This high-throughput microfluidic platform prepares libraries with droplet-partitioned barcodes using as little as one nanogram. A diverse collection of barcodes is individually distributed over 100,000 different partitions. Individual DNA molecules, 10s of kb in size, are identified by barcode sequence reads specific to a given partition. Based on single nucleotide variants (SNVs) in combination with barcode sequence reads, we generate Megabase-scale haplotype blocks. Subsequently, we used a novel computational approach relying on unique barcoded sequences and haplotype assignment to resolve phased structural variants and other genetic aberrations covering Mb segments. We determined Mb-scale haplotype blocks from a series of primary colorectal cancer samples in comparison to matched normal diploid DNA. Using SNVs, we generated Mb-scale haplotype blocks, determined segmental haplotype blocks that cover allelic imbalances, identified copy number variations such as deletions and resolved the structure of complex cancer rearrangements. Overall, we successfully demonstrated the feasibility, highly scalable efficiency and potential advantages of using contiguous phased haplotypes in resolving cancer genome structure and other characteristics.

1892T

A Novel Topo-Based Ligation Technology for NGS Library Preparation. *J. Zheng, Y. Li, K. Sean, G. Roma, H. Pan, C. Shi.* Molecular Cloning Laboratories, South San Francisco, CA.

Next-Generation Sequencing (NGS), a revolution in genomic science requires three major workflow steps: 1) library preparation, 2) DNA sequencing and 3) bioinformatics data analysis. It is evident that a critical step for successful sequencing is to obtain high quality DNA libraries. To meet clinical researcher's demanding needs, an ideal library preparation method should have a streamline workflow with low hands-on time, require low DNA input and result in highly complex libraries with high yields in the least amount of time. Ligation of sequencing adaptors to both ends of end-repaired fragmented DNA is akey NGS library preparation. Most commercially available Illumina library preparation kits require dA-tailing of the DNA fragments followed by adaptor ligation to achieve sequencing ready libraries. Methods relying on the use of both dA-tailing and T4 DNA ligation may result in reduced library complexity and reduced library yields. Here we introduce a novel topoisomerase-based ligation workflow for DNA library preparation, which eliminates the need for dA-tailing for enhanced ease-of-use. This workflow also reduces DNA input requirements by selectively and efficiently ligating blunt-ended DNA fragments to sequencing adaptors while simultaneously preventing adaptor dimer formation from forming. This rapid method supports whole genome, amplicon, and RNA sequencing applications. Results show enhanced coverage and reduced GC bias for human genome DNA sequencing.

1893F

Massively parallel single nucleotide mutagenesis using reversibly-terminated inosine. *G. Haller[1], D. Alvarado[1], M. Dobbs[1,4], C. Gurnett[1,2,3].* 1) Department of Orthopaedic Surgery, Washington University School of Medicine, Saint Louis, MO; 2) Department of Neurology, Washington University School of Medicine, Saint Louis, MO; 3) Department of Pediatrics, Washington University School of Medicine, Saint Louis, MO; 4) Shriners Hospital for Children, St. Louis, MO.

Accurate, inexpensive and efficient methods of large scale mutagenesis of specific DNA sequences have the potential to alloware needed to rapidly assess the functional effects of genetic variation. for rapid assessment of the effects of single nucleotide changes on protein function. However, many currentThese methods to generate mutational libraries are remain limited by the need to synthesize oligonucleotides *en masse* to clone directly or to use as templates for mutagenesis. Here we demonstrate the construction of a systematic allelic series (SAS) with, an oligo-free method of massively parallel single nucleotide mutagenesis using that utilizes reversibly-terminated deoxyinosine triphosphates (dITP). We created mutational libraries containing every single nucleotide mutation within the N-propeptide of the type I (I) collagen gene *COL1A1* and within the full coding region of the *b-lactamase* gene (>1kb) which confers resistance to ampicillin, each in less than a day in a single reaction for <$30. Sequencinga 1kb SAS library revealed that, >90% of molecules harbored one and only one mutation, making library construction by this method extremely efficient and cost-effective. SAS-based mutational screening has the ability to greatly expedite interpretation of human disease genetic variantions, *in vitro* protein evolution, development of biopharmaceuticals and the determination of critical domains for proper protein function.

1894W

Pushing the Limits of Small RNA-Seq with Chemically Modified Adapters. *S. Shore, J. M. Henderson, A. P. McCaffrey, G. Zon, R. I. Hogrefe.* Research & Development, TriLink BioTechnologies, San Diego, CA.

Next-generation sequencing (NGS) is an evolving tool for nucleic acid sequence identification. NGS is widely used for basic research and discovery as well as clinical diagnostics. It is critical to develop the most accurate and reproducible library preparation techniques, especially when dealing with patient samples. Both DNA and long RNA-Seq have advanced, low input, PCR-free, and automatable library preparation techniques. In the continuously growing field of small RNA (including microRNA, piwiRNA, tRNA, and snoRNA) sequencing methods still have limitations including high input requirements and difficulties in automation. Separation of tagged library from adapter-adapter ligation products (adapter dimer) has been a persistent challenge due to the minimal size difference between products, which hinders a gel-free workflow. In order to resolve this problem we have screened several chemical modifications of adapter oligos to suppress adapter dimer formation. We introduce a novel method of small RNA library preparation that applies these modified CleanTag™ Adapters and creates the potential for automation without the need for a gel separation to exclude adapter dimer. We replace the gel purification step with a two-step magnetic bead purification for size selection since adapter dimer no longer dominates the reaction. This could be carried out by a liquid handling robot for a fully automated library preparation procedure. Current commercial library preparation kits recommend inputs of 100-1000 ngs. Reduction in adapter dimer has allowed for library preparation from much less starting material (as low as 1 ng). Using 10 ngs of total brain RNA input in conjunction with CleanTag™ Adapters, adapter dimers were reduced by over 3,000,000 reads compared to a commercial kit. Additionally, libraries using modified adapters resulted in 1,000,000 more miRNA reads than a commercial kit at this low RNA input. With increased library coverage and decreased side products the data generated from modified adapters brings us closer to analysis on a single cell level. Our method of small RNA library preparation results in high quality libraries from various tissues, cells, and low RNA input biofluids (*e. g.* plasma and urine). The ability to use very low inputs and to automate library preparation will allow high throughput sequencing and improved clinical assays.

1895T

Improving the information content of long fragment read technology by expansion to 5,000 compartments. *O. Wang[1,2], Y. Xie[1,3], W. Huang[1], C. Chang[1], J. Tang[1], X. Cheng[1], Y. Fu[1], W. Zhang[1], X. Liu[1], H. Jiang[1,2], B. A. Peters[1,4], R. Drmanac[4], X. Xu[1].* 1) BGI-Shenzhen, Shenzhen, Guangdong, China; 2) Department of Biology, University of Copenhagen, Copenhagen, Denmark; 3) School of Bioscience and Bioengineering, South China University of Technology, Guangzhou, China; 4) Complete Genomics Inc. , Mountain View, California, USA.

Individual humans genomes are diploid in nature, with half of the homologous chromosomes being derived from each parent and varying at millions of positions. How these variants act upon each other within each parental chromosome can have profound effects on gene expression and ultimately the health of an individual. As human genome sequencing moves towards clinical applications, determining which sets of variants are transferred in blocks from each parent (Haplotypes) is increasingly being recognized as critically important. Haplotype information can be generated either by computational approaches that rely on population based haplotype information, sequencing both parents or by laboratory-based experimental methods. Computational methods are cheap and easy, but have lower accuracy the farther apart variants are found and cannot place rare variants or de novo mutations into haplotypes; whereas experimental methods can potentially make up for all of these shortcomings. Almost all of them rely on dilution and compartmentalization of long DNA fragments such that individual compartments contain less than 20% of a haploid genome. This results in a statistically low probably of overlapping fragments from the maternal and paternal chromosomes in any given compartment. This strategy was first proposed by Drmanac and ultimately led to the development of long fragment read (LFR) technology; a process which can start from as little as 10 cells, provides excellent performance in phasing, and the highest level of accuracy in SNP calling. However, the current LFR is based on using 384 compartments and there would still be up to 20% of a haploid genome in each well. Increasing the number of compartments and thus lowering the fraction of a genome per well results in less overlap between parental fragments and fragments containing overlapping repeat sequences. In this study we sought to expand the current 384 compartments to 5184. This allows us to decrease the haploid genome content to about 0. 1% per well; significantly reducing the possibility of overlapping of long fragments. To achieve this number of compartments required the development of a novel combinatorial barcode ligation approach. With these modifications, we have generated a more informative library with a very low error rate and demonstrate how this type of data might be used in a clinical setting.

1896F

Single Tube, Whole Genome Phasing and Assembly using Bead-based Index Partitioning. *F. Zhang[1], L. Christiansen[1], R. Jackson[2], N. Morrell[2], N. Gormley[2], J. Shendure[3], K. L. Gunderson[1], F. J. Steemers[1].* 1) Advanced Research Group, Illumina, San Diego, CA, USA; 2) Technology Development Department, Illumina, Little Chesterford, Essex, UK; 3) Department of Genome Sciences, University of Washington, Seattle, WA, USA.

The ability to accurately phase and assemble genomes is dependent on long range sequence information. Here we show a sequencing library preparation method that effectively generates long indexed-linked reads of 30-50 kb in length in which each DNA molecule is converted to multiple short libraries with the same index. The libraries labeled with the same index share genomic proximity information and therefore infer the long haplotype structure. We show for the first time that index partitioning of DNA molecules (and parental copies) can be performed in a single tube without the need to use many physical compartments in order to separate, index and sequence the individual molecules. The key concept of the approach is based on the macroscopic transfer of indices attached to a single clonal indexed bead to a long DNA molecule. As a DNA molecule diffuses and binds to clonal indexed immobilized transposomes on a single bead, indices transfer through tagmentation to the DNA molecule. Because intra-tagmentation and /binding reactions on beads are significantly faster than intermolecular reactions between beads, individual DNA molecules are clonally labeled, while simultaneously labeling many long DNA molecules in the population with different i. e. , 96-100,000+ clonal indexed beads in a single tube. As such, we have greatly simplified the process of individual parallel indexed library preparations of fractions of the genome. As a proof-of-concept, we apply this method to phase over 99% of heterozygous SNPs from a HapMap sample into long, accurate haplotype blocks with N50 block sizes >1. 2 Mb using only 88 Gb of extra sequencing over the reference unphased VCF file. Additionally, we demonstrate utility of contiguity-preserving transposition (CPT-seq [1]) on beads in assembly applications greatly improving the N50's [2] and ability to detect structural variants, including gene-fusions, indels, (compound heterozygous) deletions, and paralogs. These technologies provide a simple, rapid, scalable and highly automatable route towards accurate haplotype-resolved sequencing and assembly of the genome[1] Amini, S. et al. Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. Nat. Genet. 46, 1343–1349 (2014). [2] Adey, A. et al. In vitro, long-range sequence information for de novo genome assembly via transposase contiguity. Genome Res. 24, 2041 (2014).

1897W

Sequencing the B-cell and T-cell repertoire. *M. Karaca[1], J. N. H. Stern[2], G. Yaari[3,4], J. A. Vander Heiden[5], S. H. Kleinstein[3,5], D. A. Hafler[2,6], K. C. O'Connor[2], A. W. Briggs[7], T. Gilbert[7], C. Clouser[7], W. Donahue[7], L. M. Apone[1], F. J. Stewart[1], S. Russello[1], T. B. Davis[1], E. Dimalanta[1].* 1) New England Biolabs Inc. , Ipswich, MA; 2) Department of Neurology, Yale School of Medicine, New Haven, CT; 3) Department of Pathology, Yale School of Medicine, New Haven, CT; 4) Bioengineering Program, Faculty of Engineering, Bar-Ilan University, Ramat Gan, Israel; 5) Interdepartmental Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT; 6) Department of Immunobiology, Yale School of Medicine, New Haven, CT; 7) AbVitro Inc, Boston, MA.

Immune sequencing, which allows for the study of complex immunological diseases by sequencing millions of V(D)J combinations from B-cell antibody and T-cell receptors, has gained popularity due to recent throughput and read length improvements in next-generation sequencing (NGS) technologies. However, structural and sequence complexities of antibody genes have made reliable targeting approaches challenging. We have developed and optimized a method for accurate sequencing of full-length immune gene repertoires of B-cells and T-cells. This allows for exhaustive somatic mutation profiling across complete V, D and J segments, full isotype information analysis (IgM, IgD, IgG, IgA and IgE), and the possibility for synthesis and expression of complete antibody chains for downstream immunological assays. The method uses a unique barcoding scheme specifically designed to tag every mRNA molecule so that all PCR copies of each mRNA fragment can be collapsed into a single consensus sequence, making the assay extremely accurate by resolving PCR bias and sequencing errors, as well as allowing quantitative digital molecule counting. One example of how our sequencing technology has been applied along with advanced bioinformatics to gain important biological insights is in a recent study exploring B cell traffic patterns in multiple sclerosis (MS), an inflammatory disease of the central nervous system (CNS)1. This work demonstrated that B cells traffic freely across the CNS tissue barrier, with a majority of B cell maturation occurring within secondary lymphoid tissue outside of the CNS. These findings identify a pathway required for B cell accumulation in the CNS and may aid in designing better therapeutics to combat this debilitating disease. 1J. N. H. Stern, G. Yaari, J. A. Vander Heiden, G. Church, W. F. Donahue, R. Q. Hintzen, A. J. Huttner, J. D. Laman, R. M. Nagra, A. Nylander, D. Pitt, S. Ramanan, B. A. Siddiqui, F. Vigneault, S. H. Kleinstein, D. A. Hafler, K. C. O'Connor, B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. *Sci. Transl. Med.* **6**, 248ra107 (2014).

1898T

High quality library construction and reliable quantitation with NEBNext reagents. *E. Yigit, P. Liu, N. Tanner, J. Borgaro, L. Apone, V. Panchapakesa, D. Rodriguez, B. Langhorst, J. Menin, C. Sumner, C. Chater, F. Stewart, N. Nichols, E. Dimalanta, T. Davis.* New England Biolabs, Ipswich, MA.

The field of NGS has matured significantly over the past few years. While the use of NGS in the clinic was simply a possibility a few years ago, today it is routinely used in some areas of diagnostics. An expanded role for NGS in the clinic and research will depend on continued improvement of the upstream processes required to produce high quality NGS data. In particular, maximizing data output and minimizing instrument run failure are imperative. In this poster, we present data showing: 1) the improvements we've made in library preparation with the development of the NEBNext Ultra II Library Prep kit; and 2) the development of the NEBNext Library Quant Kit for Illumina, a simple and robust method for quantitation of Illumina libraries. In the first part, we show that libraries made with the Ultra II kit have low DNA input requirements; significantly improved library yields and reduced sequence bias. In addition, the workflow is simple and streamlined, greatly reducing the time required to produce high quality DNA libraries and the possibility of errors. In the second part, we demonstrate the effectiveness of the NEBNext Library Quant Kit for a broad range of library types and sizes as well as advantages offered by qPCR quantitation for obtaining optimal cluster density and user-to-user consistency. The NEBNext Quant Kit offers an efficient and cost-effective qPCR library quantitation workflow for users looking to optimize both sequencing yield and throughput.

1899F

A recombination-based technology in *Escherichia coli* for cloning large and specific human DNA sequences. *L. Brunelli, H. Shanmugam, G. T. Lyozin.* University of Utah School of Medicine, Salt Lake City, UT.

In addition to being driven by DNA sequencing and genome engineering, the Precision Medicine Initiative recently unveiled by the National Institutes of Health requires the complete functional analysis of the human genome. However, such comprehensive characterization remains a major challenge due to hundreds of thousands of poorly understood distal transcriptional regulatory elements in the human genome. To overcome this barrier, further advances in *in vivo* DNA cloning and engineering are urgently needed. We recently developed one of the few approaches for the precise, seamless engineering of bacterial artificial chromosomes (BACs) and P1 phage–derived artificial chromosomes (PACs), two cloning systems particularly suited for the analysis of mammalian genes (Lyozin G. T. et al. *Nat Methods* 2014). Nonetheless, current BAC and PAC library recombinants have on average <200 kb inserts with random boundaries. This limitation prevents their use in the analysis of specific regulatory variants distributed over large chromosomal regions and involved in human disease as is the case for *GATA4* in congenital diaphragmatic hernia (CDH), a severe structural birth defect (Arrington C. B. et al. *Am J Med Genet* 2012). Here we report the development of pP1RV2, a new high-capacity PAC-derived vector that can be used for seamless megabase-range gap repair DNA cloning and engineering. To show its applicability, we separately cloned two 190 kb BAC inserts with precise boundaries from overlapping BACs in the *GATA4* region. Cloning efficiency was ~30%, remarkably better than previously reported for high-copy number vectors such as pBR322 (~2% for an 80 kb insert, Lee E. C. et al. *Genomics* 2001; Copeland N. G. et al. *Nat Rev Genet* 2001). To our knowledge this is the first time that a high-capacity vector has been used in gap repair DNA cloning, allowing the precise manipulation of large and specific DNA segments in *E. coli*. Our long-term goal is to reconstruct a 1 Mb *GATA4* regulatory region relevant to CDH on pP1RV2. A technology for seamlessly engineering large, specific DNA sequences is essential to complete the functional resolution of the human genome and streamline the efforts of the Precision Medicine Initiative.

1900W

Unique molecular indexes to remove PCR bias in bisulfite sequencing. *I. lee[1], D. Wu[1], J. Lu[2], A. Lejeune[2], W. Timp[1].* 1) Biomedical Engineering, Johns Hopkins School of Medicine, Baltimore, MD; 2) Biomedical Engineering, Johns Hopkins University, Baltimore, MD.

Bisulfite sequencing is a high-throughput technique to measure 5mC DNA methylation, in which unmethylated cytosine residues are converted to uracil, affectively amplifying as thymine residues. This process exaggerates the variability in the DNA library's GC content, a known cause of bias in PCR amplification prior to next-generation sequencing. Such bias is a significant obstacle in sequencing analysis, precluding quantification of the observations and always requiring controls. We propose the addition of molecular barcodes into sequencing library preparation to remove the GC-content bias and increase the accuracy of methylation calls. Unmethylated *E. coli* DNA was treated with M. SssI CpG methyltransferase to produce a stock of DNA methylated at all C's in CpG contexts. Then, samples containing varying ratios of methylated and unmethylated DNA were prepared to assess the potential of molecular barcodes in decreasing GC-content bias. Standard bisulfite conversion and library preparation procedures (Pico Methyl-Seq™ or NEBNext® Ultra™) were used to prepare the libraries. Molecular barcodes were incorporated in the form of custom-made adaptors containing six to eight randomized nucleotide residues. The sequenced reads were analyzed with Bismark bisulfite mapping program. The reference location of each mapped read, its sequence, and/or its corresponding molecular barcode were used to identify PCR duplicates. For each set of duplicates, the best read, with corrections in methylation calls, was selected as the representative read. The methylation analysis resulted in universal but uncorrelated error in overall methylation, deviating up to 10 % from their original percentages. Removal of duplicates without molecular barcodes decreased the error to 7 %. Incorporation of molecular barcodes in the duplicate removal further decreased the error to as low as 2 %. Simulation of low-input condition by preparing the library with large number of amplification cycles showed increased resemblance of high-input results with molecular barcodes. These results indicate that molecular barcodes can be used to remove bias in sequencing introduced by PCR duplicates. In this study, molecular barcodes were used for bisulfite sequencing; however, this method can be applied in any NGS setting, especially when the input amount is low and the samples must undergo a large number of amplification cycles, as is the case in many clinical experiments.

1901T

Rapid extraction of high yield, high quality DNA from tissue samples. *D. O'Neil, S. Schroeer, M. Sprenger-Haussels.* QIAGEN GmbH, Hilden, NRW, Germany.

Genetic and genomic analysis from tissue samplesrequires the extraction ofhigh quality DNA. Mechanical disruption methods such as bead millingprovide high yield from tissue samples, but cause damage to thenucleic acids. Purely enyzmatic methods such as proteinase K digestion canextract nucleic acid without damage, but require long incubation times, often proceeding overnight, and without approaching theyields achieved by mechanical disruption techniques. Thus a method is needed which can provide a rapid extraction of high yield, high quality DNA from tissue samples. We have developed a method which combines mechanical, chemical, and enyzmatic disruption of tissue samples to provide rapid extraction of high yield DNA without causing significant damage to the nucleic acid. Mechanical disruption is achieved by vortexing the tissue in a container with a specially shaped milling bead. The lower power of a vortexer compared to a bead mill prevents shearing of the DNA. Low chaotropic conditions support a mild lysis of the sample and provide a synergistic effect for the simultaneous proteinase K digestion and mechanical shearing. DNA extracted from fresh-frozen and from stabilized tissue samples using the new method showed up to 10 fold increase in yield compared to a standard proteinase K digestion or a mechanical milling method alone. The entire disruption procedure was complete in 15 minutes for softer tissues such as liver or kidney, and within 30 minutes for tougher tissue such as stabilized lung or trachea samples. DNA from fatty tissues such as brain as well as fibrous tissues like muscle was extracted efficiently by this method. The new method provides extraction ofDNA from tissue samples superior to existing methods, on the criteria of DNA quality, yield, and time required for extraction. Technologies such as next-generation sequencing make the extraction and analysis of DNA from tissue samples more important, and a method which provide the most convenient and efficient extraction of DNA from tissue will support theroutine genomic analysis enabled by these technologies.

1902F

Pan-microbial detection using Axiom® Genotyping Solution from Affymetrix. *P. Rack[1], K. McLoughlin[2], S. Gardner[2], M. Mittmann[1], A. Pirani[1], C. Sheppy[1], C. Jaing[2], L. Bellon[1], T. Slezak[2], M. Shapero[1].* 1) Affymetrix, Inc. , 3420 Central Expressway Santa Clara, CA 95051; 2) Lawrence Livermore National Laboratories, Livermore, CA 94551.

BackgroundDifferent human body sites harbor distinct, diverse microbial populations, including eukaryotes, archaea, bacteria, and viruses, which define the human microbiome. Major efforts are underway to describe the establishment and flux of these dynamic microbial populations, with the goal of elucidating links to human health and disease. With the expansion of microbial genome sequence information, microarrays are well-positioned to capitalize on this evolving information content. Here we present an overview of Axiom® Genotyping Solution [mgb1]and a description of the design, development, and testing of a microbial detection array (MDA). **Experimental design** Axiom® Microbial Detection Array contains probe sequences targeted towards more than 12,000 organisms, using 133,500 target sequences. Probe sequences were designed with the guidance of data derived from proof-of-concept studies that informed probe parameters including length and tolerance to mismatches. This design enables species/strain-level resolution of the entities comprising a sample using a Composite Likelihood Maximization algorithm (CLiMax). To investigate models for relative quantitation, Langmuir adsorption curves were fitted using Latin-square data generated by target spikes of defined concentrations. Furthermore, defined samples of increasing complexity were used to interrogate and measure array sensitivity and specificity. Lastly, the Axiom® MDA was tested for the ability to identify and quantify microorganisms within complex biological samples, and results were compared to orthogonal approaches. **Conclusions**In summary, Axiom Microbial Detection Array offers a solution for the identification and enumeration of microbial entities comprising complex biological samples. The unbiased design of the Axiom MDA is agnostic of sample type, lending to its use and utility in the context of bioterrorism, environmental monitoring, food safety, and human and animal health. Furthermore, scalable sample throughput enabled by 24-, 96-, and 384-array layouts coupled with laboratory automation allows processing of tens to thousands of samples per week with minimal manual intervention and is consistent with the need for cost-effective detection and examination of microbiomes.

1903W

Automation of Micro RNA and Total RNA Purification from FFPE using the Beckman Coulter Agencourt FormaPure Kits and Biomek liquid handler. *B. Lee, J. Palys.* Beckman Coulter, Brea, CA.

Micro RNAs are small, naturally-occurring non-coding ribonucleic acids with sizes between 18 and 40 nucleotides (nts) that have been demonstrated to play a significant role in the regulation of gene expression. As a result, interest in smaller RNA species such as miRNA has increased. This poster describes the purification of miRNA and total RNA from challenged Formalin Fixed Paraffin Embedded (FFPE) samples using the Beckman Coulter SPRI (Solid Phase Reverse Immobilization) magnetic bead based chemistry and the Biomek automated extraction method. The FormaPure RNA 96 demonstrated method enables automated purification of total RNA, including miRNA and other small RNAs, from 1–96 samples on a Biomek Span 8 workstation. Total RNA and miRNA can be purified from very small amounts of FFPE samples using the LCM procedure. The Biomek automated SPRI method is an easy, high yielding and robust nucleic acid purification process that does not require centrifugation and vacuum filtration steps. Purified nucleic acids are easily eluted from the magnetic beads under aqueous conditions, which provide maximum flexibility for downstream applications. The data shows that the samples extracted using the FormaPure kit gave higher RNA yield, as compared to samples extracted from a column purification method.

1904T

Targeted SMRT® Sequencing and Alternative Splice Detection of cDNA using Roche NimbleGen SeqCap Enrichment. *D. Raterman[1], L. Feng[1], M. Brockman[1], Q. Meng[2], M. Wang[2], H. V. Doddapaneni[2], A. C. English[2], K. Walker[2], Y. Han[2], A. Roy[3,4], D. W. Parsons[2,4,5], D. M. Muzny[2], R. A. Gibbs[2,5], T. Hon[6], E. Tseng[6], T. Clark[6], T. Richmond[1], D. Burgess[1].* 1) Roche NimbleGen, Madison, WI; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030; 3) Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX 77030; 4) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030; 5) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; 6) Pacific Biosciences, Menlo Park, CA.

It has been estimated that between 74% and 100% of human multi-exon genes are alternatively spliced. Alternative splicing gives cells the ability to create different protein isoforms with only a single gene and is a key factor contributing to increased cellular and functional complexity in higher eukaryotes. Current RNA sequencing protocols rely on high-throughput sequencing of short reads. However with an average transcript length in the range of 1. 5-2. 0 kb, advanced analysis tools are required to reconstruct full-length transcripts, making accurate annotation of each gene locus and transcript isoform quantitation challenging. Longer reads are expected to improve performance for transcript assembly, transcript quantification and gene fusion detection. With the current advances in sequencing technologies, it is now feasible to sequence full-length cDNAs. It is not efficient, however, to sequence whole RNA-seq libraries. Most tissues express between ~6,000-10,000 mRNA genes and the vast majority of transcripts are derived from a very small subset of genes. These highly expressed genes predominate most RNA-seq experiments and make detecting rare variants difficult. Our research details a novel method to target specific large fragment cDNAs using Roche NimbleGen capture probes for sequencing on the long-read sequencing platform from Pacific Biosciences. Only transcripts from genes of interest are captured and sequenced, greatly increasing the discovery power of a given experiment. cDNA was prepared from a brain sample and then hybridized using Roche NimbleGen's Neurology Panel Design. SMRTbell™ adapters were then ligated onto the captured products and sequenced on a PacBio® RS II machine. In addition, control samples that were not enriched were sequenced. Average on-target rates were 77% compared to the control samples where on-target rates were <7%. Further R&D efforts will focus on extending the utility of this novel method to the entire coding region, making it applicable to a broad spectrum of diseases. Full-length mRNA transcripts from patients with cancer will be enriched by HGSC VCRome exome probe set (~42 Mbp capture size) for transcriptome analysis. With unique advantages provided by long-read sequencing and target capture, we believe the new method will produce unprecedented comprehensive alternative transcript splicing, novel fusion structures, as well as mRNA nucleotide variants.

1905F

Improved sequencing and transcript detection utilizing an RNase H-mediated ribosomal RNA depletion method. *D. Sanjai[1], J. Pavlica[1], R. Wadsworth[2], M. Ranik[2], P. Jones[2], N. Hapshe[1], R. W. Kasinskas[1], E. van der Walt[2].* 1) Kapa Biosystems Inc, Wilmington, MA; 2) Kapa Biosystems (Pty) Ltd, Cape Town, South Africa.

As transcriptome sequencing becomes more ubiquitous, the development of new methods is needed to optimize library preparation, and thus achieve high-quality sequencing results. Ribosomal RNA (rRNA) accounts for the vast majority of total RNA, but contributes minimally to understanding gene expression. Therefore, efficient removal of rRNA is a critical step in RNA-Seq library preparation to maximize the yield of less abundant, biologically-informative, transcriptome-derived reads. In this study, we present an alternative method for rRNA depletion, which retains more precursor and long noncoding RNA species than a capture-mediated method, giving a more comprehensive view of the transcriptome. **Methods:** To address the needs of this growing field, we have developed the KAPA Stranded RNA-Seq Kit with RiboErase, which includes a highly effective workflow for the targeted enzymatic depletion of cytoplasmic and mitochondrial rRNA through the use of complementary oligonucleotide probes and RNase H. Our method is compatible with human, mouse, rat, and other related species, and is automation-friendly. In this study, we sequence Universal Human Reference RNA, comparing our RNase H-mediated method against a bead-based capture-mediated rRNA depletion technology. KAPA Stranded RNA-Seq Kit with RiboErase results show more efficient depletion of rRNA (up to 99. 9%), lower duplication rates, increased gene and transcript detection, and better coverage of GC-rich and low abundance transcripts. Moreover, our workflow demonstrates improved detection and coverage of precursor and long noncoding transcripts, which were better correlated to an undepleted total RNA control. By utilizing a targeted enzymatic method for depletion, our workflow enables superior reduction of rRNA and a more complete representation of the transcriptome, including precursor and long non-coding RNAs. Along with improved sequencing performance, the KAPA Stranded RNA-Seq Kit with RiboErase has been proven to be highly robust and reproducible across varying input amounts. Through the use of evolved enzymes, including the KAPA HiFi DNA polymerase, high-quality reagents, and an optimized “with bead” workflow, the kit has demonstrated improved coverage of difficult transcripts, including those that are GC rich and in low abundance.

1906W

Isolation and processing of 1000s of single cells for genomics applications using Nanowell Arrays. *M. Srinivasan, J. Dunne, R. Acob, S. Anandakrishnan, L. Chan, P. Espinoza, H. Hubschle, M. Ho, S. Husain, P. Lin, I. Mann, A. Mir, T. Schaal, H. Shapiro, M. Slater, K. Swaminathan.* WaferGen Biosystems, Fremont, CA.

Single Cell RNA Sequencing has quickly become an established method to understand cell specific differences and how those differences manifest both temporally and spatially in various tissues. Despite the immense progress, there is continued demand for new technologies that can miniaturize and automate the isolation of thousands of individual cells and process these individual cells for nucleic-acid sequencing or qPCR. A 5184-well nanoarray called the SmartChip and a dispenser that can dispense nanoliter volumes accurately into the nanowells are used to dispense cells into nanowells and process the cells for RNA-seq. In one variation, the Single Cell RNA Barcoding and Sequencing (SCRB-seq, Soumillion et. al. , http://dx. doi. org/10. 1101/003236 2014) method has been adapted to the SmartChip for 3’ tag counting applications. Furthermore, an image-analysis software called Cellector was designed, which enables the selective processing of individual wells based on a researcher’s needs - a feature that can minimize data analysis noise from multiple cell containing wells. Using Poisson dispense method where on average 1 cell is deposited into each well, Cellector routinely identifies about 1500 individual cells that are viable as judged by staining for DNA and dead cells using Hoechst 33342 dye and propidium iodide, respectively. The output from the image analysis software is uploaded to the dispenser and the dispenser adds reagents only to wells that are either manually and/or automatically selected. To maximize the utilization of the SmartChip real estate and decrease the cost/cell we have also developed a targeted deposition method to deposit cells at >75% occupancy rate. We present data to demonstrate the reproducible dispensing of individual cells in Poisson and targeted dispense methods. In both approaches, individual cells were barcoded with unique molecular identifiers, by adapting SCRB-Seq, to estimate differential gene expression. System simplicity, ease of use, number of sample addressable wells, system throughput, reagent flexibility and cell type compatibility make SmartChip an ideal candidate system for use in a wide variety of single cell genomic applications.

1907T

High-Throughput, High-Precision, Single-Cell Gene Expression Analysis Using a New BD FACS™ Cell Sorter. *X. Wang, J. Goglio, K. Dembski, T. Le, P. P. V. K. , S. Ghanekar.* R&D, BD Biosciences, San Jose, CA.

Gene expression studies performed on bulk samples may obscure the understanding of complex samples. Gene expression analyses performed on single cells, however, can offer a powerful method to resolve sample heterogeneity and reveal hidden biology. Previously, we demonstrated that BD's fluorescence activated cell sorting (FACS™) technology could accurately target and isolate single cells or a small number of cells for use in next generation sequencing (NGS) or transcriptome analyses. Gene expression data from the individual cells correlated strongly with the cell phenotypes used for isolating the cells. Here, we demonstrate applications of a new, entry level BD FACS cell sorter that rapidly selects and isolates single cells or cell populations, with high precision, for molecular analysis. The sorter employs innovative technologies recently developed by BD to automate many operating steps that are traditionally performed manually, making fast and easy cell selection and isolation possible for more researchers. The sorter has one laser for illumination and four parameters for detection (one composite scatter channel and three fluorescent channels), and collects cells into plates or tubes. In this study, we demonstrated the applications for: 1) elimination of dead cells from collection by positive propidium iodide (PI) staining; 2) isolation of single cells from different cell cycle phases (G0/G1, S, G2) with Hoechst dye staining; 3) accuracy of sorting cells by depositing different numbers of cells (1 cell, 5 cells, 10 cells, 20 cells, and 50 cells) per well in a 96-well plate; 4) isolation of single cells from a mixed cancer cell sample based on their specific cell surface marker staining. The gene expression of the sorted cells was analyzed, and gene expression profiles were generated. These studies demonstrate a simple, fast method using a new, easy-to-use sorter to select and isolate single cells for high-resolution genomic studies, including next generation sequencing, qPCR, and microarray analysis. It provides a powerful tool for studying complex samples at the individual cell level.

1908F

Integrated DNA and RNA sequencing of the same cell. *J. He.* Department of Biology, South University of Science and Technology of China, Shenzhen, Guangdong, China.

Interest in single-cell whole genome analysis is growing rapidly, especially for profiling rare or heterogeneous populations of cells. Various amplification protocols have been developed to sequence the genome or transcriptome of single cells. However, the current single-cell RNA-seq method or DNA sequencing method do not permit analysis of the transcriptome and genomic sequence of individual cells simultaneously. In this study, we developed a strategy that allows us to use next-generation sequencing to simultaneously obtain both full genome and transcriptome information from an individual cell. Using mouse oocytes as an example, we demonstrated that the integrated DNA and RNA sequencing of single oocytes can deliver both high coverage of genome and reproducible low-bias transcriptome. Our technology will help to fulfill the central goal of biology and medicine, which is to connect the genotype and phenotype of individual cells under physiological or pathological conditions.

1909W

Novel Luminex assay for Telomere Length measurement does not show Well Position Effects like qPCR. *M. G. Kibriya, F. Jasmine, S. Roy, H. Ahsan, B. L. Pierce.* Public Health Sciences, The University of Chicago, Chicago, IL.

Background: Telomere length (TL) is a potential biomarker of aging and short TL is associated with increased cancer risk. In large scale studies, qPCR is typically used for TL measurement, primarily due to its low cost and low DNA input requirements. Recent study showed that the well position of the sample in the qPCR plate impacts the TL measurement. Recently we have developed a novel non-PCR amplification, probe-based assay with signal amplification on Luminex platform for TL requiring ~50ng DNA. Here, we report for the first time a comparison of TL measures obtained from two single-plex (SP) qPCR, one multiplex (MP) qPCR, and our novel Luminex (LUM) assay for TL, with emphasis on sample positioning. **Material and Methods:** For qPCR, DNA from two individuals (A and B) were placed in 48 wells – DNA from person-A in the upper left (24 wells) and lower right quadrant (24 wells) and DNA from B in other two quadrants of BioRad CFX96. We conducted two types of SP-qPCR assays, using two different primer sets. For each SP-qPCR assay, two plates (one for Telomere and one for reference gene) were used. For the MP-qPCR and LUM assay, the telomere and the reference genes were assayed from the same well. In LUM, 24 wells were used for standards, and thus DNA from the same two individuals was replicated in 36 wells. **Results:** Coefficient of variation (CV) of TL for A-DNA was 8. 9%, 14. 9%, 24. 3% and 8. 4% for SP-qPCR set1, SP-qPCR set2, MP-qPCR and LUM respectively. Similarly CV for B-DNA was 11. 4%, 13. 8%, 19. 7% and 7. 2%. Irrespective of person, samples in the left and right most columns showed statistically lower TL (ANOVA, p=3. 4x10E-6, p=0. 012, and p=0. 02 for SP-qPCR set1, SP-qPCR set2, and the MP-qPCR assays respectively), but not in the novel LUM assay (p=0. 83). Similarly, samples in the rows of the upper and lower edges also showed slightly lower TL using SP-qPCR set1 and the MP-qPCR assays (ANOVA, p=0. 01, p=0. 001, respectively), but not in SP-qPCR set 2 (p=0. 42), or in the novel LUM assay (p=0. 96). For SP-qPCRset1, 28. 2% of the variation in TL was explained by column-to-column variation and 28. 9% was explained by person-to-person variation where as for the LUM assay only 2. 65% of the variation in TL was explained by column-to column variation and 52. 2% was explained by person to person variation. **Conclusion:** Our novel Luminex assay for TL had good precision and did not show positional effect of the sample that was seen with all the three qPCR assays tested.

1910T

Fusion Gene Detection and Gene Expression Analysis of Circulating RNA in Plasma. *J. Gu, K. Lea, X. Fang, H. Saunders, K. Bramlett.* Thermo Fisher Scientific, Austin, TX.

The presence of circulating (cell-free) nucleic acids in the bloodstream of individuals with cancer offers a potential non-invasive approach to monitor disease status and guide treatment options. In the past years, increasing interests have been shown on circulating RNA, especially circulating small RNAs for their application as biomarkers for potential toward more effective diagnosis and prognosis. However, widespread inconsistencies have been observed among the studies due to biases generated during sample collection, handling, RNA extraction and analysis. We developed a complete workflow that includes blood collection, plasma preparation, circulating RNA extraction, followed by expression analysis and gene fusion detection on Ion TorrentTM Next-Generation Sequencing platforms. Blood plasma samples from 2 healthy donors and 4 individuals diagnosed with Non-Small Cell Lung Cancer were utilized for circulating RNA isolation following a TRIzol® LS Reagent and mirVanaTM miRNA Isolation Kit-based method to maximize circulating RNA recovery. AmpliSeq library preparation was performed on purified circulating RNA using either AmpliSeqTM Transcriptome panel for expression profiling of 21K coding and non-coding genes, or an Ion AmpliSeq™ panel targeting fusion transcript detection from RNA. AmpliSeq transcriptome data was analyzed using ampliSeqRNA plugin in Torrent SuiteTM Software. ~3000 genes were detected in cfDNA from plasma samples with high correlation (r>0. 8) observed between healthy donors. Ion ReporterTM Software was used to analyze fusion transcript panel data. Detection of fusion gene transcripts was demonstrated by spiking trace amount of RNA from fusion positive cell line into circulating RNA from healthy donor, indicating high sensitivity of the detection system. In summary, this study demonstrated the feasibility of gene expression profiling and gene fusion detection from circulating RNA in plasma samples on Ion TorrentTM NGS platforms.

1911F

FlightDeck: a self-service web-portal for reproducible genomics research in the Cloud via Docker container. *J. Kim[1], T. Bath[1], C. Farcas[1], O. Harismendy[1], A. Koures[1,2], S. Ko[1], L. Ohno-Machado[1,2].* 1) Department of Biomedical Informatics, University of California San Diego, La Jolla, CA; 2) Clinical and Translational Research Institute, University of California San Diego, La Jolla, CA.

With increasing size of data and analysis complexity in genomics, the community relies heavily on robust software. Yet, there are several challenges for reproducible research: 1) missing or obsolete source code, 2) undocumented or unexpected dependencies to install and configure applications, 3) undisclosed values of the parameters used in published analyses, and 4) requirements for querying and pre-processing external reference datasets. To address these issues, we present FlightDeck, a self-service web-portal that encapsulates scientific software as Virtual Machines (VMs) and enables automated VM provisioning and configuration based on published workflows and recipes. With just a web-browser, users can upload experimental data, search a recipe of interest, provision a VM instance, start the analysis, and generate reproducible results. FlightDeck is different to existing projects such as Galaxy or GenePattern in that it allows users to run both prebuilt and user-defined custom workflow. A recipe includes a version of the source code, dependent libraries, and external reference data. Recipe authors also provide the VM specifications to ensure fully automated deployment. High demand VMs, such as alignment or variant calling are pre-provisioned via Docker and are available as VM images, reducing the total time to generate data. After uploading input experimental data to FlightDeck, users can run multiple recipes that apply different softwares/workflows and varied parameter values on the same data for a comparison. Software developers who want to focus on inventing a new algorithm, still have to go through a lengthy process of example data preparation, configuration and running existing softwares for an evaluation. FlightDeck makes all these steps as one-time-job by turning into a recipe and VM-snapshot, reducing the total software development time. FlightDeck is leveraging two virtualization technologies, VMware and Docker. Moreover it offers flexibility to run workflows on public (Amazon EC2) or private (iDASH – HIPAA compliant) clouds. FlightDeck offers a viable solution toward reproducible genomics research in a user-friendly manner. FlightDeck is developed as an iDASH resource, supported by NIH/NHLBI National Centers for Biomedical Computing (U54HL108460).

1912W

Web Tool to Guide Submission to ClinVar. *M. J. Landrum, J. M. Lee, G. R. Brown, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, J. Hoover, W. Jang, K. Katz, M. Ovetsky, G. Riley, A. Sethi, R. Villamarin, D. R. Maglott.* NCBI/NIH, Bethesda, MD.

ClinVar (www. ncbi. nlm. nih. gov/clinvar/) is an archive of interpretations of variants and their relationship to disease, along with the evidence that supports the interpretations. Records are submitted to the database from clinical testing and research laboratories, locus-specific databases, OMIM, GeneReviews, as well as expert curation panels and groups who provide practice guidelines. Submissions range from hundreds or thousands of variants submitted by clinical testing laboratories to single variant submissions by researchers publishing their investigation of a clinically relevant variant. The Variation Submission Portal at NCBI was developed to facilitate the submission process to ClinVar, dbSNP, and dbVar. Via the Submission Portal, information about individuals and their organizations is provided once, rather than with every submission, which saves time for all types of repeat submitters. The Submission Portal also integrates information about laboratories that submit to variation databases and the NIH Genetic Testing Registry, thus making it easier to submit data to more than one resource. Until recently, use of Submission Portal allowed registration of the submitter and uploading of data files. We now announce the addition of a guided interface for direct data entry, targeted to research laboratories that infrequently want to submit a small number of records. The wizard prompts the user through the steps of defining the variant, defining the condition about which the interpretation is being submitted, the interpretation itself, and the observations or experiments provided as evidence. The wizard ensures that data types required for submission are provided and encourages use of data standards like OMIM, Orphanet, and HPO identifiers. Submitters are guided through the possibilities for data types with multiple options; for example, a submitter may provide a disease as part of the assertion, or a set of clinical features or the indication for testing as part of the observations. The Submission Wizard is designed to support all types of submissions to ClinVar, including structural variants, pharmacogenomics variants, somatic variants, as well as interpretations based on functional rather than clinical significance.

1913T

Ingenuity Variant Analysis, leveraging the Knowledge Base and HGMD®, achieves over 30x enrichment in biologically relevant variants from whole genome and exome sequence data from patients with rare disease. *S. Shah[1], D. Bassett[1], K. Boycott[2], J. Devaney[3], G. Eley[4], R. Felciano[1], S. E. Hofherr[3], A. Joecker[1], K. Kernohan[2], A. Krämer[1], B. W. Meltzer[3], A. Muthiah[1], K. Patel[1], D. Richards[1], M. B. Seprish[3], B. Solomon[4], A. Subramanian[1], J. G. Vockley[4], R. Yip[1].* 1) QIAGEN Bioinformatics, 1700 Seaport Blvd, Third Floor, Redwood City, CA, 94063, USA; 2) Children's Hospital of Eastern Ontario (CHEO), 401 Smyth Road, Ottawa, Ontario K1H 8L1; 3) Children's National Medical Center, 111 Michigan Ave NW # 800, Washington, DC 20010; 4) Inova Translational Medicine Institute, 8110 Gatehouse Road, Falls Church, VA, 22042, USA.

Advances in next generation sequencing technology have enabled sequencing of the whole exome and genome to identify genetic variants underlying rare inherited diseases. QIAGEN's Ingenuity Variant Analysis™ (www. ingenuity. com/variants) is an application that enables users to extract valuable insights from the large amount of genetic variation data generated by next generation sequencing. The platform integrates advanced informatics tools with biological knowledge from the literature and databases to enable real-time interactive filtering and rapid prioritization of variants to interpret the biology accurately and in more detail. Ingenuity Variant Analysis leverages the Ingenuity Knowledge Base, a repository of expert-curated mutation findings, biological interactions and functional annotations created from millions of individually modeled relationships between proteins, genes, complexes, cells, tissues, drugs, and diseases. To support users focused on hereditary disease, QIAGEN now has fully integrated HGMD® Professional, a unique resource that provides comprehensive data on human inherited disease mutations, with Ingenuity Variant Analysis. This unique combination of analytical tools plus content allows researchers to access HGMD reports for variants present in their sequence data through the Ingenuity Variant Analysis interface. To further improve the platform and improve causal variant detection, we designed preset filter cascades, implementing best practices for rare Mendelian disease genetic data analysis. To assess case solve rate, we analyzed the sequence data from 48 patients, afflicted with severe congenital abnormalities, for which the causal variant was previously known. For each sample, we applied the predefined filters, along with the observed medical signs and symptoms in the patient. By leveraging the analytical tools and Ingenuity Knowledge Base and HGMD, we achieved over 30x enrichment in the biologically relevant variants. In 96% of cases (46 out of 48), the causal variant was identified from among the top 75 variants. Furthermore, the causal variant was among the top 20 and 5 variants, in 81% and 44% of cases, respectively. These findings showcase the robust state-of-the-art capabilities and biological knowledge underlying QIAGEN's Ingenuity Variant Analysis, enabling fast and accurate discovery of rare disease causing variants within exome or whole genome data.

1914F

Impact of Genetic Variation on Three Dimensional Structures and Functions of Proteins. *R. Bhattacharya[1], P. W Rose[2], S. K Burley[2,3,4,5], A. Prlić[1].* 1) Bioinformatics and Medical Informatics, San Diego State University, San Diego, CA; 2) UCSD San Diego Super Computer Centre, San Diego , CA; 3) RCSB Protein Data Bank, Center for Integrative Proteomics Research and Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, New Brunswick, New Jersey; 4) BioMaPS Institute for Quantitative Biology and Cancer Institute of New Jersey, Rutgers, The State University of New Jersey, New Brunswick, New Jersey; 5) Skaggs School of Pharmacological Sciences, University of California, San Diego, San Diego ,CA.

Statement of Purpose:

The Protein Data Bank (PDB) currently contains more than 100,000 experimentally-determined 3D structures of biological macromolecules. Of particular interest in the PDB archive are protein structures for genetic variants. These structures can be utilized to gain deeper insight into the molecular mechanisms that are changed by genetic variation. Here, we explore these proteins and manually annotate 375 SNVs for which atomic level resolution data are available. We present an overview of the observed effects of SNV on the structure and function of proteins.

Methods:

Our dataset consists of 375 unique SNVs from human that have PDB coordinates. We systematically review the available literature and performed searches in several databases (RCSB PDB, dbSNP, NHLBI Exome Variant Server, Pubmed) to identify the structural and functional effect of the SNV.

Summary ofResults:

We observe a wide range of possible changes caused by single point mutations, starting from changes in enzyme activity, protein aggregation, structural stability, binding and dissociation, finally a number of SNVs change the arrangement of large biological assemblies. To highlight two examples:

Arylsulfatase A (gene: *ARSA*, OMIM *607574) is a protein that that breaks down sulfatides. The mutation P426L causes defective oligomerisation (dimer instead of octamer) that increases the susceptibility to lysosomal cystein proteinases and results in a severe reduction in half-life.

Delta-aminolevulinic acid dehydratase (gene: *ALAD*, OMIM *125270) catalyzes an early step in the biosynthesis of tetrapyrroles. The mutation F12L causes ALAD Porphyria (OMIM #612740), a rare autosomal recessive disease. The variant changes the protein assembly from an octamer to a hexamer. The optimal pH for the enzyme is shifted from pH 7 (wild-type) to pH 9 in the mutant. As a consequence the enzyme is barely active under physiologic conditions.

Conclusion:

An increasing number of software methods attempt to predict the consequences of genetic variation. Our results show that the variety of consequences is much larger than assumed by many of these methods and that correct prediction remains a significant challenge. A comprehensive understanding of three-dimensional structure, dynamics, and biophysics will be required to write software tools that can make accurate predictions about the consequences of genetic changes manifested at the atomic level in protein and RNA gene products.

1915W

Simultaneous Quantification of HDR and NHEJ Alleles Induced by Genome Editing Using ddPCR. *J. Berman[1], Y. Miyaoka[2], S. Cooper[1], B. Zhang[1], A. Chan[2], S. Mayerl[2], G. Karlin-Neumann[1], B. Conklin[2].* 1) Digital Biology Center, Bio-Rad Laboratories, 5731 W. Las Positas Blvd, Pleasanton, CA 94588; 2) Gladstone Institute of Cardiovascular Disease, 1650 Owens St, San Francisco, CA 94158.

Genome editing tools such as TALENs and the CRISPR/Cas9 system have revolutionized our ability to edit genomes of any cell including human induced pluripotent stem cells (iPSCs). Sequence-specific nucleases induce double strand breaks or nicks at target sites, activating the DNA repair pathways of non-homologous end joining (NHEJ) or homology-directed repair (HDR). Through the production of small insertions or deletions, error-prone NHEJ is useful for disrupting gene function. However, for many applications, HDR is more desirable than NHEJ, since HDR utilizes homologous donor DNA to produce precise gene repair, while NHEJ causes unpredictable damage. Unfortunately, cell populations edited by HDR often contain alleles damaged by NHEJ. Since NHEJ and HDR involve different repair enzymes, it is conceivable that conditions could be achieved with high HDR and low NHEJ. To investigate conditions that favor HDR over NHEJ, we developed a rapid assay using droplet digital PCR (ddPCR) that sensitively quantifies both HDR and NHEJ at endogenous genomic loci. Combination of ddPCR and fluorescent oligonucleotide probes specific to wild-type, NHEJ, and HDR alleles allowed us to simultaneously detect NHEJ and HDR events induced by site-specific nucleases in a sensitive and quantitative manner. Using this method, we measured in human cells the relative contributions of HDR and NHEJ under multiple genome editing conditions including different types of sequence-specific nucleases and donor DNAs. We are currently optimizing mutagenesis conditions to activate HDR with minimum NHEJ in human iPSCs. Our system will result in significant improvements to genome editing technology.

1916T

Correcting a dysfunctional splicing site of the *HLA-DRB1* gene from CTD-2510D15 bacterial artificial chromosome (BAC) using the recombineering method. *K. C. Chen[1], P. L. Chen[1,2,3,4].* 1) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 3) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan; 4) Research Center for Developmental Biology and Regenerative Medicine, National Taiwan University, Taipei, Taiwan.

Site-directed mutagenesis is a common method used to change specific sequence of a gene in vectors. However, if the plasmids or bacterial artificial chromosomes (BACs) are too huge to be manipulated or ligated, it becomes challenging to accomplish site-directed mutagenesis. In 2005, Warming *et al*. (Nucleic Acids Res, 2005, 33(4): e36) reported a recombineering method, taking advantage of galK-base positive-negative selection system. Through, researchers can insert selectable or nonselectable maker and clone DNA fragments without using restriction enzymes. Here, we identified a dysfunctional splicing site at the beginning nucleotide of intron1 of *HLA-DRB1*08:03:02* derived from CTD-2510D15 BAC clone. We retrieved *HLA-DRB1*08:03:02* and flanking DNA fragments from CTD-2510D15 BAC and cloned the whole segment into a pL253 vector, resulting a 29 Kb plasmid. Through the galK-based positive-negative selection recombineering method, we successfully replace the critical single nucleotide. Our results showed that galK-base positive-negative selection method is a useful and efficient recombineering method to make a single nucleotide engineering.

1917F

Long range phasing of cardiac disease genes using new long read sequencing technologies. *A. Dainis[1], M. Wheeler[2], E. Tseng[3], J. Fu[3], A. Hendel[4], E. Ashley[1,2].* 1) Genetics, Stanford University, Stanford, CA; 2) Cardiovascular Medicine, Stanford University, Stanford, CA; 3) Pacific Biosciences, Menlo Park, CA; 4) Pediatrics, Stanford University, Stanford CA.

Phasing, the process of determining which variants are inherited together on the same allele, has traditionally required familial genetic information or multifaceted barcoding and computational strategies. With the advent of long read sequencing technologies, multiple SNPs can be captured in a single read from a single DNA molecule, permitting phasing with little bioinformatic intervention. We used long read technologies, including Pacific Biosciences SMRT Sequencing and Oxford Nanopore MinION Sequencing, to perform long range phasing of RNA transcripts and genomic loci of cardiac disease genes *MYH7* and *MYBPC3* from left ventricular human heart samples. We found that long read sequencing can be an efficient and effective method of haplotyping without additional family information. We present a streamlined analysis pipeline to turn sequencing data into VCF files for easy analysis and manipulation. We also found variability across sequencing strategies in phasing ability. Fast, efficient phasing of cardiac disease genes will provide information about disease haplotypes as well as uncover potential targets for allele-specific therapeutic strategies.

1918W

Development of a Target Capture-Based Multi-Gene Next Generation Sequencing Assay for Detection of Variants in Solid Tumors. *R. Kanchi Ravi, M. Panjikaran, T. Wang, I. Mitchell, Y. Xia, R. Tseng, A. Papoutsis, S. Gadient, C. Scott, G. Sims, J. Falk, A. Graber.* R&D Assay Development Commercial Ops Team, Genoptix, A Novartis Company, Carlsbad, CA.

Introduction: Next generation sequencing (NGS) is increasingly used in clinical laboratories as one of the most cost effective and robust methods to detect and identify single nucleotide variants (SNVs), insertions/deletions (InDels), copy number variations (CNVs) and translocations in cancer patient solid cancer samples. Amplicon-based NGS assays frequently only cover hotspot mutations and do not provide a lot of information on genomic alteration types such as CNVs and translocations. In this study we evaluate the development of a target capture-based multigene panel NGS assay in order to detect SNVs, InDels, CNVs, and translocations in clinical tissue samples. **Methods:** First we identified regions of interest for clinically relevant gene variants through comprehensive bioinformatic analysis and designed appropriate probe baits to capture actionable genomic alterations across 173 genes. The gene panel was validated with a target capture-based parallel NGS method developed at Genoptix, which used DNA/RNA isolated from characterized reference cell lines and formalin-fixed paraffin embedded (FFPE) tissue, and analyzed through custom built bioinformatic pipeline tools. Variants detected by the capture-based NGS assay were confirmed using the Ion AmpliSeq CHPV2 50 gene panel for SNVs, Nanostring nCounter® method for CNVs, and Oncomine Fusion Gene Panel® from Life Technologies for the structural variants/translocations. **Results:** Validation results support a minimum average sequencing depth of 500x coverage, a lower limit of detection (sensitivity) of 5% mutant alleles for SNVs and translocation fusions, 10% for insertions/deletions, and CNV detection (>6 copies for amplification and <0. 3 for gene deletions). The target capture-based NGS assay was tested using molecular standards, reference cell lines and FFPE tissue samples, achieving >95% accuracy for all variant calls. Variants detected by the NGS assay were successfully confirmed in independent confirmation tests for SNPs and InDels, CNVs and translocation fusions. **Conclusions:** Overall, this study provides thorough analytical validation of the 173-gene NGS panel and the analysis tools that can be used in a clinical laboratory for routine testing with the ability to sequence multiple patient samples including those with only small amounts of FFPE DNA. We also observed the assay to be specific and sensitive for SNV mutation analysis, InDels, CNVs, and translocations.

1919T

CancerDIRECT: Targeted enrichment of cancer-associated genes from FFPE samples. *K. Patel[1], B. Galvin[1], N. Henig[1], T. Davis[2], S. Russello[2], C. Hendrickson[2].* 1) Directed Genomics, Ipswich, MA; 2) New England Biolabs, Ipswich, MA.

Next generation sequencing is a robust platform for analysis of cancer genomes and an important tool for the diagnosis, prognosis and selection of therapeutics of this disease state. To further these applications, targeted enrichment of the most important cancer-related genes must be fast, simple and cost-effective for clinically relevant samples, such as FFPE, which are typically of low-input and extensively compromised. We have developed a novel target capture strategy that offers advantages over traditional in-solution hybridization and multiplex PCR protocols. Using our DIRECT target enrichment method, DNA fragments are rapidly hybridized to short probes that are captured by beads. The off-target sequence of the captured DNA, is removed via enzymatic digestion. The resulting targeted sequences are ligated to NGS platform specific adaptors and amplified by PCR. With the DIRECT method, genomic DNA can be converted into libraries ready for sequencing within 7 hours. To determine the efficacy of this methodology, we designed the CancerDIRECT panel to interrogate 50 genes commonly mutated in cancer, and tested it on FFPE samples. NGS analysis on libraries generated from these compromised samples, demonstrates that our approach captures targeted sequences with high specificity and uniformity, and detects variants with high accuracy and sensitivity.

1920F

Development of research genetics panel of 28 genes known to be associated with Primary Ciliary Dyskinesia (PCD) using Ion Torrent Technology. *M. A. Zariwala[1], M. W. Leigh[2], M. L. Daniels[3], W. E. Wolf[3], H. Dang[3], P. Mieczkowski[4], N. R. Schuch[4], A. Kraltcheva[5], D. Mazur[5], P. B. Vander Horn[5], M. R. Knowles[3].* 1) Marsico Lung Institute, Department of Pathology & Laboratory Medicine, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA; 2) Marsico Lung Institute, Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA; 3) Marsico Lung Institute, Department of Medicine, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA; 4) High Throughput Sequencing Facility, Department of Genetics, University of North Carolina School of Medicine, Chapel Hill, NC 27599, USA; 5) Life Sciences Solutions, Thermo Fisher Scientific, Carlsbad, CA 92008, USA.

A defective ciliary structure, function or biogenesis leads to primary ciliary dyskinesia (PCD); a rare, recessive, genetically heterogeneous disorder. PCD diagnosis relies on identifying defective ciliary ultrastructure, and/or ciliary dysmotility/immotility, and/or clinical phenotype coupled to nasal nitric oxide measurements. Approximately, 30% of PCD cases do not have ultrastructural defects; thus, defining biallelic PCD-causing mutations is critical to aid diagnosis. Currently, mutations in 32 genes account for 2/3rds of all PCD cases. Development of a targeted gene panel with next generation sequencing assay for PCD may circumvent locus heterogeneity. A custom Ion AmpliSeq™ research panel was designed to sequence 28 (of the 32) PCD, plus 4 candidate genes using the Ion Torrent Personal Genome Machine (PGM™). A validation test-run included 15 cases with 28 known variants (7 base changes, 13 small indels, and a 220-kb deletion) across 14 genes. Twenty-four of 28 (86%) variants were called correctly, solving 12 of 15 (80%) cases by identifying biallelic mutations. Two (of 4) variants were missed due to amplification failure, 1 was missed due to its location (end of an amplicon), and 1 was 220-Kb deletion the panel was not designed to detect. The assay was refined as follows: Ion Library Equalizer™kit was used to normalize library input from different samples, 11 samples (instead of 20) were run on each 318™chip, and a list of 'hot spot' for known mutations was added to the Torrent Suite analysis pipeline to avoid missing known mutations. To discover mutations, 32 unrelated PCD cases with various ultrastructural defects were analyzed. The results revealed: 13 (41%) had two loss-of-function (LOF) mutations across 8 genes (*DNAI1*, *DNAH5*, *DNAH11*, *CCDC39*, *CCDC40*, *HYDIN*, *RSPH4A*, *SPAG1*); 6 (19%) had 1 LOF mutation and 1 rare missense variant across 4 genes (*DNAI1*, *DNAH5*, *DNAH11*, *DNAAF1*); 2 (6%) had two rare missense variants across 2 genes (*DNAH5*, *DNAAF3*); 4 (12%) had only 1 LOF mutation across 3 genes (*DNAH11*, *HYDIN*, *SPAG1*); and 7 (22%) cases were negative. In conclusions, biallelic variants in 21 (66%) of 32 PCD cases were identified using Ion AmpliSeq™ panel; thus, the panel is useful for the test translation into the lab. Additionally, the 11 (34%) negative cases will be useful for novel discoveries and large deletion/duplication analysis. This abstract was funded by 5U54HL096458-06 (NIH/ORDR/NHLBI), 5R01HL071798 (NIH/NHLBI), and UL1 TR000083 (NIH/NCATS).

1921W

Clinical Whole Exome Sequencing from Dried Blood Spot Identifies Novel Genetic Defect Underlying Asparagine Synthetase Deficiency. *A. Abhyankar[1], M. Lamendola-Essel[1], K. Brennan[2], C. Esteves[1], V. Felice[1], R. Wapner[2], V. Jobanputra[1].* 1) Molecular Diagnostics, New York Genome Center, New York, NY; 2) Department of Obstetrics & Gynecology, Columbia University, New York.

Purpose: Dried blood spots (DBS) are collected worldwide as part of neonatal screening programs and can prove invaluable in investigating etiologies underlying rare and common genetic disorders. We show that DNA isolated from archived DBS can be used to perform clinical-grade whole exome sequencing (WES) with accuracy comparable to similar assays performed using DNA extracted from whole blood. **Methods:** We received archived DBS for a premature baby delivered at 32 weeks of gestation, presenting with clonus tremors shortly after birth, microcephaly, blindness and seizures. The baby passed away at 15 months of age because of respiratory problems/seizures. DNA was isolated from DBS using Extract-N-Amp Blood PCR kit (Sigma). Whole genome amplification was performed in duplicate using the Repli-g Amplification kit (Qiagen). Peripheral blood drawn from both parents was used to isolate DNA. Library preparation was performed using SureSelectXT Human All Exon V5+UTR kit (Agilent). Paired-end WES was performed on the HiSeq 2500 (Illumina). Data was analyzed using in-house developed bioinformatics pipeline. **Results:** Exome sequencing performed on DNA obtained from DBS and parents identified Gly283Glu and Val160Ala compound heterozygous variants in *ASNS* gene. Parental studies demonstrated that the Gly238Glu variant was maternally inherited and the Val160Ala variant was paternally inherited. Computational tools predicted the amino acid changes Gly283Glu and Val160Ala to be damaging and deleterious (SIFT and Polyphen; CADD scores of 27. 1 and 20. 7 respectively). The Gly283Glu variant has not been reported in large public data sets (1000 Genomes Project and NHLBI GO Exome Sequencing Project). The Val160Ala is a rare variant, reported at frequency of 0. 0002 in the 1000 Genomes Project with no homozygous G/G genotype reported. Identified variants in the *ASNS* gene are considered likely pathogenic and explain the clinical phenotype. **Conclusion:** We report novel compound heterozygous variants underlying Asparagine Synthetase Deficiency and that DNA isolated from DBS stored for several years can be used to perform accurate clinical WES assays.

1922T

Sherloc: Evaluation of a Scalable Score-based Implementation of the ACMG 2015 Clinical Variant Interpretation Guidelines. *M. Anderson, K. Nykamp, S. Yang, Y. Kobayashi, M. Powers, S. Lincoln, S. Topper, Invitae Clinical Genomics Group.* Invitae, San Francisco, CA.

Recently, the American College of Medical Genetics (ACMG) published updated guidelines for the interpretation of sequence variants (ISV) with an evidence checklist and rules for combining evidence (Richards *et al.* , *Genet Med* 2015). While these guidelines represent a significant step forward in the standardization of variant classification across clinical laboratories, they are a general framework and require each laboratory to establish more specific procedures for their particular clinical setting. Moreover, the impact of these guidelines has not yet been thoroughly established in the literature. Using the draft ISV checklist (August 2013) as a starting point, we developed a weighted, score-based classification system designed to be scalable across a large clinical variant interpretation team. This system, called Sherloc, has been iteratively revised 4 times based on our experience interpreting more than 11,000 variants. These revisions increased the number of criteria from 22 to 87 in order to explicitly handle evidence details not directly addressed by the ISV guidelines. We also adjusted thresholds to improve classifications, created rules to properly account for redundant and conflicting evidence, and added extensive usage notes to increase consistency. To assess the concordance of Sherloc with current community standards, we compared Sherloc classifications of over 800 variants to a consensus classification derived from ClinVar entries with multiple submitters. We find that Sherloc interpretations are often (92. 2%) in the consensus majority, indeed more so than other ClinVar submissions when benchmarked equivalently (85. 7%). Most differences with ClinVar were not clinically impactful (e. g. , between likely benign and benign). We also find Sherloc to be highly reproducible across different team members when providing independent interpretations. By contrast, interpretations based on strict adherence to the ISV criteria have lower agreement with the community consensus (61%) and a higher VUS rate. We find Sherloc to be highly adaptable, efficient, and consistent across multiple disease areas and a large variant interpretation team. This score-based classification system adheres to the ISV guidelines, but also illustrates the specific application and evolution of the ACMG ISV criteria in a clinical molecular laboratory. We are in the process of publishing and releasing Sherloc for community feedback and in the hope that it may be useful to others.

1923F

Development of Next Generation Sequencing Based Genetic Diagnosis and Screening Panel for Hereditary Periodic Fever Syndromes. *T. Avsar[1,2], C. Holyavkin[1,2], N. Saygili[1,3], E. Yilmaz[1,3], M. Öztürk[1], E. Keleş[1], I. V. Bodur[1].* 1) R&D Department, Done Genetics and Bioinformatics, Istanbul, Turkey; 2) Istanbul Technical University, Molecular Biology, Genetics and Biotechnology Program, Istanbul, Turkey; 3) Istanbul Universty, Institute of Experimental Medicine, Istanbul, Turkey.

The hereditary periodic fever syndromes (HPFS) refer to a unique and complex disease group with hereditary genetic defects and periodic clinical symptoms. The prototypic and probably most common HPFS is Familial Mediterranean Fever (FMF) that is common in Turkey and other Mediterranean countries. Currently there are more than 20 HPFS diseases, some are rare and some are common in different populations. Diagnosis and prognosis of HPFS diseases differ based on the clinical manifestations and genetic mutations. An accurate genetic diagnosis will reduce the number of patients who having non specific treatments and living without a correct diagnosis. The aim of this study was to develop a next generation sequencing based genetic test for HPFS related 9 genes and validate them. To this aim, amplicons targeting 9 related genes were designed for all exons, some introns, exon-intron boundary and UTR regions. Primers were designed for long-PCR applications and, in silico analysis were performed in order to develop 7 multiplex PCR for 32 amplicons. Following library preperation, adapter and index tagmentation of PCR fragments were performed via in house designed oligos suitable for Illumina NGS systems. 120 previoulsy diagnosed HPFS samples were studied by using in house developed test. Results revealed common and shared genetic variants associated with HPFS (Table 1). Sensitivity, specificity, accuracy, positive/negative predictive values of the test were given in Table 2. The mean covarages of each genes were above 220X readings. The diagnostic yield of the tested samples were approximately 80%. Furthermore frequency of known SNPs were also compared in our Turkish sample cohort and it was shown that some rare phenotypes in different populations common in Turkish population and significantly associated with the disease phenotypes. Selected SNPs were confirmed by using Sanger sequencing. Additionaly, we developed a standardized bioinformatic analysis pipeline using Galaxy web server specific for HPFS panel. This strategy also facilitates the development of additional disorder-specific gene panels with the same standardized workflow. We evaluated the performance characteristics sequencing workflow, and found high reproducibility and accuracy, with 100% detection of HPFS genetic variants in our validation samples. We also explore the advantages and technical challenges of this strategy to further develop clinically relevant gene panels for genetically complex disorders.

1924W

A custom-designed next generation sequencing panel detecting extremely low-level mosaic mutations in overgrowth syndromes. *F. Chang[1], L. Liu[2], E. Fang[2], G. Zhang[2], T. Chen[2], M. M. Li[1].* 1) Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Overgrowth syndromes are genetically heterogeneous diseases caused by germline or postzygotic somatic mutations of different genes. Recent studies have shown that a group of mosaic overgrowth syndromes, such as congenital lipomatous overgrowth with vascular, epidermal, and skeletal anomalies (CLOVES), Proteus syndromes, and two overlapping disorders, megalencephaly-polymicrogyria-polydactyly-hydrocephalus (MPPH) and megalencephaly-capillary malformation (MCAP), are caused by somatic activating mutations in the genes involved in the PI3K-AKT signaling pathway. Due to the mosaic nature of these mutations, routine Sanger sequencing often yields negative results. We have developed a next generation sequencing (NGS) panel that covers all known mutations associated with these overgrowth syndromes in multiple genes involved in the PI3K-AKT pathway. Thirty three cases including two prenatal cases and thirty one postnatal cases suspected of overgrowth syndromes were tested using the NGS sequencing panel. Seventeen of the 33 cases showed activating mutations in PIK3CA with mutant allele frequencies ranging from 1. 5% to 49. 2%. The mutations were predominantly located in the kinase domain (10/17) followed by helical domain (4/17) similar to the mutation distribution reported in cancers. These PIK3CA mutations were only present in affected tissues in the majority of the cases demonstrating causal role of the mutation. In vitro cell culture showed significant enrichment of the cells harboring mutant alleles, suggesting that the activating mutation renders growth advantages to the mutant cells both in vivo and in vitro. Preliminary phenotype-genotype correlation analysis showed that mutations at residue 1047 are often associated with CLOVES syndrome while G914R mutation is preferentially seen in MPPH/MCAP. Our experience demonstrates that NGS technology is highly sensitive for the detection of low-level mosaic mutations and can be used for the diagnosis of these mosaic overgrowth syndromes in both prenatal and postnatal settings.

1925T

Analysis of sample multiplexing in massively parallel sequencing of the human clinical exome. *J. Devaney, B. Meltzer, K. Cusmano-Ozog, J. Campos, S. Hofherr.* Laboratory Medicine, Children's Natl Health System, Washington, DC.

The quality of data from massively parallel sequencing (MPS) can be affected by DNA quality. Another factor that can affect the quality of a MPS analyses is the number of samples that can be pooled in a sequencing run while achieving the desired coverage level. Our lab sought to optimize the number of samples that can be studied on a run an MPS system using the human clinical exome that consists of ~ 4800 genes. Genomic DNA was purified from peripheral blood using standard extraction conditions as recommended by the EZ1 DNA Kit (Qiagen) using an EZ1 Advanced XL. The MPS was completed using the TruSight One panel (4,813 genes) on a NextSeq 500 (Illumina). To find an optimized number of samples that can be multiplexed in each run, we examined the percent Qscore > 30, mean coverage depth, percent target coverage at 20X, confirmation of known variants determined by sequencing at reference labs, and the number of fragments under 20X coverage. We analyzed 48, 36, and 24 samples using a high output chip and 15, 12, and 9 samples using a mid-output chip on a NextSeq 500. The concentration of our library for all analysis was 0. 8 pm. The TruSight One (TSO) panel discovers around 8000 variants in the NA12878 (Genome in a Bottle) with over 99% present in the dbSNP141 database and typically discovers around 150 insertions with 88% in dbSNP and 210 deletions with about 78% appearing in dbSNP. We found that the Qscore > 30 was the same across all the runs. The mean coverage increased as the number of multiplexed samples decreased going form 80X to 151X on the high output chip while on the mid output chip went from 91X to 133X on the mid output chip. The percent target coverage at 20X went from 93. 6% (48 samples) to 97. 3% (24 samples) using the high output chip and from 94. 6% (15 samples) to 97. 3% (9 samples) with the mid output chip. For all the samples, we were able to confirm the known variants (SNVs and indels). Finally, we saw the number of fragments with coverage less than 20X drop from 2. 0% (1249 fragments; 48 samples) to 0. 8% (496 fragments; 24 samples). We were able to confirm all variants from reference labs in the samples regardless of the multiplexing level. Interestingly, the percent duplicate paired reads increased as the number of samples multiplexed decreased. In our lab, the optimal multiplexing of TSO panel samples is 24 for a high output chip and 9 for a mid output chip with analysis using a Nextseq 500.

1926F

Augmenting clinical exomes with low pass genome sequence to identify copy number variation. *M. O. Dorschner, M. A. Weaver, E. A. Phillips, M. B. Dauer, D. S. Hanna, C. J. Hale.* Center for Precision Diagnostics, Department of Pathology, University of Washington, Seattle, WA.

Clinical exome sequencing is rapidly being applied to the molecular diagnosis of human genetic disease. While exome sequencing has several advantages over small gene panels or even genome sequencing, it is not effective at identifying copy number variants (CNVs). Algorithms designed to detect CNVs from read depth or split read mapping do not provide adequate sensitivity or specificity for clinical testing. We recently launched an exome-based assay for the diagnosis of dementia, Parkinson's disease and amyotrophic lateral sclerosis. Several genes, including *PARK2* and *SNCA* frequently undergo genomic rearrangement. Rather than concurrently, or reflexively running a deletion/duplication array to detect potential rearrangements, we decided to evaluate the addition of low pass genome sequencing to exome data. We initially sequenced eight samples carrying known CNVs to a read depth of 8X. Six samples possessed CNVs in *PARK2* and two in *SNCA*. All CNVs, ranging from single- to multi-exon deletions, as well as whole gene duplication and triplication of *SNCA*were detected. To determine the lowest genome coverage that could be applied to the detection of these CNVs, we subsampled the 8X genome data down to 1X . At 1X, all CNVs were confirmed. We recently applied this approach to the molecular diagnosis of an early onset Parkinson's patient. This patient possessed a 40 base deletion in exon 3 of one *PARK2* allele, which appeared homozygous. Knowing that exons in this region of *PARK2* are commonly deleted or duplicated, we suspected an overlapping deletion of the remaining allele. Low pass sequence data of this patient's genome revealed a 175kb deletion, encompassing exons 3 and 4 of the other allele. Additional studies to investigate a broader range of CNV sizes, and sequence depths are under way to determine the sensitivity of combined low pass genome and exome sequencing. Low pass genome data represents a potential, cost effective method for expanding the sensitivity of exome sequencing and the scope of detectable disease-causing variants.

1927W

Multi-ethnic *APOL1* G1 and G2 allele frequencies and clinical validation of a personalized medicine *APOL1* genotyping assay for non-diabetic chronic kidney disease risk assessment. *A. M. Fedick[1], J. Zhang[1,2], G. Zhao[1], L. Edelmann[1], E. P. Bottinger[3], R. Kornreich[1], S. A. Scott[1].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY.

The incidence of chronic kidney disease (CKD) significantly varies by ancestry and ethnicity, with African Americans (AAs) having a three to four-fold higher rate than European Americans (EAs). In addition to environmental causes, genetic studies have revealed that variant alleles in the apolipoprotein L1 (*APOL1*) gene, termed G1 (c. [1072A>G;1200T>G]) and G2 (c. 1212_1217del6), are strongly associated with increased risks for hypertensive end stage renal disease (ESRD), focal segmental glomerulosclerosis, HIV-associated nephropathy, and higher rates of non-diabetic CKD in the AA population. These data have prompted the implementation of *APOL1* genetic testing to identify at-risk AA patients for counseling to modify other risk factors (e. g. , hypertension) in an effort to reduce the progression of CKD to ESRD. To facilitate this personalized medicine program, a targeted *APOL1* genotyping assay was developed based on multiplex allele-specific primer extension using the Luminex platform. The *APOL1* genotyping assay was clinically validated using 58 positive and negative control samples, which resulted in complete concordance between genotyping and Sanger sequencing, suggesting a test sensitivity and specificity approaching 100%. The *APOL1* genotyping assay was found to be very robust, as both triplicate inter- and intra-run reproducibility testing results were completely concordant. To determine the *APOL1* G1 and G2 allele frequencies of the multi-ethnic population from the New York City metropolitan area, 7462 DNA samples (5463 AA, 1148 Hispanic, and 460 Asian) were subjected to genotyping. The *APOL1* G1 and G2 allele frequencies were 0. 22 and 0. 13, 0. 037 and 0. 025, and 0. 013 and 0. 004 for the AA, Hispanic, and Asian populations, respectively. Importantly, approximately 14%, 7%, and 3% of AA, Hispanic, and Asian individuals, respectively, carried two of the G1 and/or G2 alleles and are at elevated risk for developing CKD. Taken together, this novel *APOL1* G1 and G2 genotyping assay is robust and highly accurate, and represents one of the first personalized medicine clinical genetic tests beyond the currently available clinical pharmacogenetic assays.

1928T

To Confirm or Not Confirm, That is the Question: A Rigorous Approach to Evaluating the Importance of Sanger Confirmation of Clinical NGS Findings. *S. Lincoln[1], B. Funke[2,3], J. Zook[4], S. Yang[1], D. Watson[1], M. Bowser[3], S. Aradhya[1], S. Kulkarni[5], H. Rehm[2,3].* 1) InVitae, San Francisco, CA; 2) Department of Pathology, Harvard Medical School, Boston, MA; 3) Laboratory for Molecular Medicine, Cambridge, MA; 4) National Institute for Standards and Technology, Gaithersburg, MD; 5) Washington University, Saint Louis, MO.

Clinical next-generation sequencing (NGS) is becoming an established diagnostic tool. Because test results can influence significant medical decisions, many labs rule out false positive calls by confirming variants with an orthogonal method such as Sanger sequencing. The collective experience of many laboratories doing so is that the great majority of NGS calls indeed confirm, and therefore the need to broadly apply orthogonal confirmation has been questioned. However, identifying the highest confidence NGS calls and quantifying the degree of confidence in these calls can be challenging. Our cross-laboratory collaboration has developed a framework for addressing this challenge which can be used with different assay targets, NGS protocols, bioinformatics algorithms, and QC criteria. In brief, our framework involves: (1) assembling confirmation data from each laboratory across clinical and reference samples, (2) determining laboratory-specific QC thresholds for classes of variants [KT1]that each data set adequately describes, and (3) using proper statistical metrics to quantify the accuracy of variant calls that meet these thresholds. We use multiple thresholds: a strict tier, over which the likelihood of a false positive is exceedingly low, and a lower tier established to ensure sensitivity. Reported variants from the lower tier would need orthogonal confirmation while confirmation of the higher-tier variants may depend on clinical factors or operational criteria. To date, we have applied this framework to data from four laboratories. For example, in one dataset for 184genes in an NGS panel, 3082 variants were assembled as a complete and representative set of calls that (a) met that lab's top-tier QC criteria (e. g. , read depth and other factors), and (b) had high-quality orthogonal data available. All 3082 were confirmed as true positives. Further bioinformatic analysis identified a subset of 2394 with homogeneous properties in terms of genomic context, variant types and local sequence properties. The observed analytic false-discovery rate (FDR) for the 2394 remained 0. 0% and statistically these data demonstrate an FDR of at most ~0. 1% for such variant calls (at p=0. 05). Our framework is the first to combine data across labs to help evaluate the value of confirmation and determine the appropriate burden of proof needed potentially change practice. We believe these results and this framework can contribute to the ongoing community dialog on this subject.

1929F

Classification of truncating variants in the 3' end of genes requires a detailed analysis of C-terminal protein structure and function. *C. Perreault-Micale, K. Robinson, V. Greger.* Good Start Genetics, Cambridge, MA.

Truncating variants that introduce premature stop codons (nonsense variants and insertions/deletions causing a frameshift) or disrupt canonical splice sites (positions +/- 1 or 2) are generally considered to lead to loss of gene function. However, truncating variants that are near the 3' end of a gene may escape nonsense-mediated decay (NMD) and produce a protein that is at least partially functional. Genome-wide assessments typically use the last 5-10% of genes as a rough cut-off for pathogenicity due to NMD (i. e. , only truncating variants in the upstream 90-95% of the gene are considered pathogenic). However, a more detailed approach is required for the analysis of specific genes. Our goal was to assign gene-specific 3' cut-offs after which truncating variants should not be considered pathogenic for the purpose of genetic carrier screening. We performed an analysis of the C-terminal protein structure and function using information available, and also considered the last known truncating pathogenic variant reported in a patient, for 15 genes underlying severe autosomal recessive disorders in which loss of function is a known disease mechanism. Our results indicate that using a 3' cut-off of 5-10% for truncating variant pathogenicity is not the most accurate approach because it can lead both to pathogenic variants being missed, and to likely benign variants being mischaracterized as pathogenic. For example, in HEXA (Tay-Sachs disease) there is a crucial disulfide bond at p. Cys522 (7 amino acids from C-terminal end). Thus, biochemical and structural evidence, and the fact that this is a highly conserved residue, indicate variants affecting the extreme C-terminus of HEXA should be considered pathogenic. In BCKDHA (MSUD, type A) p. Tyr438 (7 amino acids from C-terminal end) is needed for proper complex assembly. In DLD (dihydrolipoamide dehydrogenase deficiency) regions up to p. Glu492 (17 amino acids from C-terminal end) correspond to important oxidoreductase and dimerization domains. Conversely, the high incidence of truncating variants in the cytoplasmic domain of PCDH15 (Usher syndrome, type 1F), which makes up nearly one-third of the gene, suggests they are unlikely to cause disease, possibly due to redundant isoforms. In *PCDH15*, many truncating variants are not necessarily loss of function variants. Thus, analysis of truncating variants in the 3' end of genes should be done in a gene-specific manner to improve variant classification.

1930W

Clinical validation of a NGS based in vitro diagnostic (CE-IVD) kit for targeted detection of actionable gene rearrangements in lung cancer specimens. *J. Schageman[1], J. Costa[2], O. Sheils[3], J. Glassco[4], D. Chi[4], J. Cooc[4], E. Ballesteros-Villagrana[1], R. Petraroli[1], J. Sherlock[1], J. Bishop[1], K. Bramlett[1].* 1) Thermo Fisher Scientific, , UK, Italy, Netherlands, USA; 2) Institute of Molecular Pathology and Immunology, University of Porto, Porto Portugal,; 3) Sir Patrick Dun Research Laboratory, St. James Hospital, Dublin, Ireland; 4) Life Technologies Clinical Services Lab, West Sacramento, CA USA.

In recent years, advances in next-generation sequencing (NGS) technologies have enabled faster and cheaper methods for uncovering the genetic basis of disease. For cancer, NGS based screening for known tumour subtypes can inform diagnosis and allow the clinician to tailor a specific therapy based on testing outcome. Here we present the validation of one such NGS based kit approved for CE-IVD* use to screen for specific chromosomal translocations in non-small cell lung cancer (NSCLC) samples by targeting specific breakpoints in known fusion transcripts.

The kit tested (Oncomine™ Solid Tumour Fusion Transcript Kit) included a single primer pool containing amplicon designs to simultaneously screen for over 75 specific rearrangements involving the receptor tyrosine kinase (RTK) genes ALK, RET and ROS1 as well as NTRK1. The panel was compatible with formalin-fixed paraffin-embedded (FFPE) lung tumour samples and achieved high-sensitivity down to 10 ng of RNA input. In addition, amplicon assays designed at the 5' and 3' ends the RTK genes provide non-specific evidence that a translocation exists in a sample by comparing expression imbalance between the two ends.

Validation testing was carried out at three external clinical laboratories (CLIA, CAP, INAB). In addition to positive and negative control samples, each site contributed FFPE lung tumour samples for which ALK fusion status was known prior to NGS library preparation carried out using the Ion AmpliSeq™ workflow. For site-specific samples (n=144, 16 samples per sequencing run), high concordance, sensitivity and specificity were measured at 97. 2%, 90. 5% and 98. 4%, respectively. Validation testing was carried out both in the United States and Europe and this kit isCE marked and registered in accordance with the European*in vitro* Diagnostic Medical Devices Directive (IVDD) (98/79/EC).

1931T

Non-ceruloplasmin-copper involvement in Alzheimer's disease and Mild Cognitive Impairment. *R. Squitti[1], M. C. Ventriglia[1], M. Cortesi[2], F. Papa[2], C. Vaccarella[2], M. B. Majolini[2], M. CA. Rongioletti[2].* 1) Department of Neuroscience -AFAR division, San Giovanni Calibita Hospital, Rome, Italy; 2) Clinical Patholgy Department, San Giovanni Calibita Hospital, Rome, Italy.

Numerous studies indicate a specific role of copper in aging and in Alzheimer's disease (AD) mechanisms. Specifically, it has been proposed that the hypermetallation of the Abeta peptide, which forms plaques in the AD brain, can be at the basis of redox cycles of oxidative stress, toxicity, Abeta oligomer formation and precipitation (reviewed in Squitti and Polimanti, 2013). We have contributed to this topic demonstrating in living AD patients that an increase in the serum copper fraction that does not bind to ceruloplasmin (Non-Cp-Cu) correlates with the AD typical deficits (Squitti, et al. , Neurol 2002, 2005), cerebrospinal fluid (CSF) markers of AD (Squitti et al. , Neurol 2006), and with a worse prognosis for of the disease (Squitti et al. , Neurol 2009). Furthermore, we have shown that loci of susceptibility for AD lie in genes pertinent to copper metabolism (Squitti et al. , 2013; Bucossi et al. , 2012; Squitti and Polimanti 2013, Bucossi etal J Alzh Dis 2013), in particular in the *ATP7B* gene, which codes for the ATP7B-pump controlling copper excretion through the bile and ceruloplasmin biosynthesis. Moreover, we have demonstrated that copper and non-Cp-cu levels are higher in AD patients vs. healthy controls by means of meta-analyses (Squitti et al. , J Alzh Dis 2014). Non-Cp-Cu helps in properly classifying subjects with Mild Cognitive Impairment (MCI) from healthy ones (Squitti et al. , J Alzh Dis 2011) and provides prognostic information about the conversion to full AD (Squitti et al. , Ann Neurol 2014). We have recently developed a patented devise to directly measure serum Non-Cp-Cu [Colabufo, N. and R. Squitti, P. E. European Patent Office (EPO) (RO/EP), 2012], highly reliable.

1932F

Whole Exome Sequencing for Molecular Diagnosis of Chinese Patients with Rare Genetic Disorders. *J. Wang[1], Z. Wang[1], L. Yin[1], X. Huang[1], T. Yu[1], R. Yao[1], J. Geng[1], N. Li[1], Y. Qing[1], X. Wang[2], Q. Fu[1], Y. Shen[1,3], Rare Diseases Collaboration in Shanghai Children's Medical Center.* 1) Molecular Diagnostic Center, Shanghai Children's Medical Center, Shanghai, Shanghai, China; 2) Guizhou Provincial People's Hospital, Guiyang, China; 3) Boston Children's Hospital, Boston, MA, USA.

Background: Over 7000 genetic disorders have been reported in Online Mendelian Inheritance in Man (OMIM). Most of them are rare and often difficult to be recognized and diagnosed at clinics in a timely manner. Whole exome sequencing (WES) has increasingly been utilized both as a disease gene discovery tool and a molecular diagnostic tool for rare genetic diseases. We set to evaluate the technical and clinical validity of WES as a diagnostic tool in a rare and complex disease clinic of a tertiary pediatric hospital in China. **Methods:** (1) We established the workflow of a WES test including SureSelect exome target capture and library construction, high-throughput sequencing with Illumina Hi-Seq or Mi-seq platforms, data analysis and variant processing with NextGENe® software and Ingenuity Variant Analysis online tool. (2) 72 children with various rare conditions of suspected genetic basis were selected for WES testing. We used Sanger sequencing for confirming candidate variants in proband. Additional family members were Sanger sequenced for the variants for co-segregation analysis. **Results:** (1) More than 99% of targeted regions were successfully captured and aligned to reference human genome. > 70% of total reads were within targeted regions using Agilent SureSelect V5 kit. >92% of targeted regions had >20X coverage when the medium coverage for the whole exome reached 100X. (2) We identified 47 pathogenic variations in 38 of the 72 patients, achieving a 52. 78% molecular diagnostic rate. Among the 38 patients, 24 had autosomal dominant disease, 9 had autosomal recessive disease, and 5 had X-linked disease. Of the 47 pathogenic variations, 22 were novel variants, 6 were *de novo*. 23 genes such as *ATP6V0A4, BMPR2, CPS1, CYP11B2, EDA, GORAB, KRT1, LGI1, MAP2K1, MYBPC3, MYH7, NR5A1, PDGFRB, PHEX, PLA2G6, PTHLH, SLC45A2, SMC1A, TRAPPC11, WT1* were involved with the diseases. **Conclusion:** We established a complete WES process from clinical ascertainment to variant confirmation and segregation analysis. We are able to achieve a high diagnostic yield (~53%) for the selected Chinese pediatric patients from the clinic. The time to reach a molecular diagnosis for patient is on average 28 days, with a material cost of $500 per case. The data support the utilization of WES as an effective diagnostic approach for patients with rare genetic disorders.

1933W

Strategies for calculating variant confidence by combining sequencing results. *N. Chennagiri, D. Lieber, T. Yu, J. Thompson.* Claritas Genomics, Cambridge, MA.

Based on the American College of Medical Genetics guidelines, clinical laboratories using High Throughput Sequencing Technology (HTS) should confirm all variants they report using an orthogonal technology, with many using Sanger sequencing. As whole exome sequencing becomes increasingly popular and reveals thousands of variants per run, clinical laboratories using whole genome technologies are faced with having to confirm a large number of variants which increases cost and turnaround time. At Claritas Genomics, we have developed an approach that uses an orthogonal technology and minimizes the need for Sanger sequencing. We sequence a sample using two HTSs – Illumina NextSeq and Ion Torrent Proton. We have developed methods to combine the two calls into a consensus call and assign confidence levels. We assign levels of confidence based on Positive Predictive Value (PPV) of the consensus calls calculated using NA12878 NIST reference dataset. About 85% of the total calls are concordant between the two technologies and are assigned the highest level of confidence - Orthogonally confirmed. The calculated PPV for such variants is 100%. These variants can be reported immediately without additional Sanger sequencing. Other variants are assigned lower confidence levels based on their PPV such as likely true positives or likely false positives and, if appropriate, can be Sanger confirmed or dropped from consideration if irrelevant to the phenotype being examined. The confidence levels provide a method for prioritizing variants to confirm and minimizing variants needing Sanger sequencing thereby significantly reducing turnaround time.

1934T

***CLIP-Cap*: *C*ombined *L*ong-*I*nsert *P*aired-End and *Cap*ture Sequencing for precise and comprehensive analysis of complex genomic rearrangements.** *C. Purmann[1,2], J. Hallmayer[1], J. Bernstein[3], A. E. Urban[1,2].* 1) Department of Psychiatry & Behavioral Sciences, Stanford University, Palo Alto, CA; 2) Department of Genetics, Stanford University, Palo Alto, CA; 3) Department of Pediatrics, Stanford University, Palo Alto, CA.

Currently, the study of complex genomic rearrangements requires several complementary yet disparate assays to characterize the exact nature of the rearrangement. Techniques that can reduce the number of experiments while achieving more accurate identification are highly sought after. Toward this aim, we have developed a method that combines Long-Insert Paired-End Sequencing and chromosome-wide targeted capture into one experimental workflow. This new genomic tool enables the resolution of complex genomic rearrangements in a single assay. For proof-of-principle, we analyzed a Phelan-McDermid patient with a complex and mosaic rearrangement on chromosome 22q13. A combination of four different assays – karyotyping, high-density microarray, fluorescence in situ hybridization (FISH), and digital droplet PCR from multiple subcloned cell lines - had revealed the presence of three different cell populations in the patient sample. Half of the patient cells carried a heterozygous terminal 22q13. 3 deletion. The other half of the cells was split into further two subgroups, both showing the presence of an isodicentric chromosome 22, but with two different breakpoints. CLIP-Cap, using an oligomer capture-design representing all of chromosome 22q, and a single sequencing run on an Illumina MiSeq instrument (2x300 nt and ~10M mappable PE-reads) allowed us to completely resolve this complex genotype with a single procedure, including determining the heterozygous terminal 22q13. 3 deletion and the different isodicentric breakpoints. We have been applying CLIP-Cap to additional cases from both patients with developmental abnormalities and cancer. Here, we present CLIP-Cap as a novel and widely applicable technology to resolve complex chromosomal abnormalities at highest resolution and with procedural and cost requirements that are comparable to, or an improvement on, current less highly resolving approaches.

1935F

Next-generation sequencing-based genetic testing of malignant hyperthermia. *H. Yeh[1], M. Liao[2], W. Yang[3,4], P. Chen[2,3,4].* 1) Department of Anesthesiology, National Taiwan University Hospital, Taipei, Select a Country; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei 100, Taiwan; 3) Graduate Institute of Medical Genomics and Proteomics, College of Medicine, National Taiwan University, Taipei 100, Taiwan; 4) Graduate Institute of Clinical Medicine, College of Medicine, National Taiwan University, Taipei 100, Taiwan.

Malignant hyperthermia (MH), also known as malignant hyperpyrexia, is a potentially fatal pharmacogenetic disorder triggered by certain drugs used for general anesthesia, such as inhalation anesthetics and/or succinylcholine. In susceptible individuals, these medications can induce skeletal muscle calcium regulation abnormality, and subsequent uncontrolled skeletal muscle hypermetabolism, which overwhelms the body's capacity to regulate body temperature and eventually leads to circulatory collapse and death if not treated quickly. Its unpredictability and high mortality make MH a disease urgently needing a predicting tool, such as a reliable, cost-effective genetic diagnostic test. Up to now, two known genes and at least 4 other loci can cause malignant hyperthermia, and as a whole only account for proximately 50-70% of cases. To make the situation more troublesome, the known causative genes, *RYR1* and *CACNA1S*, are both huge with numerous exons, which pose severe difficulties for genetic testing using traditional technology. Next-generation sequencing (NGS) for research and for clinical applications is revolutionizing genetics in recent years, and in this study we applied NGS to genetic testing of MH. We designed a capture-based targeted enrichment panel using the Roche NimbleGen SeqCap EZ Choice library covering the *RYR1* gene (chr19:38,924,340-39,078,204, hg19) and the *CACNA1S* gene (chr1:201,008,640-201,081,694, hg19). Data cleaning, filtering, analysis and verification were performed using BWA, SAMtools, Picard, GATK, ANNOVAR and IGV. Sanger sequencing was used for confirmation. Six clinically diagnosed MH patients were tested, and we identified the causative variants for four (66. 7%) of them. The identified causative *RYR1* variants were c. 7373G>A (p. R2458H) (two MH patients), c. 7304G>A (p. R2435H) and c. 6502G>A (p. V2168M), respectively. Four additional individuals (two with osteogenesis imperfecta and one with muscular dystrophy with complicated anesthesia courses, and one normal control) were also tested, and none of them carried potential causative variants at *RYR1* or *CACNA1S*. For those two MH cases without identifiable causative variants by the NGS-based test, we performed Sanger sequencing covering the coding regions of *RYR1* and *CACNA1S* to make sure that no variant was missed. In conclusion, the NGS-based method provides a sensitive, reliable, fast and inexpensive genetic diagnostic platform for MH genetic testing.

1936W

An integrated method for extraction of high-molecular-weight DNA and preparation of genomic sequencing libraries using agarose gels. *C. Boles, E. S. Abrams, D. Yun, T. Barbera.* Sage Science Inc. , Beverly, MA.

Purpose: Several new DNA sequencing methods can provide read lengths greater than 10,000 base pairs (bp). This trend has motivated researchers to reevaluate standard DNA sample prep methods which usually produce genomic DNA fragments of only up to 50,000 bp in length. To this end, we have developed a new workflow and equipment for rapid purification of high-molecular-weight (HMW) DNA (>50,000 bp) from whole blood and cell suspensions. The system can also be adapted to perform both DNA extraction and transposase-mediated sequencing library preparation in a fully integrated, automated process. **Methods:** Our method is loosely based on the original agarose-based sample preparation method of Schwartz and Cantor (1984). We embed cell suspensions in agarose, followed by lysis and enzymatic processing steps which are carried out in the sample-embedded gel. However, we dramatically accelerate the process by using electrophoresis to remove contaminants, and to exchange DNA processing reagents. Electrophoretic processing is enabled by the fact that the genomic DNA produced in our process is many megabases (mb) in size, and cannot be electrophoresed out of the sample gel. After purification and processing, the DNA is lightly and randomly cleaved with a non-specific nuclease, allowing the processed DNA sample to be electroeluted from the gel into a liquid-filled elution chamber. **Summary of Results:** Genomic DNA from whole mammalian blood can be electrophoretically purified in less than 30 minutes. Several random cleavases can be used to release the HMW DNA from the sample gel. Depending on the fragmentation conditions used, it is possible to recover purified HMW DNA fragments ranging from 50,000bp to 800,000bp in size. In addition, the post-fragmentation electroelution process can be used to perform size selection on the DNA products. Finally, we show that transposasome complexes carrying sequencing adapters can be used for the fragmentation step, thereby integrating DNA extraction, purification, and library construction into one automated workflow.

1937T

Accessing the Full Spectrum of Polymorphisms in HLA Class I & II Genes without Imputation for High Throughput Disease Association and Evolutionary Research. *S. Ranade[1], W. Lee[1], J. Harting[1], K. Eng[1], L. Hepler[1], B. Bowman[1], N. Westerink[2].* 1) Pacific Biosciences, Menlo Park, CA; 2) GenDx, Utrecht, The Netherlands.

The HLA system plays central role in the mechanisms governing the outcomes of disease, health and survival. Specifically the HLA class I genes HLA-A, -B, -C, and class II genes HLA-DR, -DQ, and -DP play critical role as factors responsible for organ transplant rejection and are important targets for clinical and drug sensitivity research. They have been associated with more than 100 diseases by direct or linkage-based association through extended LDs in the MHC region. However, due to their highly polymorphic diversity originating from exonic combinations as well as recombination events, HLA gene sequencing has always been a challenge. A large number of new alleles are expected to be encountered when HLA genes are sequenced in their entirety, emphasizing the need for allele-level resolution. Until now, technological challenges and lack of complete understanding about the impact of polymorphisms outside of peptide binding domains of HLA genes, most of the studies have remained exon centric and focused on these regions. To access the full spectrum of polymorphisms across the entire gene or complete length of the amplicon sequenced, without limitation from sequencing technology read length or availability of reference sequence, we have used, Single Molecule, Real-Time (SMRT®) technology from PacBio to interrogate the HLA system. In the present work, we employed DNA barcode tagged SMRTbell™ sample preparation approach to multiplex 96 samples for sequencing all of the HLA class I and II genes. Commercially available NGSgo® reagents (GenDx) for full-length HLA Class I and relevant exons of class II genes were amplified for high-resolution HLA sequencing. Sequence analysis was performed with PacBio® LAA software and the NGSengine® (GenDx) software package was used for HLA typing. In this study, PacBio sequencing was found to be a highly effective tool for generating accurate, fully phased, allele-specific sequences of full-length alleles of HLA genes without imputation. With long read lengths (average >10 kb) and concensus accuracy exceeding 99. 999% (Q50), we were able to confidently generate four field typing information and take a comprehensive snapshot of variants in exons as well as non-coding regions that may have impact on gene regulation and provide invaluable genomic insights to advance causality association research.

1938F

Comparison of third-generation sequencing with Sanger for HLA haplotyping in the context of a clinical study. *E. Palescandolo[1], C. Van Hove[1], W. Talloen[1], I. Lonjon-Domanec[2], J. Aerssens[1].* 1) Janssen R&D, Beerse, Belgium; 2) Janssen EMEA Medical Affairs, Paris, France.

HLA typing is routinely accomplished via Sequence-Based Typing (SBT, Sanger being the gold standard), but requires a high number of reactions to resolve uncommon types. The development of massively parallel sequencing combined with the recent improvements in read lengths holds promise to improve HLA typing efficiency and accuracy which might have impact in clinical settings. We compared the performance of SBT (Sanger) versus Pacific Biosciences' (PacBio) long-range single molecule sequencing method (SMRT) in the context of a clinical study that required assessing the correlation between rash reactions and HLA haplotype. The aims of our study were to 1) assess the feasibility of HLA typing with the SMRT sequencing technology, 2) determine its accuracy compared to SBT and 3) evaluate potential association between HLA and rash in the clinical study. In Telaprevir's HEP3002 early access program, patients chronically infected with HCV genotype 1 and with severe fibrosis or compensated cirrhosis were treated with Telaprevir in combination with peg-interferon and ribavirin (PR) for 12 weeks followed by 12 or 36 weeks of PR. We selected 33 patients suffering severe rash, grade 3 or 4 and 66 matched control patients without rash. Patients' DNA underwent Sanger and PacBio sequencing for HLA-A, -B, -C, -DRB1, -DQB1. For full haplotype sequencing we used published primers able to amplify a full locus in either one or two (overlapping) PCRs. Fisher's exact test was utilized to evaluate the association between HLA and rash occurrence. Compared to the current gold standard SBT, PacBio revealed correct alleles for 94% of the samples. The non-matching haplotypes were attributed as lack of typing in 2% (i. e. both alleles were not identified by PacBio), homozygous in 1% (i. e. PacBio found one allele whereas Sanger two) and actual mismatches 3%. Mismatches occurred because of SBT's inability to disambiguate alleles, lack of primer specificity or PCR allelic imbalance. In conclusion, 1) we demonstrated that PacBio can be used for HLA haplotyping and is superior to Sanger since it can resolve complex haplotypes mixtures and determine types up to the field four; 2) we found no association between rash susceptibility and Telaprevir based therapy.

1939W

Chromosomal microarray as a clinical diagnostic test for undiagnosed rare disease. *H. Kim[1,2,3], H. Lee[2], HJ. Kim[1], E. Cho[4].* 1) Dept. of Ped. Konyang Uninv. College of Med. Daejon, South Korea; 2) Medical Genetics, Ajou Univ Med Sch, Seoul, South Korea; 3) Dept. of Genetic Counseling, Gachon Univ. Gil Med. Center, Incheon, South Korea; 4) Green Cross Genome, South Korea.

PurposeGenomic imbalance is known as a major cause of rare diseases. This study is aimed to evaluate the clinical utility of chromosomal microarray (CMA) as a diagnostic test for undiagnosed rare diseases patients. Materials and MethodsCMA performed on 179 samples of 108 patients and 71 family members from 37 patients in a single clinical laboratory referred by a clinical geneticist during 2. 5 years period (Dec. 2012-May. 2015). The results of 179 CMA were analyzed for CNVs and classified into pathogenic CNVs, variant of uncertain significance (VOUS), benign CNVs and normal. The comparison between patients and family member was made for the significance of CNVs and phenotypes. ResultsCMA results of 108 patients were revealed pathogenic CNVs in 34 cases (31. 5%), VOUS 10 cases (9. 3%), benign CNVs 16 cases (14. 8%) and normal 48 cases (44. 4%). One case of pathogenic CNV coexisted with VOUS and three cases with benign CNVs. CMA results of 71 family members showed pathogenic CNVs in 6 cases (8. 5%), VOUS 7 cases (9. 8%), benign CNVs 13 cases (18. 3%) and normal 45 cases (63. 4%). One out of 6 pathogenic CNVs coexisted with VOUS. 19 out of 71 family members showed abnormal phenotype, and 19 abnormal phenotypes were associated with 6 pathogenic CNVs (31. 6%), 3 VOUS (15. 8%), 3 benign CNVs (15. 8%) and 7 normal (36. 8%). Pathogenic CNVs included deletion (8 cases, 23. 5%), long contiguous stretches of homozygosity (LCSH) >5Mb (11 cases, 32. 4%), duplication (6 cases, 17. 6%), del/dup (5 cases, 14. 7%), aneuploidy (3 cases, 8. 8%), uniparental disomy UPD (1 cases, 2. 9%). Sites of genomic alteration occurred in all chromosomes but chromosomes 19, 20 and Y. Most common changes found in chromosomes 2, X, 3, 10 and 21 in frequency in order. An identical pathologic CNVs were found in 20 family members from 37 patients, ten of them were found to have abnormal phenotypes as well. Conclusion-This study of CMA analysis detected high rate of pathogenic CNVs (31. 5%) in 108 patients. CMA is especially useful tool as a diagnostic test of patients with normal karyotypes, yet clinically chromosomal aberration suspected. Our comparison study of CNVs found among patients and family members indicated the importance of family study with phenotype examination and pediagree analysis to be included for the accurate interpretation of CNVs found by CMA test of proband and appropriate genetic counseling.

1940T

Analysis of genetic variation and haplotypes of CYP2D6 amplicons by long-read nanopore DNA sequencing. *M. A. Kennedy[1], S. Cree[1], A. L. Miller[1], E. W. Chua[1,2], S. Maggo[1].* 1) Dept Pathology, University of Otago, Christchurch, Christchurch, Canterbury, New Zealand; 2) Faculty of Pharmacy, Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia.

The liver enzyme cytochrome P450 2D6 metabolises many drugs including antidepressants, anticoagulants and chemotherapeutics. Its gene (*CYP2D6*) is amongst the most variable of human genes, but current assays detect only the most common variants. We are striving to develop methods that allow more comprehensive analysis of *CYP2D6* variation. One technology we are exploring is the MinIon (Oxford Nanopore Technologies, UK). This tiny DNA sequencer uses nanopore sensing, whereby the passage of a single DNA strand through each pore generates changes in ionic current, allowing potentially very long sequence reads on single molecules. These characteristics mean the MinIon has potential for rapid analysis of long PCR amplicons, and for direct phasing of genetic variants. A preliminary report demonstrates that the MinIon can read a long CYP2D6 amplicon and detect genetic variation (Ammar *et al*. 2015; F1000Res, 4:p17). For our analysis *CYP2D6* was PCR amplified on a 6. 6kb amplicon, in order to avoid potentially misleading pseudogene sequences. Sequences were extracted from the MinIon HDF5 files generated by Metrichor analytical software (ONT), then aligned using Lastz. As kit chemistry has improved (from R6 to R7. 3) we have seen corresponding increases in mean length of aligned segments (792bp – 6431bp), proportion of 2D reads (4-12%), and proportion of bases in aligned regions (62%-84%). In general we see blocks of good data interspersed with poor data, and known variants can be discerned. Errors appear to be randomly distributed within reads. On our most recent runs, we have successfully used oligonucleotide indexing to carry out simple multiplexing of samples. Our conclusions are that it is indeed possible to read *CYP2D6* amplicon sequences with the MinIon, and variants can be detected, although not yet called *de novo* with high confidence. We have demonstrated multiplexed analysis using oligonucleotide indexing. Remaining questions we are now addressing are simpler workflows, increasing confidence in variant calls, establishing the degree of sample multiplexing we can achieve, and quantifying the degree to which template switching during PCR may give rise to misleading haplotype determinations. Nanopore sequencing of CYP2D6 is a promising approach to rapid, simple and cost effective analysis of this important pharmacogene, although still in need of optimisation.

1941F

Newborn Screening for Spinal Muscular Atrophy. *W. Hwu, Y. Chien, S. Chiang, L. Hsu, N. Lee.* National Taiwan University Hospital, Taipei, Taiwan.

Background:Because of the needs for early diagnosis and early treatment of spinal muscular atrophy (SMA), newborn screening for SMA is attempted. Methods:Dried blood spot (DBS) samples were obtained from routine newborn metabolic screening samples taken at the age of day 3. DNA was extracted from a 3. 2mm punch from each DBS sample. A real-time quantitative PCR (qPCR) assay using primers and probes targeting IVS7+100 was conducted in 96-well plates. Newborns with a positive screening result, showing the absence of IVS7+100C, was requested for a confirmatory test, and at the same time, a second 3. 2mm punch from the original DBS was subjected to qPCR (with another set of primers and probes) and Droplet Digital PCR (ddPCR). The confirmatory process included clinical evaluation and a whole blood sampling for MLPA test. Results:From Nov. 17, 2014 to May 31, 2015, we have screened more than 46,000 newborns and have found 2 SMA patients. In addition, we experience 2 false-positive cases, one with borderline initial screening values, while the other suffered from a polymorphism at the targeted nucleotide (IVS7+100C>T). Therefore, the false-positive rate was low, and the positive-prediction rate was 50%. The theoretically false-negative rate of the current screening method was 5% because we will miss SMA patients caused by point mutations of the SMN1 gene. Conclusion:Newborn screening of SMA by qPCR method is feasible and can provide patients early diagnosis. Further screening on more newborns are necessary to provide the current incidence of SMA, and to explore the benefit of early treatment.

***CYP2D6* genotyping with the PacBio RSII using multiplex targeted long amplicon sequencing.** *S. J. White[1], H. Buermans[1], T. van der Straaten[2], R. Vossen[1], S. Y. Anvar[1], J. Swen[2], J. T. den Dunnen[1,3].* 1) LGTC and Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Department of Clinical Pharmacy & Toxicology, Leiden University Medical Center, Leiden, Netherlands; 3) Department of Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands.

CYP2D6 is among the most important enzymes involved in the metabolism of many routinely prescribed drugs. Specific variants in *CYP2D6* are associated with changes in the activity and amount of the CYP2D6 enzyme. *CYP2D6* genotyping is complicated because of homology with pseudogenes *CYP2D7/2D8,* and information on which variants are located on the separate alleles cannot easily be determined with standard assays. Targeted long amplicon sequencing using the PacBio RSII sequencing platform holds several advantages over second generation NGS systems, including the ability to sequence multi-kb amplicons without the need for fragmentation steps, obtain high accuracy consensus sequences, and deliver fully phased variant information for separate alleles. For optimal implementation of PacBio sequencing for variant profiling, efficient sample barcoding strategies are needed. Standard methods introduce sample barcodes either via ligation of barcoded SMRTbell adapters to amplicons, or via PCR, by using a set of fusion primers linking the barcode sequence directly to the locus specific primer. Downsides of these barcoding methods are that they are expensive and lack flexibility. We devised a PCR-based multiplexing strategy for long amplicon variant profiling using the PacBio RSII platform. A 6. 6 kb DNA fragment containing *CYP2D6* is first amplified with a pair of gene-specific primers containing forward and reverse M13 sequence tails. A sample barcode is subsequently introduced in a second PCR using a set of re-usable M13-tailed barcode primers. Barcoded samples are then pooled in equimolar amounts and processed for PacBio sequencing. Using this setup, we sequenced the complete *CYP2D6* gene for 12 individuals with previously established *CYP2D6* genotypes, called with the (now discontinued) AmpliChip CYP450 Test. Full length *CYP2D6* sequences were obtained for all individuals. Predicted genotypes from the PacBio RSII were in agreement with the previous calls. Two different allele sequences were evident in four individuals, providing the exact distribution of multiple heterozygous SNPs over the two separate *CYP2D6* alleles. In conclusion, we have developed a versatile and simple multiplexing method for obtaining fully phased allele sequences for *CYP2D6*. With minor modifications this method could be applied for targeted long amplicon sequencing of other loci.

1942W

1943T

Mutation screening of the SLC26A4 gene in Brazilian non-syndromic sensorineural prelingual deaf individuals. *S. C. S. Carvalho[1,2], C. H. P Grangeiro[1,4], C. G. P. Albuquerque[1,4], T. O. Anjos[2], J. R. Plaça[2,3], L. F. Araujo[1,2], B. R. Muys[1,2], E. M. Souza[1,2], G. A. Molfetta[1,2], W. A. Silva Jr[1,2,3], V. E. F. Ferraz[1,4].* 1) Department of Genetics, Ribeirão Preto Medical School, Ribeirão Preto, São Paulo, Brazil; 2) Center for Genomic Medicine - University Hospital of Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil; 3) Center for Cell-Based Therapy (CEPID/FAPESP); National Institute of Science and Technology in Stem Cell and Cell Therapy (INCTC/CNPq), Regional Blood Center of Ribeirão Preto, Ribeirão Preto, São Paulo, Brazil; 4) Department of Medical Genetics of University Hospital of Ribeirão Preto Medical School, São Paulo University, Ribeirão Preto, São Paulo.

Hearing is the main source for speech and language learning during childhood, and deafness or auditory deprivation can result in both emotional and social difficulties to those affected individuals. Approximately 360 million people in the world are affected by hearing loss, and hereditary deafness is the most common sensorineural disorder in humans with a prevalence of 1 per 1,000 live births. In most cases, inherited sensorineural hearing loss is heterogeneous and has a large number of genes involved. Studies have shown the important role of the SLC26A4 gene in the physiology of the inner ear, and mutations in this gene have been reported as a cause of hereditary hearing loss. Thus, the aim of this study was to investigate the prevalence of mutations in coding regions of SLC26A4 in non-syndromic prelingual sensorineural hearing loss (SNHL) patients assisted by the Department of Medical Genetics of University Hospital of Ribeirão Preto Medical School. To achieve this goal, were analyzed 79 DNA samples from unrelated deaf individuals, which were previously screened to GJB2/GJB6 mutations and had an inconclusive molecular diagnosis. Samples were analyzed by the High Resolution Melting (HRM) screening technique in order to detect alterations in melting curve patterns, which were validated by Sanger sequencing. The prevalence of mutations in SLC26A4 was 27. 8% (22/79) with 10 mutations found in this gene. Three of these mutations are previously undescribed mutations: p. Gly139Arg, p. Ile254Val, p. Asn382Lys. Among these 10 mutations, the most prevalent was chr7:g. 107301238C>G on exon 1 (the SLC26A4 promoter region), found in 11 deaf patients (13. 9%) in heterozygosity, as well as in compound and double heterozygosity (with GJB2) genotypes. The 1000 Genomes Project data suggests it as a polymorphism, but pathogenicity information on this mutation remains inconclusive. Four of these 10 mutations found in SLC26A4 have shown pathogenic scores by *in silico* analyses: p. Gly139Arg; p. Ile300Leu; p. Asn324Tyr and p. Asn. 382Lys; suggesting that these four mutations may constitute the cause of genetic deafness in 8. 9% of the patients. This data highlights the importance of the SLC26A4 gene for molecular diagnosis of hereditary hearing loss and gives strength to its potential contribution to the genetic counseling process. However, it is suggested that functional tests and segregation studies are needed in order to elucidate the role of SLC26A4 mutations to the phenotype.

1944F

Is Sanger Sequencing Still the Gold Standard? *T. F. Beck[1], N. Hansen[2], J. Mullikin[1,2], L. Biesecker[1].* 1) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) NIH Intramural Sequencing Center, Rockville, MD.

Next-generation sequencing (NGS) costs are rapidly decreasing, and more clinical and research labs are using NGS for many types of evaluations. We focus on our experience in clinical translational genomics research. While some of these data are generated purely for research, it is often the case that medically relevant variants can be identified which we would like to return to the participant. Currently, the standard of care is to verify variants using Sanger sequencing before returning the results, which can be costly and time-consuming. We decided to evaluate the utility of Sanger validation given the current quality of NGS. In this study, we compare the variants found using exome sequencing on a set of five medically-relevant genes in over 250 participants against those found using Sanger sequencing data from the same cohort, namely the ClinSeq® cohort. Out of over 5,600 variants found in these participants using exome sequencing, only two variants were not validated by Sanger sequencing, and those two variants had exceedingly low most probable genotype (MPG) scores (4 and 10, respectively). Our findings indicate that Sanger validation may be unnecessary in a clinical setting, given the accuracy of current next-generation sequencing technologies.

1945W

A 4,500 Gene, Short-turn-around-time, NGS Neonatal Intensive Care Diagnostics Panel. *S. Dames[1, 3], D. Nix[1], C. Miller[1], S. Pew[1], C. VanSant Webb[1], J. Stocks[1], R. Mao[1, 2].* 1) ARUP Laboratories Institute for Clinical and Experimental Pathology, Salt Lake City, UT; 2) University of Utah Department of Pathology, Salt Lake City, UT; 3) University of Utah Department of Biomedical Informatics, Salt Lake City, UT.

Introduction: Rapid turnaround for NICU patients is critical for diagnosis and treatment, as well as reducing costs associated with NICU hospitalization. IDT has developed an inherited disease panel comprised of 4,500 genes—most of which have mutations described in HGMD. Using the Illumina NextSeq 500 platform 2x150 mid-throughput kit, we have developed a STAT, 4,500 gene panel. Proband and parents, when available, are sequenced, but only variants associated with the proband are reported. Our goal is a 7-10 day turnaround from receipt of specimen to preliminary report. **Methods:** Illumina libraries are prepared on a Beckman-Coulter SPRI-TE. Samples are hybridized between 4-16 hours. Post hybridization wash and PCR clean up are performed using an Agilent Bravo A. All other library prep and quantification steps are performed manually. Samples were sequenced using version 1 and 2 chemistries on the NextSeq 500. A UCSC bed. file (+/- 10 bp CDS) was derived using HGMD reference transcripts for each gene (when defined) or HUGO reference transcripts for genes. An intersect between the IDT and UCSC bed files was preformed since all exons are not captured in the IDT panel for a subset of genes. A total of 54,517 targeted regions are queried. The ARUP validated clinical pipeline was comprised of BWA, SAMtools, GATK, Freebayes, and snpSIFT/eff. Average processing time from sample to fastq is approximately 72 hours, with alignment and variant analysis an additional eight hours. **Results:** Samples NA12878 and ARUP's wild-type female/male controls were sequenced. Initial experiments using version 1 chemistry displayed greater than 99. 5% targeted coverage > 15-fold (Q-score >30 73%). NextSeq 500 background was higher compared to the HiSeq and MiSeq (up to 3% error in read 2) despite optimum cluster density. While this may adequate for germline mutations, it will be problematic for somatic or mtDNA assays. Subsequently the samples were run with version 2 chemistry and displayed greater than 99. 5% targeted coverage, Q-score >30 at 90%, with a flat error rate of <0. 5 percent. Concordance in raw vcf variant calls was > 99% for all three samples between versions 1 and 2. **Conclusions:** Preliminary experiments demonstrate that our proposed timeline for bench work can easily be met; however, a dedicated server is required for optimal turnaround for STAT tests. Despite the quality difference between versions 1 and 2 chemistries, raw vcf variants are greater than 99% concordant.

1946T

Clinical applicability of direct-to-consumer exome analysis. *J. Vengoechea[1], MP. Alfaro[1,2], J. Kelsay[1].* 1) Division of Genetics, University of Arkansas for Medical Sciences, Little Rock, AR; 2) Arkansas Chidlren's Hospital, Little Rock, AR.

A 30 year-old-man presented to the Adult Genetics clinic with results from a direct-to-consumer (DTC) next generation sequencing exome test. The patient pursued testing to explain his fatigue after taking medication for attention deficit/hyperactivity disorder. He had no other chronic conditions. The patient brought the results as VCF files. Cartagenia Bench Lab NGS 4. 0 was used for analysis. There were 55,645 variants from the whole exome; these were restricted to variants in genes with an OMIM phenotype leaving 15,795 variants. None met standard quality criteria used by our lab. After loosening the quality criteria to only include variants with ≥20X coverage, 14,125 variants remained. Next, variants with population frequency ≥ 3% in 1000 Genomes, dbSNP and/or ESP6500 were excluded. The remaining 1,631 variants were filtered to include only those that altered protein length, were reported in ClinVar or HGMD as pathogenic or were predicted to be pathogenic by 4 out of 5 *in-silico* prediction tools. The goal was a specific but not sensitive analysis. This left 58 variants for manual review. After excluding genes that caused syndromes that the patient clearly did not have, heterozygous variants for conditions with autosomal recessive inheritance and known benign variants, there were three variants of interest. The first was a frameshift variant in *MYF6*, a gene causative of centronuclear myopathy type 3. The variant was confirmed by a CLIA-certified lab. Only one affected family has been reported previously. The patient's CK level was normal. A muscle biopsy is possible, but the description of this syndrome is too limited to make formal recommendations. The second was a missense variant reported in HGMD and ClinVar in *PROS1,* the gene that encodes protein S. The patient had normal protein S activity. The third was a missense variant in *PKD1* predicted to be pathogenic by *in silico* tools. A renal ultrasound was ordered. This project showed that although it may be feasible to use DTC exome data to perform a clinical analysis, the great number of variants necessarily limits the analysis to known pathogenic mutations in autosomal dominant conditions while relying on clinical databases. Feasibility is further limited because there is no code to bill for the analysis of DTC results. The DTC test and our analysis did not answer our patient's chief complaint, and the low cost of DTC testing may be offset by the cost of confirmatory clinical studies.

1947F

Effect of Formalin Fixation on Targeted Sequencing and Genomic Analysis in Cancer. *R. K. Alla[1], E. Helman[1], M. J. Clark[1], S. M. Boyle[1], S. Luo[1], S. Kirk[1], P. Sripakdeevong[1], M. Karbelashvili[1], D. M. Church[1], M. Snyder[2], J. West[1], R. Chen[1].* 1) Personalis Inc, Menlo Park, CA; 2) Department of Genetics, Stanford University, Stanford, CA.

Tumor biopsies are often Formalin-Fixed and Paraffin-Embedded (FFPE) for histological staining, genetic testing and archival purposes. Formalin treatment preserves tissue by crosslinking proteins, but also leads to mutation of the nucleic acid bases and poses a challenge to identification of true variants in the tumor using next-generation sequencing (NGS) methods. Studies have shown that it is possible to isolate high quality nucleic acids from good FFPE samples and to profile small variants using NGS. These studies used tissue fixed with 10% neutral buffered formalin for 24 hours, a standard protocol in the pathology field. In our initial handling of FFPE samples we found that the quality of the isolated DNA and subsequent sequencing results vary widely. We hypothesized that this may be due to deviations from the standard protocol, such as inaccurate logging of the fixation protocol, variation in the fixation time, and varied storage conditions of the samples. To understand the role of formalin fixation on the quality of variants called, we performed an augmented target enrichment and sequencing assay on fresh frozen (FF) and FFPE samples from different sources. We assessed raw DNA quality, library quality, sequencing metrics (alignment rate, duplication rate, on-target efficiency, etc.) and variant concordance profiles between FF and FFPE. Preliminary results showed that formalin fixation causes an increase in global CG to TA mismatch rate from 0. 025% to 0. 04%. This led to additional variants detected at low (<5%) allele frequencies (AFs). We observed that about 45% of low AF mutations in FFPE samples were CG to TA transitions. We also noted that the severity of molecular damage across different tissue types correlated with the pathology lab preparing the FFPE sample. To directly test the effects of formalin, we subjected tissue to different FFPE protocols and compared their sequencing profiles to those of adjacent FF tissue. We varied fixation protocols to reflect the conditions from different pathology labs, adjusting formalin concentration, fixation time and storage conditions. Using the results of this study, we intend to develop fixation protocols optimized for sequence analysis to facilitate tumor clinical molecular profiling. We also demonstrate how a deeper understanding of the effects of formalin can improve sequencing analysis results from formalin-fixed tissues, especially at lower AFs where formalin-related errors have the greatest impact.

1948W

Defects in *TANGO2* cause episodic muscle weakness, rhabdomyolysis and cardiac arrhythmia. *J. A. Rosenfeld[1], S. R. Lalani[1], P. Liu[1], C. Miyake[2], M. Shinawi[3], L. Emrick[4], S. Schelley[5], M. K. Koenig[6], N. Memon[7], T. Chiang[1,8], T. Gambin[1], M. K. Eldomery[1], S. N. Jhangiani[1,8], N. Hanchard[1], J. S. Orange[9,10], L. B. Watkin[9,10], D. M. Muzny[1,8], H. Northrup[11], C. Bacino[1], F. Scaglia[1], P. E. Bonnen[1], J. Duis[12], G. H. B. Maegawa[12], B. Graham[1], J. McGill[13], C. M. Eng[1,14], F. Xia[1,14], J. R. Lupski[1,15], R. A. Gibbs[1,8], Y. Yang[1,14].* 1) Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Pediatric Cardiology, Texas Children's Hospital, Houston, TX; 3) Division of Genetics and Genomic Medicine, Washington University School of Medicine, St. Louis, MO; 4) Department of Neurology, Baylor College of Medicine, Houston, TX; 5) Division of Medical Genetics, Stanford School of Medicine, Stanford, CA; 6) Division of Child & Adolescent Neurology, Department of Pediatrics, The University of Texas Health Science Center at Houston, TX; 7) Division of Cardiology, DCH Regional Medical Center, Tuscaloosa, AL; 8) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 9) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 10) Texas Children's Hospital Center for Human Immuno-Biology, Houston, TX; 11) Division of Medical Genetics, Department of Pediatrics, The University of Texas Health Science Center at Houston, TX; 12) McKusick-Nathans Inst. of Genetic Medicine & Department of Pediatrics, The Johns Hopkins University School of Medicine, Baltimore, MD; 13) Department of Metabolic Medicine, Lady Cilento Children's Hospital, South Brisbane, Queensland, Australia; 14) Baylor Miraca Genetics Laboratories, Baylor College of Medicine, Houston, TX; 15) Department of Pediatrics, Texas Children's Hospital, Houston, TX.

Whole-exome sequencing (WES) is facilitating an increasing pace of human disease gene discovery. Leveraging the power of extensive genotype and phenotype data from a large cohort of individuals undergoing clinical WES (n>5500), we identified 12 individuals with bi-allelic defects in *TANGO2* [transport and golgi organization 2 homolog (*Drosophila*)] who shared clinical features including episodic muscle weakness with rhabdomyolysis, cardiac arrhythmia, intellectual disability, and seizures. Two individuals carried a homozygous loss-of-function variant, one nonsense and one affecting a splice site. Five individuals, all of Hispanic ancestry, were homozygous for the same missense variant and had evidence of a shared haplotype. Control data show this variant to be most common in the Latino population, with no homozygotes detected, and the frequency of homozygotes in our cases is enriched compared to simple population genetics predictions. The eighth individual carried the common missense variant on one allele and a large intragenic deletion on the other allele. The remaining four individuals were homozygous for large deletions affecting several *TANGO2* exons. The large deletions were identified by comparing normalized depth of coverage from clinical WES samples with that from individual patients. Following informed consent from a majority of these individuals, an in-depth phenotypic analysis showed a consistent phenotype among those with bi-allelic variants (sequence and copy number): recurrent metabolic crises frequently triggered by illness, progressive neurologic disease, and life-threatening cardiac arrhythmia. Two families had multiple affected siblings. Therefore, phenotype, segregation, and control data support bi-allelic pathogenic mutation of *TANGO2* as a cause of human disease. It has been reported that TANGO2 deficiency may cause morphology changes of golgi and ER membrane in *Drosophila* S2 cell lines, although specific functions of TANGO2 are poorly characterized. Preliminary cellular studies using tissue from a patient with the homozygous Hispanic missense variant suggested decreased TANGO2 expression, decreased golgi size, and increased ER stress. These results demonstrate the strength of large cohorts to facilitate rare disease discovery, particularly in minority populations, in addition to the importance of analyzing WES data for potentially pathogenic copy number variations.

1949T

Copy number variation detected by CMA- A review of 2000 cases. *Y. Hadid.* Genetics, Bnai Zion, Haifa, Israel.

Chromosomal microarray analysis is a method of measuring gains and losses of DNA (CNVs) throughout the human genome. It is considered to be a first-line test in the genetic evaluation of infants and children with unexplained intellectual disability, congenital anomalies, multiple birth defects, or autism spectrum disorder. The test is also recommended for further investigation of prenatal testing in pregnancies at risk: advanced age of mother, abnormal U. S finding, or pregnancy loss. Here we describe CNV results from 2000 diagnostic cases tested in the genomic lab (Bnai Zion medical center, Haifa, Israel). The samples are divided into three categories: Prenatal, Postnatal and tested parents for further investigation. The main conclusions from our experience about **postnatal** samples are that detection of pathogenic CNV by CMA method is relatively higher percentages more than normal karyotyping method. For **prenatal** testing, broader CMA testing will undergo a higher identification of chromosomal alterations clinically relevant, even at low-risk pregnancies. Our findings provide strong evidence for the entry of CMA diagnostic tests first line to all pregnant women undergoing invasive prenatal tests, regardless of risk factors.

1950F

Contributions of microarray data to the molecular diagnosis in 5700 consecutive clinical whole-exome sequencing (WES) cases. *W. Bi[1], P. Liu[1], J. Scull[1], H. Cui[1], T. Chiang[2], A. Hawes[2], Y. Ding[2], S. Matakis[3], D. Muzny[2], R. A. Gibbs[1,2], A. L. Beaudet[1], J. R. Lupski[1], A. Patel[1], C. M. Eng[1], Y. Yang[1].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, 77030; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, 77030; 3) Baylor Miraca Genetics Laboratories, Baylor College of Medicine, Houston, TX 77021.

Whole-exome sequencing has been utilized in the molecular diagnosis of patients referred for evaluation of suspected genetic conditions. The disease causative variants reported in exome cases are usually single nucleotide variants (SNV) or small indels. The contribution of copy number variants (CNV) and uniparental disomy (UPD) to the molecular diagnosis in clinical WES cases is not well understood. Here, we analyzed microarray data for ~5700 consecutive primarily pediatric patients tested by WES in the Baylor clinical genetics laboratory between 2011 to 2015; a molecular diagnosis was found in 1539 cases (27%). HumanOmni1-Quad array or HumanExome-12 v1 array was performed on each patient as a quality control measure. In addition, concurrent chromosomal microarray analysis using a high resolution targeted clinical array with SNPs was requested for 307 patients. Disease-associated microarray findings were evaluated based on the clinical symptoms provided by the referring physicians to determine whether they are causative or not. Patients' phenotypes can be fully explained by microarray results in 42 cases, corresponding to 2. 7% (42/1539) of molecular diagnosed cases, which consist of microdeletion/duplication syndromes such as SMS and WBS or other disease-associated large >1 Mb deletions/duplications except for 3 UPD syndromes, 1 monosomy X, and 4 small <0. 5 Mb deletions involving disease genes. Phenotypes can be partially explained by array findings in 51 cases, of which in 32 cases, WES was pursued for potential additional molecular diagnoses beyond CNVs. Therefore, causative findings that can explain either part or the entire clinical findings were present in 93 cases (1. 6%, 93/5700; 6%, 93/1539). Disease-causing CNVs that are likely to be irrelevant to the phenotypes include 28 CNVs with reduced penetrance, a 15q11q13 duplication and a STS deletion. In addition, CNVs >1 Mb in non-disease associated regions were detected in 17 cases, with one CNV de novo, two CNVs inherited and the others without parental study. For autosomal recessive diseases, molecular diagnoses were obtainedfor five cases with a pathogenic SNV in one allele and a deletion involving the other allele of the disease genes. This study showed that microarray data contribute to 2. 7% to 6% of molecular diagnosis of whole exome sequencing and demonstrated that CNVs/UPD contribute to the diagnoses of autosomal dominant or X-linked diseases, as well as autosomal recessive diseases.

1951W

Next-generation sequencing and the molecular diagnosis of newborns in the neonatal intensive care unit. *H. Daoud[1], S. Luco[1], R. Li[2], O. Jarinova[1], N. Carson[1], SM. Nikkel[1,3], GE. Graham[1,3], J. Richer[1], C. Armour[1], SL. Sawyer[1,3], J. Majewski[2], KM. Boycott[1,3], DA. Dyment[1,3].* 1) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) McGill University and Genome Quebec Innovation Centre, Montréal, Québec, Canada; 3) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ontario, Canada.

Genetic diseases often present in the first days of life and are associated with significant morbidity and mortality. Recent studies have emphasized the success of Next Generation Sequencing (NGS), notably Whole Exome Sequencing (WES), for the diagnosis of rare genetic disorders but its effectiveness in critically ill neonates has not been thoroughly investigated. Here, we report the application of NGS in a cohort of twenty newborns with congenital anomalies, ascertained prospectively and retrospectively, and admitted to the Neonatal Intensive Care Unit (NICU). NGS of a comprehensive panel of genes (Trusight One Sequencing panel, Illumina) with known associated clinical phenotypes was performed in the patients and their parents. We identified disease-causing mutations in seven of the twenty newborns, achieving a 35% molecular diagnostic rate. Diagnoses included bi-allelic mutations in *CHRND*, *FTO*, *WDR19*, *ACE*, an X-Linked mutation in *MTM1*, and *de novo* mutations in *SCN1A* and *DYRK1A*. The time to diagnosis varied considerably, though a return of results was possible within a 2-3 week timeframe. These results show the potential of NGS to deliver accurate and quick molecular diagnoses in the NICU, thereby eliminating the diagnostic odyssey that many of these patients historically underwent. A rapid accurate diagnosis in the newborn period may also significantly impact patient management, and in some cases alter treatment decisions, thereby improving care for patients with rare genetic disorders. The success of this project highlights the need to translating NGS in the NICU into routine clinical care.

1952T

Clinical exome sequencing: accelerating the pace of diagnosis and gene discovery for improved patient management. *C. Eng[1,2], D. Muzny[1,3], F. Xia[1,2], M. Walkiewicz[1,2], R. Person[1,2], Z. Niu[1,2], Y. Ding[2,3], P. Ward[1,2], A. Braxton[1,2], W. Bi[1,2], N. Veeraraghavan[3], T. Chiang[3], J. Posey[1], J. Rosenfeld[1], M. Leduc[1,2], R. Xiao[1], P. Liu[1], H. Cui[1,2], W. He[2], F. Vetrini[2], J. Zhang[2], C. Buhay[3], A. Hawes[3], W. Craigen[1], E. Boerwinkle[3], S. Plon[1,3,4], J. Lupski[1], A. Beaudet[1,2], R. Gibbs[2,3], Y. Yang[1,2].* 1) Dept Molec & Human Genetics, Baylor Col Medicine, Houston, TX; 2) Baylor Miraca Genetics laboratories, Houston, TX; 3) Human Genome Sequencing Center, Baylor Col Medicine, Houston, TX; 4) Dept Pediatrics, Baylor Col Medicine, Houston, TX.

We developed and optimized technical, bioinformatic and interpretive exome sequencing pipelines in a CAP/CLIA certified laboratory to identify causative mutations underlying disease phenotypes in undiagnosed patients being evaluated clinically for genetic disorders. Several variations of the exome testing approach have evolved for specific clinical scenarios including critical exome and prenatal exome sequencing with expedited technical and analytical reporting, trio and quad exome sequencing, and customized focused panel sequencing derived from the exome. Using these approaches, clinical exomes have been completed for 5700 cases, most of which are pediatric patients with neurological manifestations, and achieved a molecular diagnosis rate of approximately 27%- 31% for known disease genes using proband or trio approaches, respectively. The inheritance patterns of the contributing Mendelian diseases included about 53% AD, 34% AR, and 12% X-linked. About 5% of the positive patients were diagnosed with two genetic disorders resulting in blended phenotypes. Genetic events such as *de novo* variants, mosaicism in patients or parents, copy number variants (CNVs) and uniparental disomy were discovered, further documenting the frequency of these events in routine clinic populations of patients suspected to have genetic disorders. Patients without a molecular diagnosis after initial reporting have several avenues to an eventual diagnosis. Our clinical testing consent process allows the option of entering research studies for novel disease gene discoveries such as the Centers for Mendelian Genomics and more recently, the Undiagnosed Diseases Network. These research mechanisms as well as internal observations in the clinical lab have led to reports of new disease genes such as *ASXL3*, *MAGEL2*, *AHDC1* and *PURA* as well as candidate disease genes such as *MIPEP, GMNN, and TANGO2*. Moreover, we perform re-analyses of likely pathogenic changes in new disease genes for previously reported cases without a definitive molecular diagnosis. The combined efforts of data re-analyses and research activities resulted in additional molecular diagnoses and improved our exome diagnosis rate from 25% to approximately 30% over time. Our four year experience with optimizing clinical exome sequencing has demonstrated utility in different clinical settings as well as providing the basis for a multi-directional effort to continue to improve diagnostic rates through research and discovery.

1953F

Diagnostic Exome Sequencing (DES) provides a diagnosis for 23% of adult patients. *K. D. Farwell Hagman, S. Li, L. Shahmirzadi, D. El-Khechen, Z. Powis, C. Gund, K. Burk, S. Tang.* Ambry Genetics, Aliso Viejo, CA.

Diagnostic exome sequencing (DES) is successful in solving the diagnostic odyssey for roughly 30% of undiagnosed patients with a broad range of underlying Mendelian disorders. Pediatric neurologic manifestations are the most common DES referral indication. However, DES has also been successful in diagnosing adult patients undergoing the diagnostic odyssey. Among the first 1000 reported DES families, 175 (17. 5%) adult patients (>18 years old) were referred for testing. Significant family history was more common among adult patients (81. 1%) than children (65. 3%) (p=1. 826e-5). Among adults tested, 79 (45. 1%) were between 18-29 years old, 50 (28. 6%) were between 30-49 years old, and 46 (26. 3%) were between 50-79 years old. The most common referral indication among adult patients was cancer susceptibility (25, 14. 3%), followed by multiple congenital anomalies (23, 13. 1%), neurodevelopmental anomalies (15, 8. 6%), skeletal anomalies (14, 8. 0%), and ataxia/spasticity (10, 5. 7%). Positive findings were uncovered in 23. 4% of these families (41 of 175), compared to 29. 3% of children (242 of 825). A novel genetic etiology was proposed for 9 families (5. 1%). The diagnostic rate among patients <30 years old (29. 5%) is significantly higher than among patients greater than 30 years old (16. 7%) (p=4. 125e-3). Among the 41 adults with positive results, over half of the gene findings (23, 56. 1%) were autosomal dominant, 12 (29. 3%) were autosomal recessive, and 6 (14. 6%) were X-linked. Overall, the rate of *de novo* and likely *de novo* findings among adult patients (17. 5%) was much lower than among children (44. 4%) (p=5. 278e-5). The highest diagnostic yields were observed among patients with ocular anomalies (3/3, 100%), integumentary anomalies (2/3, 66. 7%), muscular dystrophy (3/5, 60%), intellectual disability (4/7, 57. 1%), renal anomalies (1/2, 50%), and cardiovascular anomalies (6/15, 40%). These data, including the high DES diagnostic rate of 23% among adult patients, highlight the potential medical-economics savings to patients as well as payers given that achieving a diagnosis abrogates the expensive, time-consuming, and often invasive diagnostic odyssey.

1954W

Clinical Diagnostic Yield for Autism and X-Linked Intellectual Disability Mutation Detection by Targeted Panels versus Whole Exome Sequencing: the GGC Experience. *J. Lee, M. Jones, S. McGee, K. Kubiak, F. Abidi, J. Jones, M. Basehore, M. Friez.* Molecular Diagnostic Lab, Greenwood Genetic Center, Greenwood, SC.

With the advent of Next Generation Sequencing (NGS), it is now possible to interrogate hundreds to thousands of genes at once for a given patient, increasing the potential diagnostic yield per assay. At the Greenwood Genetic Center, routine clinical diagnostic testing for syndromic autism and X-linked intellectual disability (XLID) is currently performed using targeted NGS panels that are specifically enriched for 83 and 114 genes, respectively. Here, we discuss our experience with diagnostic testing for autism and XLID by targeted panels versus by whole exome sequencing (WES). Since the inception of offering these targeted panels, our detection rate has been approximately 5-10% for each. This relatively low detection rate is likely due to the phenotypic diversity of patients and compounded by significant genetic heterogeneity. Therefore, WES has recently been pursued more frequently to diagnose these types of cases. Using this strategy we have begun to diagnose phenotypically diverse patients referred for autism, ID, or suspected XLID, that would not have been possible based on the content of our current targeted panels. Because targeted panels are limited in the genes that can be interrogated, WES offers the benefit of analyzing the entire exome, thus enabling mutation detection in newly associated disease-causing genes and potential new gene discovery. Overall, our diagnostic yield for WES is approximately 25%, much greater than the yield for our autism and XLID panels. However, certain limitations continue to exist for WES. Here, we review data from six randomly selected patients from independent WES runs, and focus on the autism and XLID genes that are covered on the targeted panels. We find that for the autism panel 99. 95%, and for the XLID panel 99. 99%, of exons are covered at a depth of 20X. However, for the respective exons by WES, the average coverage is 98. 57% and 97. 75%. Considering the span of coverage to ensure that all bases within the exons and 20 bases into the introns were covered in their entirety by at least 20X, greater than 99% of exons met these criteria for the panels, but it is closer to 90% by WES. We conclude that the diagnostic yield seems to be greater for these phenotypes by WES than by targeted panels, even though the coverage is not as complete for the most relevant genes. Our experience is consistent with others and supports the notion of using WES for patients with a broader clinical spectrum.

1955T

Clinical Exome Sequencing as a First-Line Molecular Diagnostic Test for Mendelian Disorders. *H. Lee[1], S. Strom[1], N. Dorrani[2], J. L. Deignan[1], N. Ghahramani[1], J. Jianling[1], R. Xian[1], S. Kantarci[1], F. Quintero-Rivera[1], K. Das[1], E. Vilain[2,3], W. W. Grody[1,2,3], S. F. Nelson[1,3].* 1) Pathology and Laboratory Medicine, Univ California, Los Angeles, Los Angeles, CA; 2) Pediatrics, Univ California, Los Angeles, Los Angeles, CA; 3) Human Genetics, Univ California, Los Angeles, CA.

The CLIA-certified and CAP-accredited UCLA Clinical Genomics Center launched the Clinical Exome Sequencing (CES) test in 2012. We routinely perform CES in pediatric and adult patients with a variety of clinical indications such as developmental delay presenting as part of a complex syndrome including co-morbid diagnosis of seizures, hypotonia or dysmorphic features, neuromuscular disorders, and ataxia. Causal variants have been identified across a variety of indications in both known and novel genes with a diagnostic rate of 27% (302/1100 cases). The conclusive molecular diagnostic rate was higher when CES was performed on a trio (32%, vs 23% for proband-only cases, p=0. 002), with de novo and phased compound heterozygous variants as the two additional categories of causal variants identified in trio cases. In addition to our routine CES practice with an average 8 week turnaround time from test ordering to reporting (TAT), we have also offered rapid turnaround CES testing (rapid-CES) at our center for 14 patients with critical conditions. The median TAT was 14 days, with a range from 11 to 25 days. Of the 14 cases, a conclusive molecular diagnosis was achieved in 5 cases (36%), which is statistically similar to the diagnostic rate of our routine CES cases. Single-gene and/or gene-panel testing was clinically available in all 5 cases; however, rapid-CES provided a much quicker TAT and was able to impact clinical management. TAT for the single-gene or gene-panel testing for the 5 genes varies widely depending on the reference laboratory, but would have required at least 4 weeks, and averages (6-10) weeks. Since CES can provide a broader search of the coding regions of the genome for causal variants, we expect CES to continue to be increasingly used as a first-line diagnostic test for many Mendelian disorders. Here, we have demonstrated that CES testing can be completed quickly while providing a high diagnostic rate in known or novel Mendelian disorders.

1956F

Establishing the next generation sequencing based genetic testing for pathological jaundice related diseases. *H. Li[1], H. Chen[2], M. Chang[2], Y. Ni[2].* 1) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 2) Departments of Pediatrics, National Taiwan University College of Medicine and Hospital, Taipei, Taiwan.

Jaundice is a medical condition with yellow skin, mucosa and eyes caused by the excess of bilirubin in the blood, reflecting impaired bile flow. In human body, any abnormalities related to hemolysis, failures in bilirubin conjugation, transport or bile secretion can result in the increase of bilirubin and lead to pathological jaundice. Many hereditary diseases can cause jaundice in this manner including: (1) hematological diseases with increased hemolysis such as G6PD deficiency, (2) diseases with no or lower UDP glucuronosyltransferase activities that affect bilirubin conjugation, (3) intrahepatic or extrahepatic cholestasis such as Dubin-Johnson syndrome, Alagille Syndrome and progressive family intrahepatic cholestasis, (4) inborn errors of bile acid synthesis also led to the failure of bile secretion. Since the causes of jaundice can be very complex despite the clinical symptoms are similar, it would be necessary and helpful to clarify which disease accounts the accumulation of bilirubin or impairment of bile flow by genetic test. However, the fact that more than 50 genes have been shown to relate to diseases with pathogenic jaundice leads the genetic testing by traditional Sanger sequencing a time and labor-consuming work. In this study, we tested the possibility of using next-generation sequencing (NGS) as the genetic test platform for jaundice related diseases. Target enrichment strategy was used and the whole genomic regions of target genes related with diseases mentioned above were captured with Roche NimbleGen customer designed probes. Libraries were paired-end sequenced by Illumina Miseq system. Variants calling and annotation were performed with Miseq Reporter Enrichment Workflow and Illumina VariantStudio software. SIFT and PolyPhen2 were used to predict the biological significance of the genetic variants. The average coverage depth of target regions is 202. 3 ± 23. 7 and the mean uniformity is 92. 9% ± 0. 3% indicate that the reads distribution is even across a targeted region. As a pilot study, patients with known genetic defects were tested for validations. All of the 20 previously identified disease causing variants were confirmed by this panel. There are 8 more variants newly identified by this method which can help to validate involvement of each genes in different diseases. The high throughput and high detection rate of NGS-based method shows the great potential to facilitate the clinical genetic testing for jaundice related diseases.

1957W

Data re-analyses lead to improved molecular diagnosis rate in clinical whole exome sequencing (WES): experience in 5,700 cases. *P. Liu[1], W. Bi[1], F. Xia[1], Z. Niu[1], R. Person[1], M. Walkiewicz[1], J. A. Rosenfeld[1], T. Chiang[2], M. Leduc[1], S. Pan[1], R. Xiao[1], J. Zhang[1], X. Wang[1], L. Meng[1], W. He[1], F. Vetrini[1], N. Veeraraghavan[2], P. Ward[1], A. Braxton[1], T. Vaughn[1], E. Boerwinkle[2,3], S. E. Plon[1], A. L. Beaudet[1], D. Muzny[2], J. R. Lupski[1], R. A. Gibbs[1,2], C. M. Eng[1], Y. Yang[1].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Human Genetics Center, University of Texas Health Science Center, Houston, TX.

The clinical application of WES generates a wealth of variant information for molecular diagnosis in patients with diverse clinical phenotypes. Follow-up reanalysis may lead to additional molecular diagnoses with potential clinical utility without repeating wet lab processes and incurring related costs. Here, we re-analyzed data from patients referred for diagnostic WES between 2011 and 2015. The initial diagnostic rate was 24% (1372/5700). Of the 4328 patients without a definitive molecular diagnosis in their first WES report, 3. 9% (N=167) received a new diagnosis after clinical WES re-interpretation 1 month to 3. 5 years after generation of the original report. Factors contributing to the growth of the molecular diagnostic rate include new disease gene discovery (59%), CNV analysis (19%), additional familial studies, mostly demonstrating *de novo* origin of a variant (15%), follow-up clinical correlation or updated variant interpretation (5. 5%), and interrogating specific regions with suboptimal WES coverage by Sanger (3. 0%). To illustrate the influence of the aforementioned factors on cases analyzed at different time intervals, we split our clinical WES cohort into two specific groups, NEJM2013 and JAMA2014, which correspond to two previous published studies. The overall rate of WES molecular diagnoses was ~25% at the time of each publication. After accumulating evidence for up to 3. 5 years, the enhanced diagnostic rates became 32. 8% for NEJM2013 and 28. 4% for JAMA2014. New diagnosis triggered by recent disease gene discovery predominates in the NEJM2013 group, accounting for 89% of the total growth compared to 58% in JAMA2014, signifying the importance of constantly updating disease gene databases when performing clinical WES analysis. The impact of CNV analysis and additional familial studies only emerged in the JAMA2014 group, representing 20% and 11% of the new molecular diagnoses, potentially reflecting physicians' evolving attitudes towards ordering WES prior to CNV analysis. Overall, our experience demonstrates the clinical and cost utility of WES reanalysis over time, and the impact of both new gene discovery and improved wet-bench experimental (capture design) and computational analyses (trio designs and CNV detection) that can yield significant increases in molecular diagnostic rates. Remaining challenges include evaluating the strength of evidence for new genes, CNV calling algorithms, and communicating updated results with referring physicians.

1958T

Genetic predisposition of human herpes virus associated lymphoproliferative disorders. *H. Liu[1], X. Chen[1], Y. Zhang[1], F. Wang[1], W. Teng[1], X. Han[1], F. Jin[1], C. Zhen[1], S. Song[1], Q. Yin[1], C. Tong[1], P. Zhu[2].* 1) Medical Laboratory Division, Hebei Yanda Ludaopei Hospital, Beijing, China; 2) Department of Hematology, Peking University First Hospital, Beijing, China.

PurposeLymphoproliferative disorders (LPD), which consist of infectious mononucleosis (IM), chronic active Epstein-Barr virus infection (CAEBV), hemophagocytic lymphohistiocytosis (HLH), X-linkage lymphoproliferative diseases (XLP), post-transplant lymphoproliferative disorders (PTLD) and lymphoma, are a series of diseases shared some common manifestations. Human herpes virus (HHV) is the most common pathogen related to LPD. HHV consists of 8 subtypes, including HSV-1, HSV-2, VZV, EBV, CMV, HHV-6, HHV-7 and HHV-8. HHV is conditional pathogenic, and HHV associated LPD occur mainly in East Asian countries. So we speculate that the pathogenesis of LPD is associated with genetic background of patients. The aim of the present study is to explore the genetic predisposition of HHV associated LPD. **Methods**608 HHV associated LPD cases were collected. All types of HHV were detected except VZV. EBV infection was the most common, which was detected in 82% cases. CMV infection was detected in 11% cases. Patients with different types of HHV infection did not show obvious differences in manifestations. 50 unrelated healthy donors who were HHV-negative were as normal controls. Coding exons and flanking intron sequences of six common HLH genes including PRF1, UNC13D, STXBP2, STX11, SH2D1A and XIAP were amplified and sequenced by Sanger sequencing. **Results**One quarter (157/608) of the patients harbored mutations in PRF1, UNC13D, STXBP2, STX11, SH2D1A or XIAP and 87 novel mutations were found. HLH cases showed the highest frequency of mutations (34. 8%). Monoallelic mutations were most common, accounting for 61% of the positive cases. Besides, 9% cases harbored synergistic mutations of two distinct genes. UNC13D was the most frequently mutated gene, followed by PRF1 and STXBP2. Deletion was often found in SH2D1A and XIAP. Synergistic mutations of two distinct genes could be found in some females with heterozygous SH2D1A or XIAP mutations. In silico analysis further suggested that the novel mutations found in the present study were associated with pathogenesis of LPD. **Conclusion**The present study investigates the molecular genetics of the largest cohort of HHV associated LPD in our country to date. We find that genetic immunodeficiency gene mutations play an important role in the pathogenesis of LPD and affects the clinical manifestations, prognosis and treatment options of LPD.

1959F

Sample ID Quality Control and Methodological Improvements for NGS Gene Panels. *J. M. Marqués, L. Repetto, L. Guggeri, V. Russo, E. García, A. Agorio, A. Torres, C. J. Azambuja.* GeniaGeo, Montevideo, Uruguay.

Many labs perform variant verifications by Sanger sequencing not only to check NGS results but also to check sample ID. In our lab we established the need for checking routinely the sample ID before the final report is generated. For this purpose we developed an internal control to perform systematically. We ordered custom Ampliseq™ primers for amplification of regions containing STRs commonly used in human identification and evaluated their performance in our Ion Torrent PGM. Then, we selected six informative STRs with good performance so we were able to differentiate one sample among over a million samples. We mixed the primers corresponding to STRs with pools from panels such as *BRCA1&2* and *CFTR* community panels, and Lynch and FAP custom panels. We have previously detected in these pools a certain degree of disequilibrium between amplicons so we performed modifications to the standard procedures by adding PCR cycles to some amplicons in order to achieve better equilibrium. Standard bioinformatics pipeline was modified with a different alignment argument to allow more accurate detection of small deletions next to the amplicon ends. In parallel, in a separate workflow, we performed the identification of each sample with 19 STRs markers by classic methods: PCR–CE with fluorescent primers, to establish the alleles markers for each sample. As the NGS yielded the results, before continuing with the pipeline for analysis of variants, we checked the identity of the sample comparing the results from both methods. Introducing the STRs in pools and performing the control quality mentioned above, allowed us to detect a problem with sample IDs in our lab. It has also been useful in cases where two samples had exactly the same variants because you could distinguish different individuals through STR NGS analysis. Finally, it allowed us to identify a problem with the consistency of an NGS result with its corresponding Sanger sequencing confirmation of a sample. Modifying alignment arguments in the bioinformatics pipeline enhances the detection of heterozygous and homozygous intronic polymorphic 4 bp deletion in *BRCA2* present at the end of an amplicon. Overall, improvements had been made regarding a more comprehensive protocol to diagnose germline mutations and to control sample IDs. Finally, we have developed an internal useful quality control for NGS that only needs a systematic PCR and Capillary Electrophoresis in parallel to NGS workflow.

1960W

A Quality Assurance Framework for Supporting a Clinical NGS Laboratory. *S. Marshall, B. Breton, A. Indap, T. Mullen.* Good Start Genetics Inc. , Cambridge, MA.

Integrated informatics and data science are essential components of a quality assurance framework to support the clinical next generation sequencing laboratory. NGS has changed how we do clinical molecular diagnostics, due in part to the significant increase in the complexity, volume, and automation of the underlying procedures. However, the paradigm of developing quality assurance tools and processes in the clinical operational setting remains unchanged. For example, monitoring assay health, tracking and trending instrument performance, identifying technical problems and reacting to them in rehearsed ways, are still the goals of a quality assurance program. By leveraging recent advances in the collection, analysis, and visualization of data, these important quality assurance goals can be met. These metrics and tools can help identify contamination and sample issues, automation errors, instrumentation malfunctions which could impact genotyping accuracy. Dynamic dashboards in Tableau enable lab personnel to quickly identify various quality issues. Integration with statistical methods implemented in R/python such as breakout/anomaly detection can enable more automated ways to identify these issues. In addition, integration with other 3rd party tools like UCSC Genome Browser provide orthogonal ways of viewing NGS quality and laboratory data in a clinical setting. To approach this challenge, we have developed an extensive suite of metrics that monitor run (batch) and sample quality. The metrics go beyond routine depth of coverage and Q30 scores; instead, these metrics interrogate clinical data on multiple additional dimensions. We have applied clinical informatics with robust data science to monitor the entire analytical laboratory process. Through interactive visualization of the data and analysis, we have integrated these tools into routine geneticist review.

1961T

Deciphering the best DNA extraction method for massively parallel sequencing of the human clinical exome. *B. Meltzer, J. Devaney, K. Cusmano-Ozog, J. Campos, S. Hofherr.* Childrens National Health System, Washington, DC 20010, DC.

The clinical exome contains 4813 genes based on information in the Human Gene Mutation Database, the Online Mendelian Inheritance in Man (OMIM) catalog, GeneTests. org, and the Genetic Testing Registry. The application of this test in the clinical lab requires quantities of genomic DNA that are pure, intact, double stranded, highly concentrated, and free from bacterial contamination. Therefore, we investigated four methods for DNA extraction from blood and cheek swabs for MPS anlysis. Genomic DNA was purified from peripheral blood using three different kits. In addition, we extracted check swabs using one kit. For all kits, we used the standard extraction conditions as recommended by the manufacturer. Three of the extraction kits use a robotic system for the extraction (EZ1 DSP Blood Kit, QIAsymphony DSP DNA Mini Kit, and ORAcollect-DNA) while the Gentra Puregene Blood kit is a manual extraction. All the extractions were quantitated using a Qubit 2. 0 Fluorometer (Life Technologies). The sequencing was performed using TruSight One (TSO) sequencing panel with analysis on a NextSeq 500 system. We performed alignment to the hg19 reference genome with the Burrows-Wheeler Aligner (BWA; 0. 7. 7). BaseSpace's version of BWA uses the Genome Analysis Toolkit (GATK; 1. 6) to call genetic variants (SNVs) and short indels. We examined the percent Qscore > 30, percent target coverage at 20X, and the number of fragments under 20X coverage for each type of extraction using three samples with duplicate extractions. In addition, we used the Kraken Metagenomics application to calculate the percent of bacterial DNA in the MPS data. Results: We analyzed 24 samples on the TruSight One panel and the four different extraction methods did not show significant differences in the Qscores, the total aligned reads (p=0. 29), coverage at 20X (p=0. 40), and number of fragments not covered at 20X (p=0. 47). In addition, we did not show significant differences between samples for the number of SNVs or indels called. Interestingly, the ORAcollect-DNA extraction (check swab) showed a higher percent of reads classified as bacteria (10. 1%) compared to the amount of bacterial DNA in the other blood-based extractions (4. 6%). The ORAcollect-DNA extractions (n = 6) all showed the bacteria *Streptococcus mitisas* as the most abundent (12% of baterial reads). We show that four different extraction methods can be used for DNA for MPS analysis using a TSO panel on a NextSeq 500.

1962F

Detection of small size CNVs (< 20 genes) in patients with epilepsy and related disorders. *L. Meng[1], T. Gambin[3], P. Liu[1], C. Shaw[1], W. Bi[1], A. Breman[1], J. Smith[1], A. Pursley[1], S. Lalani[1,2], C. Bacino[1,2], A. Beaudet[1], A. Patel[1], S. Cheung[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX. , USA; 2) Pediatrics, Baylor College of Medicine, Houston, TX. , USA; 3) Institute of Computer Science, Warsaw University of Technology, Warsaw, 00-665 Warsaw, Poland.

Epilepsy is a common medical problem with a complex clinical presentation and extreme genetic heterogeneity. So far, there are at least 100 single genes implicated in the etiology of epilepsy (26 definitive epilepsy genes, 62 OMIM genes, and 14 new genes identified by whole exome sequencing). Most of the clinical genetic tests for epilepsy focus either on large copy number variations (CNVs) or single nucleotide variations. Few studies have investigated CNVs affecting one or few genes, with even less known at the exonic level. This is partially due to the technological limits to reliably detect small size CNVs by most molecular assays. We designed a chromosomal microarray (BCM v8 CMA) to include coverage for 22 definitive epilepsy genes (84. 6%, 22/26), 46 OMIM genes (74. 2%, 46/62), and 8 new genes (57. 1%, 8/14). In order to evaluate the diagnostic yield of our CMA for epilepsy and related disorders, we retrospectively analyzed 23,630 cases studied at Baylor Medical Genetics Laboratory. Among 1714 cases with a clinical indication of an epilepsy-related presentation, 527 (31%) were found to have CNVs that are pathogenic or of uncertain clinical significance. Focusing on small CNVs with less than 20 genes, 26 cases had CNVs that include known epilepsy genes (27 CNVs, ranging from 1 to 18 genes per CNV). Noticeably, 12 of these cases involve only 1 gene, highlighting the clinical importance of exonic CNV detection in this cohort. Parental studies were performed on 8 cases with 4 confirmed to be *de novo*. We also analyzed cases with incomplete test indications and found 195 additional CNVs encompassing known epilepsy genes. The most common genes with CNVs in this patient cohort are *PLCB1* (20), *CNTNAP2* (19), *CDKL5* (15), *HCN1* (13) and *CLN8* (12). In total, we identified 22 cases carrying *de novo* CNVs of known epilepsy genes including *CDKL5* (4), *MEF2C* (3), *PLCB1* (2), *SCN1A* (2), *CLN8* (1), *ASAH1* (1), *BRAT1* (1), *CACNB4* (1), *CHD2* (1), *GRIN2B* (1), *GABRB3* (1), *GABRD* (1), *KCNT1* (1), *SCN2A* and *SCN9A* (1), and *STXBP* (1). With exon-by-exon coverage for disease and candidate genes, CMA has expanded its clinical utility from detecting recurrent microdeletion/microduplications to small CNVs at the exonic level. This not only increases the diagnostic yield for epilepsy and other genetic disorders, but also greatly expands our understanding about small size CNVs, which were usually not detected in routine clinical diagnostic testing.

1963W

Prenatal diagnosis of fetal akinesia deformation sequence caused by mutations in *KLHL40*. *X. Tian[1], PL. Kuo[2], TH. Chen[3], HP. Pan[4], Y. J. Jong[5], LJ. Wong[1,6].* 1) Baylor Miraca Genetics Laboratories, Houston, TX; 2) Departments of Obstetrics and Gynecology, National Cheng Kung University Hospital,Tainan, Taiwan; 3) Department of Emergency, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan; 4) Center for Medical Genetics, National Cheng Kung University Hospital,Tainan, Taiwan; 5) Departments of Pediatrics and Laboratory Medicine, Kaohsiung Medical University Hospital, and Graduate Institute of Medicine, College of Medicine, Kaohsiung Medical University, Kaohsiung; and Department of Biological Science and Technology, National Chiao Tung University, Hsinchu, Taiwan; 6) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA.

Background: Nemaline myopathy (NEM) is a common congenital myopathy. Mutations in at least 7 genes (NEM1-NEM7): *TPM3, NEB, ACTA1, TPM2, TNNT1, KBTBD13 and CFL2* are responsible. Recently, mutations in *KLHL40*, encoding the Kelch-like family member 40 protein, have been found to be associated with the severe form of NEM characterized by severe weakness, fetal akinesis and contractures, namely fetal akinesia deformation sequence (FADS). Most of NEMs are diagnosed by muscle biopsy followed by molecular confirmation. However, prenatal diagnosis by fetal muscle biopsy followed by sequencing of candidate genes one by one is costly and time consuming. The severe form of NEM is prenatally detectable. A comprehensive panel with quick turnaround time will be helpful for prenatal diagnosis. **Case report:** The proband is an 18 weeks fetus with akinesis and ultrasound detectable arthrogryposis. The mother had two previously similarly affected fetuses aborted at 21 and 28 weeks, respectively. Parents are unrelated Taiwanese. Clinical diagnosis of FADS possibly related to severe congenital myopathy was suspected. Molecular diagnosis was performed on the proband by target capture of genes causing a broad spectrum of neuromuscular disorders followed by next generation sequencing (NGS). **Methods:** All coding exons and at least 20 bp of flanking intronic sequences of 246 neuromuscular disease (NMD) related genes including KLHL40 gene are captured followed by NGS on Illumina HiSeq2000. Subsequent confirmation and phase study are performed by Sanger sequencing. **Results:** Our NGS analysis revealed two compound heterozygous mutations: c. 602G>A (p. p. W201*) and c. 1516A>C (p. T506P), in the *KLHL40* gene of the aborted fetus. The father carried the c. 1516A>C (p. T506P) mutation and the mother carried the c. 602G>A (p. p. W201*) mutation. The finding of *KLHL40* mutations is consistent with the congenital rapidly fatal form of NEM characterized clinically by in utero presentation of FADS. **Conclusion:** The comprehensive approach of target gene capture followed by NGS analysis effectively detects the mutations causing FADS. The inclusion of newly discovered *KLHL40* in the NMD panel increases the detection efficiency of the most common severe form of congenital NEM prenatally. The identification of the *KLHL40* mutations makes early prenatal diagnosis for this couple feasible.

1964T

Diagnosing skeletal disorders in the next generation sequencing era: lessons learned from skeletal disorder panels on 175 patients. *Y. Xue, W. R. Wilcox, M. Hegde.* Emory University, Atlanta, GA.

Making a correct diagnosis for the >450 highly genetically as well as clinically heterogeneous skeletal disorders is critical for management and accurate genetic counseling, but is often difficult due to overlapping clinical and radiographic features, especially at a young age, genetic locus heterogeneity and difficulty in choosing the appropriate molecular test. To expedite the molecular diagnosis of skeletal disorders, we designed proportionate short stature panels and skeletal disorder panels. The proportionate panels consist of sequencing of disease associated genes (45 genes), SNP array, Russell-Silver testing and UPD14. Skeletal disorder panels contain four subpanels (total 216 genes), namely, disproportionate short stature (85 genes), skeletal dysplasias with increased bone density (34 genes), skeletal dysplasias with decreased bone density or osteolysis (46 genes) and limb malformations and brachydactyly (51 genes). These four subpanels can also be ordered separately or together as a comprehensive panel. To compare the clinical diagnostic yields of the panels, we analyzed data from 175 clinical tests performed at the Emory Genetics Laboratory from January 2014 to April 2015. We found that the skeletal disorder panels had an overall 39% diagnostic yield in 75 cases while the proportionate panels had 5% in 100 cases. Among the distinct skeletal disorder panels, the disproportionate short stature had the highest yield of 75% (n=8) followed by skeletal comprehensive panel (51%; n=37). The lowest was the limb malformation panel at 0% (n=13). For proportionate cases identified with genetic causes, two cases had copy number changes, the two cases had sequencing pathogenic variants in single genes and one case had methylation abnormality consistent with Russel-Silver. Our results indicate that sequencing panel testing is an efficient approach for molecular diagnosis of disproportionate short stature. The reason for the low diagnostic yield of the proportionate panels might be disorders in this panel are relatively easily recognized and diagnosed by other means such as single gene sequencing; thus panel testing was not ordered. Limb malformations have many genetic causes remaining to be defined, which may explain the low yield.

1965F

When the Gold Standard is supplanted by Platinum: Analysis of the reliability of Sanger sequencing compared to that of Next-Generation Sequencing. *A. S. Zare, J. B. Warner, N. Gebremedhin, E. M. Zare, C. E. MacNally, V. R. Tadigotla, C. M. Stanley.* Courtagen Life Sciences, Woburn, MA.

Sanger sequencing, developed in 1977, has long been regarded as the gold standard of sequencing. Though the technology has seen many improvements and iterations in its long history, progress has slowed down significantly in the last decade. Efforts on developments in sequencing have been shifted to the fiercely competitive arms race in the Next-Gen Sequencing (NGS) space, causing this new technology to become more reliable and more robust than its predecessor. Yet, regulatory agencies still mandate the use of older technology to confirm all variants on outgoing reports, even when the preponderance of evidence points to the validity of variants detected via NGS. The question then arises, what is one to do when the outdated technology does not confirm the strong evidence present in the superior data? To address these concerns we describe the retrospective analysis of over 2,000 variants, which were initially identified by NGS and were subsequently Sanger sequenced for confirmation. Using an internally derived scoring method constructed on Phred base quality, mapping quality of reads, and coverage over a particular variant we define low quality reads as being reads with less than a score of 500. We analyzed 2,092 NGS variants that we've identified in 1,003 different samples, and only found 68 low quality variants that would require a secondary method of confirmation. In the total data set, we found only 28 variants that were not confirmed by Sanger sequencing. Of these 28 non-confirming variants, 23 were low quality variants with an internal quality score less than 500. The other 5 variants did not confirm by Sanger due to the existence of private SNPs existing under one of the PCR primers designed, resulting in allele dropout. The results of this study show the reliability of NGS when compared to Sanger sequencing, and illustrate the redundant effort needed by the diagnostic labs to create a secondary pipeline, which is now obsolete compared to the superior technology that exists today. Low quality variants likely need a method of confirmation. However, we propose the use of a sequencing technology that has quality and reliability comparable to the original detection technology.

1966W

Exon-level arrayCGH identifies novel multi-exon copy number variants in the TTN gene in patients with dilated cardiomyopathy. *O. D. Cano[1], A. P. Barnes[2], B. F. Birnbaum[2], I. Klein[3], T. de Marco[3], G. Li[1], D. J. Tan[1], A. Singleton[1], S. Aguilar[1], J. L. Jenkins[2], J. Wojciak[3], D. Dorsainville[1], D. Macaya[1], G. Richard[1], D. E. Pineda-Alvarez[1], J. Meck[1], C. Antolik[1].* 1) GeneDx, Gaithersburg, MD; 2) The Ward Family Heart Center and Division of Clinical Genetics Children's Mercy, Kansas City, MO; 3) University of California, San Francisco. CA.

TTN encodes for the protein titin, which is localized in the contractile apparatus in striated muscle. Titin, the largest protein in the human proteome, spans half the distance of the sarcomere and functions both as a scaffold for sarcomere assembly and as a molecular spring that provides passive tension in muscle fibers. Within the sarcomere structure, titin is anchored in the Z-line and spans through the I- and A-bands to the M-line. Pathogenic variants in titin are known to cause skeletal myopathies and/or cardiomyopathy. Loss of function mutations in TTN identifiable by gene sequencing account for ~25% of cases of familial dilated cardiomyopathy (DCM) and ~18% of sporadic DCM. These mutations occur predominantly in the A-band region of titin, and only rarely in the amino terminal I-band. To date, a single intragenic duplication of 28 kb involving exons 72-124 has been reported in a proband with DCM (Herman DS, et. al. , N Engl J Med. 2012:366:619). The duplication occurred in tandem and was predicted to be in-frame and encompassing the I-band. Using a multigene panel, which includes sequence analysis of 76 genes and exon-level deletion/duplication analysis of 60 genes associated with various forms of cardiomyopathy, we identified 3 patients with DCM from 2 unrelated families who were heterozygous for a multi-exon copy number variant in TTN. No other reportable sequence variants were identified in TTN or in any other gene of the panel. A 16 yo female with family history of cardiomyopathy and her 45 yo father had an 11 Kb intragenic deletion spanning exons 110-125. Based on *in-silico* analysis this deletion is predicted to be in-frame and would span the I-band of titin. Another 20 yo male with DCM who also had a family history of DCM carried a 27 Kb intragenic duplication of exons 193 through 245 of the TTN gene. This duplication is also expected to be in-frame if it occurred in tandem, and would alter the I- and A-band regions of titin. Further family studies are underway to determine if this event co-segregates with DCM in the family. Our findings suggest that intragenic, multi-exon deletions or duplications might be pathogenic and associated with DCM, and hence should be included in the diagnostic testing algorithm for cardiomyopathies.

1967T

Broadening the mutational spectrum of Cornelia de Lange syndrome: identification of four novel *HDAC8* deletions including the first report of a deletion in mosaic state. *M. Helgeson[1], J. A. Lee[2], G. Hoganson[3], J. Burton[3], C. Schaaf[4], R. Laframboise[5], S. Tremblay[5], I. Krantz[6], S. Noon[6], Y. Hu[1], K. Donato[1], Z. Li[1], L. Wysinger[1], S. Das[1], D. del Gaudio[1].* 1) Department of Human Genetics, University of Chicago, Chicago, IL; 2) Greenwood Genetics Center, Greenwood, SC; 3) Division of Genetics, University of Illinois College of Medicine, Peoria, IL; 4) Department of Pediatric Medicine, Genetics, Texas Children's Hospital, Houston, TX; 5) CHU de Québec, CHUL, Québec, QC, Canada; 6) Division of Human Genetics, The Children's Hospital of Philadelphia, PA.

Cornelia de Lange syndrome (CdLS) is a genetically heterogeneous disorder characterized by distinctive facial features, hirsutism, growth retardation, upper limb anomalies, developmental delays, and intellectual disabilities. Mutations in *NIPBL* [MIM 608667], encoding a cohesion regulatory protein, account for >80% of cases with typical facies. Mutations in the *SMC1A* [MIM 300040], *SMC3* [MIM 606062], and *RAD21* [MIM 614701] genes encoding for cohesin complex proteins, together account for ~5% of patients with atypical CdLS features. Recently, loss-of-function mutations in the X-linked *HDAC8* [MIM 300269] gene have been identified in a small number of patients with growth, cognitive and facial features consistent with those caused by mutations in *NIPBL*. We report the identification, by exon-targeted array-CGH, of four novel intragenic *HDAC8* deletions in female patients with features of CdLS. The deletions were confirmed by either Real-Time quantitative PCR or breakpoint sequencing. For one case the deletion was present in a mosaic state in the peripheral blood sample from the patient and was subsequently confirmed in buccal cells. For this patient, X-chromosome inactivation was random in peripheral blood sample, but was moderately skewed in buccal cells. For the other deletion patients, X-chromosome inactivation studies demonstrated a highly skewed pattern in the blood samples, presumably towards the normal allele, suggesting a strong selection against the mutations. Consistent with the molecular analyses, these patients displayed features of CdLS, including dysmorphic features such as synophrys, thin upper lip, and posteriorly rotated ears. Other features included significant speech and motor delays, growth retardation, myopia, and atrial septal defect. This study expands the spectrum of *HDAC8*-related CdLS mutations, highlights the importance of performing comprehensive molecular testing including copy-number analysis of *HDAC8*, and provides the first description of an *HDAC8* deletion in a mosaic state.

1968F

Success of next generation sequencing usage for adults and newborns affected with neuromuscular and skeletal disorders. *K. M. Rocha[1], M. Lazar[1], G. L. Yamamoto[1,2], M. Aguena[1], V. Takahashi[1], N. Lourenço[1], M. Varela[1], S. Ezquina[1], D. Bertola[1,2], R. C. Pavanello[1], M. Zatz[1], M. R. Passos-Bueno[1].* 1) HUG-CELL Dept of Genetics and Evolutionary Biology, Institute of Biosciences - University of Sao Paulo, Sao Paulo, Sao Paulo, Brazil; 2) Genetics Unit, Children's Institute, Clinics Hospital, Faculty of Medicine, University of Sao Paulo, Sao Paulo, Brazil; 3) Faculty of Medicine, University of Sao Paulo, Sao Paulo, Brazil.

During the past two years our lab has been using the targeted next geneneration sequencing (NGS) as the primary method for genetic diagnostic of some groups of diseases. Since then, we have concluded the molecular testing of 406 patients that were referred to our service. Those patients were divided in four main groups: neuromuscular disorders (N=229), skeletal dysplasias (N=60), hereditary cancer (N=12) and other diseases, including neurodevelopmental and cystic fibrosis (N=105). We have detected mutations in 252 patients (62. 1%). The highest detection rate was found in patients grouped as neuromuscular disorders, 72. 9% (N=167), followed by skeletal dysplasias, 63. 3% (N=38), hereditary cancer, 41. 7% (N=5) and all other diseases, 40% (N=42). A significant improvement in the diagnosis rate was made through the use of a software (NextGene from Softgenetics) to detect gross deletions and duplications from NGS data. Although the majority of the cases had hemizygous deletion in *DMD* gene, we were also able to detect *DMD* duplications in females, heterozygous duplications of *PMP22* gene and heterozygous deletions of *MLH1*, *TWIST1* and *NSD1* genes. Apart from *MLH1* gene, which has not been tested by another method, all deletions and duplications were confirmed by MLPA. The use of this software has also simplified our routine, since now, we only do the MLPA as a confirmation test. Along with our routine diagnostic tests, we have also performed a pilot study with pregnant women in whom prenatal ultrasonography had showed a fetus presenting shortened or malformed limbs. These women were referred for clinical genetics prenatal consultation and received a letter of recommendation to collect blood and X-ray images from the newborn. After postnatal images evaluation, a diagnosis hypothesis was made and molecular testing performed accordingly. In the past few months, 10 cases were evaluated and 7 have completed molecular testing. Among these 7 cases, 5 newborns died in the perinatal period. Conclusive diagnostic was obtained in all 7 cases (lethal cases: otopalatodigital type II; thanatophoric type I, thanatophoric type II, achondrogenesis type II, osteogenesis imperfecta type II; non-lethal cases: achondroplasia, diastrophic dysplasia). Targeted NGS has been proved a valuable diagnostic tool in patients whose phenotype had been well characterized, mostly when the number of disease-causing genes is small and well known. Financial support: Fapesp/CEPID and CNPq/INCT.

1969W

Gross deletion of *TSPAN12* detected in patients with familial exudative vitreoretinopathy. *J. Song[1], S. Seo[1], Y. Yu[2], S. Cho[1], M. Seong[1], S. Park[1].* 1) Dept Laboratory Medicine, Seoul National Univ Hosp, Seoul, South Korea; 2) Dept Ophthalmology, Seoul National Univ Hosp, Seoul, South Korea.

Background: Familial exudative vitreoretinopathy (FEVR) is a rare hereditary disorder characterized by failure of peripheral retinal vascularization. FEVR is genetically heterogeneous, and found in various modes of inheritance. Autosomal dominant inheritance is the most common form in FEVR, with *FZD4*, *LRP5*, and *TSPAN12* known to be associated with the disease. These genes are essential components of wingless (Wnt) pathway and pathogenic mutations affecting the function of these genes result in abnormal retinal vascular formation. Until now, gross deletion of *TSPAN12* in FEVR patients had not been reported. Herein, we report cases of FEVR with *TSPAN12* deletion. **Methods:** In patients who were suspected to carry gross deletion of *TSPAN12*, gene dosage analysis of *TSPAN12* was performed. Diagnosis of FEVR was established by ophthalmic examinations. All patients had been previously screened for *NDP*, *FZD4*, *LRP5*, and *TSPAN12* mutation, which were all found to be negative. Gross deletions and duplications of NDP, FZD4 and LRP5 screened via Multiplex Ligation-dependent Probe Amplification (MLPA) were also negative. **Results:** Patient #1, who was found to carry chromosome del(7) (q31. 2q34), had peripheral retinal avascular area in both eyes. Considering the location of *TSPAN12* being 7q31. 31, it was suspected that whole gene deletion of *TSPAN12* was the pathogenic cause for patient's symptoms. Patient #2 was found to carry a rare *TSPAN12* variant of c. 484G>A as a homozygote by sequencing analysis. Her brother (Patient #3) who had the same diagnosis did not carry this variant. Since the variant did not segregate in this family, it was concluded as a less-likely pathogenic variant, but still there was a possibility of gross deletion in this exon regarding that this rare variant was detected as homozygote. Gene dosage analysis showed that all three patients carried half the amount of *TSPAN12* gene compared to normal control (ratio of 0. 45 to 0. 50), indicating that each patient carried gross deletion of *TSPAN12*. **Conclusions:** In this study, we have identified gross deletion of *TSPAN12* in patients with no mutation detected in NDP, FZD4 and LRP5. This is the first report of *TSPAN12* gross deletion detected in FEVR patients. Evaluating gross deletion or duplication of *TSPAN12* should also be considered in genetic workups for FEVR patients. ----------**Keywords**: Familial exudative vitreoretinopathy, *TSPAN12*, gross deletion.

1970T

Whole Exome Sequencing and Whole Mitochondrial Genome Sequencing for the Molecular Diagnosis of Mitochondrial Disorders. *R. Bai[1], A. Balog[1], J. Higgs[1], K. Retterer[1], D. Arjona[1], J. Juusola[1], Q. Abu Ali[1], S. F. Suchy[1], G. M. Enns[2], R. Haas[3], A. Goldstein[4], M. Tarnopolsky[5], S. Parikh[6], W. C. Copeland[7], D. Niyazov[8], M. J. Falk[9], S. J. Bale[1], W. K. Chung[10], G. Richard[1].* 1) GeneDx Inc, Gaithersburg, MD; 2) Stanford University, Palo Alto, CA; 3) Rady Children's Hospital-San Diego; 4) Childrens Hospital of Pittsburgh, Pittsburgh; 5) McMaster University Medical Center, Hamilton, Ontario; 6) Cleveland Cinic, Cleveland, Ohio; 7) NIEHS ,Research Triangle Park, NC; 8) Ochsner Health System, New Orleans, LA; 9) Children's Hospital of Philadelphia, Philadelphia, PA; 10) Columbia University Medical Center, New York, NY.

Molecular diagnosis of mitochondrial disorders (MtD) is challenging due to their clinical and genetic heterogeneity. Until now, it entailed consecutive testing for common mutations/ genes/disease-specific gene panels until a diagnosis was made or all these options were exhausted. Now whole exome sequencing (WES), which covers >96% of the gene coding regions in the nuclear genome, and whole mitochondrial genome sequencing (WMGS) are becoming widely applied to the molecular diagnosis of MtD. We performed WES on 865 patients clinically suspected of having a MtD, combined with WMGS in 673 of these cases. According to the MtD criteria (Neurology 2006, 67:1823), 86/163/616 had definite/ probable/possible MtD, respectively. In the majority of the cases (83%), WES was performed on proband-parent trios (WES-Trio), while the remaining 17% were singletons or non-trios. A positive diagnostic result was obtained in 254/865 (29%) cases tested by WES. The positive rate for trios (30. 5%) was appreciably higher than that of singletons and non-trios (23. 8%). Of the 254 positive cases, 97 (38%) had pathogenic/ likely pathogenic variants (PV) in known MtD genes, 15 (6%) had PV in genes encoding proteins involved in other mitochondrial pathways, and the remaining 142 (56%) had PV in well-known non-MtD genes with clinical features overlapping with MtD. Concurrent WMGS performed for 673 cases revealed PV in mtDNA at a clinically significant level of heteroplasmy in 26 (4%) cases, including 2 large-scale mtDNA deletions and 24 point mutations (8 homoplasmic, 16 heteroplasmic). Three of the PV were seemingly unrelated to the patient's reported phenotype, two of these cases as well as another six cases with PV in mtDNA also had a positive WES result, in a nuclear gene associated with either a MtD or a non-mitochondrial disorder. In summary, the combined positive rate of WES and WMGS in our cohort was approximately 32%, considerably higher than the positive rate (<20%) of our multi-gene panel for MtD, WHICH INCLUDES almost all the well-characterized mitochondrial disease genes. We provide evidence that PVs in mtDNA and nuclear genes may coexist in a patient, albeit this co-occurrence is rare. We suggest that WES-Trio/WMGS is effective as the first-line test for patients with a suspected mitochondrial disorder in order to achieve the best diagnostic yield.

1971F

Impact and implementation of benchmarking frequency thresholds for variant filtering in clinical diagnostics. *M. C. Dulik[1], L. K. Conlin[1,2], S. Biswas[1], J. L. Abrudan[1], V. Jayaraman[1], M. H. Li[1], I. D. Krantz[3], N. B. Spinner[1,2].* 1) Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Pathology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Allele frequency is often a significant piece of information used in classifying the pathogenicity of a variant in clinical molecular genetics and cytogenetic laboratories. Many clinical laboratories performing exome and whole genome sequencing rely on allele frequency to filter variants to a manageable list for further interpretation. As the variant interpretation is the most time consuming and resource-heavy component of such tests, reduction of this variant list is desirable to include only the most important variants. This is especially critical for secondary finding analysis, which can identify numerous previously reported variants but is not the primary reason for the test. Secondary findings often require more effort than those for primary because in many cases multiple papers with additional evidence are considered for a reported variant. As part of the NHGRI-funded Clinical Sequencing Exploratory Research Program, the Pediatric Genetic Sequencing (PediSeq) Project at The Children's Hospital of Philadelphia and the University of Pennsylvania is identifying best practices for exome sequencing in a pediatric clinical diagnostics setting. To determine appropriate frequency thresholds for filtering secondary findings, over 130 genes implicated in causing autosomal recessive conditions were selected from our established PediSeq carrier status list. More than 200 prevalent, well-known pathogenic variants were identified through literature and databases searches. The overall contribution of each variant to causing disease was recorded if known, and allele frequencies for each variant were queried from publicly available databases. A set of 50 exomes were then filtered using a list of medically actionable genes, a commercially available database of reported mutations, and different allele frequency cutoffs to assess the impact of a benchmarked value as compared to the commonly applied value of 5%. On average, use of the benchmarked threshold reduces the list of variants by an average of 64%, resulting in 10 variants per exome versus 30 with the 5% cutoff. Interpretation of variants between the benchmark and 5% threshold in a subset of exomes resulted in the identification of no additional pathogenic variants. It is shown that by considering the composition of allele frequency databases and the frequency with which known pathogenic variants occur, one can reduce the overall burden of secondary findings for clinical exome sequencing.

1972W

The Genome Clinic in Geneva: development of an efficient and patient-friendly diagnostic application of NGS. *E. B. Hammar[1], S. Fokstuen[1], P. Makrythanasis[1,2], M. Guipponi[1,2], E. Ranza[1,2], K. Varvagiannis[1], F. A. Santoni[1,2], M. Albarca-Aguilera[1], M. Poleggi[1], A. Vannier[1], C. Brockmann[1], F. Couchepin[1], A. Mauron[4], S. A. Hurst[4], C. Moret[4], C. Gehrig[1], J. Bevillard[2], T. Araud[1], S. Gimelli[1], E. Stathaki[1], A. Paoloni-Giacobino[1], F. Sloan-Béna[1], L. D'Amato Sizonenko[1], M. Mostafavi[1], H. Hamamy[2], T. Nouspikel[1], J. -L. Blouin[1], S. E. Antonarakis[1,2,3].* 1) Genetic Medicine and Laboratory, University Hospitals of Geneva, 1211 Geneva 14, Geneva, Switzerland; 2) Department of Genetic Medicine and Development, University of Geneva, Geneva, Switzerland; 3) iGE3, Institute of Genetics and Genomics of Geneva, Geneva, Switzerland; 4) IEH2, Institute of Ethics, History and Humanities, University of Geneva, Geneva, Switzerland.

The advances of next generation sequencing (NGS) technologies enable their application in clinical care. However, there are several challenges for the implementation of NGS in clinical diagnostics including technical, financial, ethical, managerial, variant interpretation, genetic counseling and quality control issues. The aim of our work is to continuously improve the diagnostic use of NGS to be most efficient in terms of accuracy of variant interpretation, turn-around time of analysis and patient information. The reimbursement of diagnostic NGS tests has recently been accepted by the Swiss Federal Office of Public Health (SFOPH). The prices are described in the List of analysis (see link: http://www. bag. admin. ch). We use whole exome sequencing (WES) followed by targeted bioinformatics analysis of gene panels. In order to optimally integrate the implementation of NGS in clinical care, we have created a multidisciplinary working group, the Genome Clinic Task Force (GTCF), which meets once a week. During these meetings, clinical cases and results are presented, the degree of pathogenicity of variants is determined and the final laboratory reports are critically discussed. Reimbursement and ethical issues such as the content of the informed consent form, the disclosure of results as well as the management of incidental findings and variants of unknown significance (VUS) are also addressed. Since January 2013, a total of 116 cases have been included in the GCTF workflow, and 71 cases are finished. We found pathogenic variants (class 4 or 5) in 23% (6/26) of patients with developmental delay and in 57% (26/45) of patients suffering from various mendelian disorders, which gives us an average detection rate of (likely) causative variants of 46 % (33/71). Variants of unknown clinical significance (VUS, class 3) were found in 23 % (6/26) of the patients with developmental delay and in 7 % (3/45) of the patients with other phenotypes. VUS are disclosed only if they are closely related to the clinical phenotype. The mean duration of a complete analysis was 193 days. In order to render our diagnostic use of NGS even more efficient and patient-friendly, our main aims are 1) to improve different steps of the workflow in order to accelerate the analysis process 2) to integrate new and more reliable methods of variant interpretation (as described in S. Richards et al. 2015) and 3) to improve the genetic counseling.

1973T

Pre-clinical cancer: an example of how genomics could reshape the prognostic paradigm. *E. G. Seaby[1], R. D. Gilbert[2], G. Andreoletti[1], R. J. Pengelly[1], S. Ennis[1].* 1) Genomic Informatics, University of Southampton, Southampton, Hampshire, United Kingdom; 2) Wessex Regional Paediatric Nephro-Urology Service, Southampton Children's Hospital, Southampton, Hampshire, United Kingdom.

Whole exome sequencing (WES) has proven success in rare disease of unknown aetiology, yet one of its greatest strengths lies in the ability to return to and manipulate data for prognostic benefit. We present the case of a 4-year-old Caucasian male with atypical haemolytic uraemic syndrome (aHUS), moya moya and a non-specific dysmorphology. In 2013, this patient was referred for WES to elucidate a cause for his aHUS; no pathogenic variants were identified. In 2015, this case was revisited following clinical review. Phenotypic information concerning splenomegaly and persistent thrombocytopaenia informed a revised analysis. Filtering parameters reduced 51 variants to one heterozygous splicing variant in *CBL* (c. 1096-1G>T), later confirmed *de novo* by Sanger sequencing. Germline *CBL* mutations are associated with: Noonan-like syndrome; moya moya; vasculitis; and the progression to juvenile myelomonocytic leukaemia (JMML) following somatic loss of heterozygosity (LOH) on chromosome 11q. JMML is an aggressive, childhood myeloproliferative disease cured only by stem cell transplant, yet haematological heterogeneity is reported; some haematological abnormalities spontaneously resolve, while others progress aggressively. This considered, we scrutinised available data for LOH across chromosome 11 in the region of the *CBL* locus; facilitated by DNA extracted from peripheral blood. We retrospectively assessed LOH by plotting B-allele frequency ratios across the exome. Significant allelic imbalance across chromosome 11q strongly suggested a clonal expansion of peripheral lymphocytes consistent with myeloproliferation and a transformation to JMML. These results pose a difficult diagnostic dilemma; our patient is displaying evidence of myeloproliferation and progression to JMML despite not manifesting clinical leukaemia (a bone marrow aspirate and blood film were unremarkable). This pre-clinical finding hugely challenges the appropriate action(s) to take, especially since a decision to proceed to transplant is not without significant risk, yet may prove curative. This has necessitated a multidisciplinary convergence of genomic informaticians and clinicians into discussions concerning best practice. In the ever emerging genomics era, cases such as these will become ever prevalent and the interpretation and translation of genomic data within clinical medicine has the potential to force a paradigm shift in clinical diagnostics with substantial prognostic impact.

1974F

Prevalence of incidental findings of potentially actionable variants in exome chip in the CHARGE consortium. *M. K. Puurunen[1], Y. -P. Fu[2], C. J. O'Donnell[2] on behalf of the CHARGE Exome Chip Working Group.* 1) Framingham Heart Study, Boston University School of Medicine, Framingham, MA; 2) Division of Intramural Research, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA.

BackgroundWith next-generation sequencing (NGS) methods, disease variants with potential clinical importance are increasingly detected (i. e. incidental findings). Knowledge of 'actionable variants' can lead to a clinical intervention or personal decisions. However, estimates in large general populations of the prevalence of incidental findings expected from contemporary genome analysis are lacking. AimIn a large consortium of community-based populations we aim to assess the burden and distribution of potentially actionable variants using Illumina Infinium HumanExome BeadChip. We focus on a panel of genes listed as disease-causing with possible adult onset by an expert panel (UW Clinical Sequencing and Exploratory Research "NEXT Medicine" Return of Results Committee). MethodsThe study includes 85,101 adults across 17 community-based cohorts within the Cohorts for Heart and Ageing Research in Genomic Epidemiology (CHARGE) Consortium [54. 5% female; 15. 9% non-European descent; mean age 57 years (range 18-101)]. In a preselected set of 112 'actionable' genes, we focused on 254 coding single nucleotide variants (SNVs) [251 nonsynonymous, 1 synonymous, 2 3'slice site] observed in the NHLBI Exome Sequencing Project (ESP) and classified by a trained annotation panel as pathogenic (n=6), likely pathogenic (n=10), variants of unknown significance (VUS, n=150), or likely benign (n=88) based on allele frequency and review of literature. ResultsThe overall burden of pathogenic or likely pathogenic SNVs was 0. 34%. The number of individuals with actionable genotypes of pathogenic variants, likely pathogenic variants, likely benign variants, and VUS were 41, 250, 16266, and 15107, respectively. Pathogenic or likely pathogenic genotypes (prevalence ranging from 0. 001% to 0. 096%) were noted in 10 genes that predict a range of diseases, including *SCN5A* (ventricular arrhythmia), *SERPINA1* (alpha-1 antitrypsin deficiency) and *BRCA1/2* (breast cancer). After accounting for cohorts, no age or sex differences were observed. Significant allele frequency differences between persons of European and African ancestry were noted for SNVs in *SERPINA1, SLC7A9* and *LDLR*. ConclusionsA significant number of participants with potentially actionable SNVs were observed in community-based studies genotyped by exome chip, some with important differences depending upon ancestry. Further research is warranted to determine the clinical importance of these variants.

1975W

Exome sequencing is an efficient method for the molecular characterization of rare Mendelian disorders. *A. E. Bale, D. J Dykas, A. E. Bailey, J. C. Sinson, S. Mane, R. P. Lifton, P. Tsipouras.* Genetics, Yale School of Medicine, New Haven, CT.

Whole Exome Sequencing (WES) is increasingly utilized in clinical diagnosis of genetic conditions due to its efficiency, comprehensiveness, and agnostic nature. Being one of the first to introduce WES in clinical practice we set to evaluate its utility by reviewing our first two years of use. Exonic sequences in DNA samples from blood or saliva were captured (Roche/Nimblegen SeqCap EZ Human Exome Library v2. 0) and sequenced on the Illumina HiSeq platform. Mean coverage of the human exome was 75-100X with 97% of bases covered at least 8 times. The resulting sequence was analyzed for genetic variants of likely relevance to human disease. We reviewed the clinical notes and lab reports from a total of 1,082 cases spanning a 24 month period (2013-2014). Major diagnostic categories of testing included acute metabolic disorders in newborns and infants, dysmorphic, metabolic, and neurologic conditions presenting in childhood, intellectual disability and autism, adult-onset cardiac, vascular, neurological, hepatic, hematological and connective tissue disorders, and hereditary cancer predisposition. The referrals were primarily through Yale faculty and affiliated community physicians. We detected a pathogenic variant in 242/1,082 (22%), a variant of uncertain significance (VUS) in 295/1,082 (27%), and incidental findings in 15/1,082 (1. 3%) of cases. These statistics are in agreement with other large scale studies. The greatest number of positive tests were detected in cardiomyopathies (pos 44%, VUS 39%, neg 17%), seizures (pos 25%, VUS 42%, neg 33%), and hematological disorders (pos 25%, VUS 25%, neg 50%). Samples submitted for connective tissue disorders showed the lowest rate (pos 3%, VUS 6%, neg 91%). During the two year period, the number of tests ordered per month increased three-fold, reflecting primarily an increase in the spectrum of clinicians ordering testing for large panels of genes. WES has disrupted molecular diagnostics. A single platform covering almost all hereditary conditions allows for cost-effective, comprehensive, phenotype-based diagnostics rather than an inefficient and lengthy gene-by-gene hunt for a diagnosis.

1976T

An integrated approach to genetic diagnosis: Genomic research and clinical care at the TGen Center for Rare Childhood Disorders. *A. M. Claasen[1,2], K. M. Ramsey[1,2], N. Belnap[1,2], S. Szelinger[1,2,3], A. L. Siniard[1,2], I. Schrauwen[1,2], J. J. Corneveaux[1,2], B. E. Hjelm[1,2], A. L. Courtwright[1,2], R. F. Richholt[1,2], M. De Both[1,2], S. Rangasamy[1,2], M. J. Huentelman[1,2], D. W. Craig[1,2], V. Narayanan[1,2].* 1) Neurogenomics Division, Translational Genomics Research Institute, Phoenix, AZ; 2) Center for Rare Childhood Disorders, Translational Genomics Research Institute, Phoenix AZ; 3) Molecular and Cellular Biology Interdisciplinary Graduate Program, School of Life Sciences, Arizona State University, Tempe, AZ.

Personalized genomic medicine is changing the landscape of diagnosis of rare diseases. At TGen's Center for Rare Childhood Disorders, our integration of a clinical diagnostic center and a genomics research laboratory provides a unique opportunity to significantly impact the lives of children with rare, undiagnosed genetic disorders, and their families. We have enrolled 914 individuals across 290 families into our study to date, and completed whole-genome or whole-exome sequencing of over 800 individuals. A further 42 families sequenced at external sites have been analyzed using our in-house bioinformatics pipeline and annotation. Our study takes a multi-omic approach to diagnosis, including sequencing whole-exomes of family trios, with complementary application of mRNA-seq, methylation arrays, and structural variant detection for difficult cases. Where appropriate, patients are consented for collection of fibroblasts for reprogramming and differentiation to appropriate tissue types to better understand the biology, particularly as new gene-disease associations are made. Currently, RNAseq has been completed for 102 individuals. We implement frequent re-analysis of genomic data using updated annotation and prioritization methods. Here we report a summary of the families currently in our study. Criteria for enrollment include a previously undiagnosed condition with childhood onset, ambiguous genetic origin, and negative or inconclusive genetic tests. The majority of patients present with some form of neurological phenotype, and suspected conditions include mitochondrial disease, Aicardi Syndrome, movement disorders, and epilepsy, among others. Identified variants are characterized by their likelihood, and in approximately 36% of cases a presumed causal or likely causal variant has been found, and in about 9% of cases we have identified candidate variants which require additional testing and functional validation. Following on from diagnosis, we have further supported families by facilitating enrollment of a number of diagnosed families into clinical trials, connected them with experts in the field, and provided alternative treatment options. Research initiatives include the development of animal models, and drug screens with FDA-approved compound libraries for selected causal genes. Our center provides a unique experience for families, through our personalized approach to diagnosis and follow-through care.

1977F

Lessons from Clinical Exome Sequencing in the Montana Genetics Program. *A. F. Elias, K. Berry, J. A. Camacho, M. Clark, J. Foster, C. Hudson, J. P. Johnson, T. Pitman, S. Wakefield, D. Xu.* Department of Medical Genetics, Shodair Children's Hospital, Helena, MT.

Shodair provides clinical and laboratory genetic services through an outreach model across Montana and bordering regions. Integration of genomic approaches into patient care faces particular challenges within the health care setting of Montana with its mostly rural communities and few urban centers separated by vast geographic spaces. At the same time, the state draws a growing number of specialists, allowing increasingly for local management of patients with complex health problems. NGS technologies including whole exome sequencing (WES) have led to a substantial increase in diagnostic yields. Here, we present our experience with WES over the past 3 years. From 1/2012 to 6/2015, samples from 39 patients were submitted to AmbryGenetics. Results of 27 patients were available at the time of this abstract. Fifteen samples were submitted between 2012 and 2014 compared to 24 in 2015 corresponding to a 10-fold increase. Initially, the majority of patients had undergone repeated unsuccessful evaluations. More recently our patients underwent WES earlier in the diagnostic process. Most patients had developmental delay, but the clinical presentations span a broad spectrum of phenotypes including autism, dysmorphism, congenital anomalies, metabolic disturbances and epilepsy. In 9 of 27 patients mutated alleles highly likely to be causative were identified (diagnostic rate of 33%). Inheritance patterns were AD (6), AR (2), XL (1). All 7 of the AD and XL mutant alleles occurred *de novo*. Eight of the 9 genes were associated with Mendelian disorders: *ANKRD11*;*BBS10*; *CASK*;*GLI3*; *KMT2A;MEF2C*;*SETD5*; *TCF4*. Compound heterozygous mutations detected in one patient in a gene without disease association were found in other patients with similar presentation using GeneMatcher. Variants in 334 genes not known to underlie a genetic condition and 51 variants of uncertain significance (VUS) without actionable information were reported. The large amount of variants can increase workload and complicate genetic counseling. As a result, we are implementing a NGS platform to sequence about 4,800 genes with known disease associations. This approach has the potential to maximize diagnostic yields while minimizing the number of VUS. In a setting like Montana, where timely specialist evaluations may be limited, a phenotype-genotype-guided evaluation early in the diagnostic process rather than diagnostic odysseys preceding genetic testing may be cost efficient and improve patient care.

1978W

Utility and limitations of exome sequencing as diagnostic tool for bilateral sensorineural hearing loss. *V. Jayaraman[1], J. Abrudan[1], M. Li[2], I. Slack[1], A. Sasson[1], M. Dulik[1], S. Biswas[1], E. DeChene[1], M. Luo[1], I. Scarano[1], R. Tilton[1], S. Noon[1], A. Wilkens[1], X. Ying[1], L. Conlin[1], J. Pennington[1], N. Spinner[1], I. Krantz[1].* 1) Children's Hospital of Philadelphia, Philadelphia, PA; 2) University of Pennsylvania, Philadelphia, PA.

Bilateral sensorineural hearing loss (BLSNHL) is one of the most genetically heterogeneous diagnoses with more than 200 associated genes. Current testing for BLSNHL is either done by Sanger sequencing or next generation sequencing (NGS) of gene panels. These tests can be cumbersome and are not comprehensive. We report the results of applying WES for the diagnosis of BLSNHL. This work was carried out through the CHOP/UPenn Pediatric Genetic Sequencing (PediSeq) project, an NHGRI funded (part of the Clinical Sequencing Exploratory Research (CSER) consortium), to study the application of exome and genome sequencing to clinical care. To evaluate coverage and clinical utility of the test we exome sequenced 269 samples (from a diverse pediatric cohort sequenced for various clinical indications) using exome capture (Agilent SureSelect V4), and sequencing (Illumina HiSeq 2000) at an average depth of 100X. Reads were mapped to the human genome GRCh37 using Novoalign; post processing and analysis was performed using a combination of Picard and GATK. A list of 247 BLSNHL genes was used as a filter for analysis. Our goals were to determine 1) the coverage for all 4,421 exons in the 247 genes; 2) the number of BLSHNL variants that could be potentially missed from randomly generated list of 454 variants from HGMD; 3) the diagnostic yield within our BLSHNL cohort. The average percentage of exons completely covered at 20X was 77. 71%. The average percentage of exons partially covered and not covered were 16. 09% and 6. 20% respectively. In order to determine the clinical utility of exome sequencing as a potential diagnostic test for BLSHNL, we had randomly selected 454 BLSHNL variants from HGMD within the 247 known hearing loss genes and calculated the percentage of capture at various read depth using this test. Out of the 454 variants, 318 (70%) were covered with a minimum read depth of 20X. 81 (17. 8%) of them had a read depth between 10X and 20X, 45 (9. 9%) had less than 10X and 10 (2. 2%) were not captured using this test. For our diagnostic yield rate with this test, we have successfully identified the genetic cause in 12 out of 47 probands. We have negative results on 14 probands and 21 probands have a variant of unknown significance in one of the 247 BLSHNL gene. We will present several interesting illustrative examples as part of this presentation highlighting the utility and limitations of exome sequencing for BLSHNL.

1979T

Clinical Exome Sequencing Experience in 4866 Consecutive Cases. *J. Juusola[1], K. Retterer[1], M. T. Cho[1], F. Millan[1], J. Neidich[1,2], K. G. Monaghan[1], G. Douglas[1], R. E. Schnur[1], L. B. Henderson[1], A. Begtrup[1], A. Lindy[1], D. Pineda-Alvarez[1], Q. Abu Ali[1], D. McKnight[1], R. Bai[1], S. Suchy[1], B. Friedman[1], A. Telegrafi[1], D. Copenheaver[1], J. Tahiliani[1,3], G. Richard[1], T. Brandt[1], W. K. Chung[4], E. Haverfield[1,3], S. Bale[1].* 1) GeneDx, Inc. , Gaithersburg, MD; 2) Pathway Genomics Corp. , San Diego, CA; 3) Invitae, San Francisco, CA; 4) Columbia University, Departments of Pediatrics and Medicine, New York, NY.

Since January 2012, we have analyzed 4866 unique WES cases with the goal of identifying a molecular diagnosis. Of those cases, 18% were submitted as proband-only, 6. 5% with one additional family member, 69% with two additional family members, most often parents, and 6. 5% with three or more additional family members. WES was performed in three family members to maximize sensitivity and specificity of variant calling and to enable identification of *de novo* variants and determination of phase. Mitochondrial genome sequencing was also performed in 39% of cases. Patients were referred for WES primarily for neurological indications (46%) or multiple congenital anomalies (22%). A definitive diagnosis was made in 29. 7% of cases, with 24. 5% of cases definitive when analyzing the proband only, increasing to 31. 7% when three family members were tested. With an increasing number of cases analyzed, the identification of recurrent *de novo* likely damaging variants in candidate genes has enabled us to identify new disease genes, suggesting that the yield of WES will continue to increase over time. Upon request, 179 cases were reanalyzed after one year and 11. 7% of these analyses yielded a positive diagnostic result, and another 9. 5% had variants in strong candidate genes identified by mining our internal database for individuals with similar phenotypes. When analyzing by mode of inheritance for cases with a definitive diagnosis, 36. 4% were autosomal dominant and *de novo*, 27. 5% autosomal recessive, 11. 6% X-linked, 8. 0% autosomal dominant and inherited, 9. 1% autosomal dominant of unknown origin, 3. 5% mitochondrial encoded, and 3. 9% unknown inheritance. Copy number variation detected directly from WES data accounted for 2. 7% of diagnoses. Uniparental disomy was identified in 13 cases, and a definitive diagnosis was established in 6 of these cases. Mosaicism for pathogenic variants was identified in 17 cases, 9 of which represented parental mosaicism. The overall diagnostic yield for patients whose symptoms included intellectual disability or developmental delay was 32. 4% (N=2650), 29. 9% (N=1403) for seizures, and 22. 9% (N=849) for autism spectrum disorder. Secondary findings for the 56 genes in the ACMG list were analyzed in 3815 cases, and 3. 7% were found to have reportable variants. The experience with our first 4866 WES cases suggests that analysis of trios improves the diagnostic yield and facilitates identification of novel candidate genes.

1980F

Exome sequencing as a diagnostic test to establish the molecular cause in renal disorders. *D. Lugtenberg[1], I. Rood[2], E. Cornelissen[3], J. Schoots[1], J. Wetzels[2], R. Roepman[1], EJ. Kamsteeg[1], H. IJntema[1], E. Bongers[1].* 1) Genetics, Radboudumc, Nijmegen, Netherlands; 2) Nephrology, Radboudumc, Nijmegen, Netherlands; 3) Pediatric nephrology, Radboudumc, Nijmegen, Netherlands.

Single gene disorders are estimated to account for ~30% of children and ~10% of adult patients attending renal outpatient services. In the Radboudumc diagnostic exome sequencing for a broad spectrum of isolated- and syndromic renaldisorders has been developed. The approach consists of a two-tier analysis in which the first step is to screen for pathogenic variants in genes that are known to be mutated in renal diseases (187 genes) or (renal) ciliopathies (125 genes). If causative mutations are not identified in the first step, the complete exome data set can be analyzed with informed consent. Here we present the results from exome seqeuncingusing arenal disease gene panel in74 unrelated patients with undiagnosed renal disease. This cohort includes patients with glomerular disease (34 cases), cystic renal disease (16 cases), electrolyte disorder (11 cases) and renal insufficiency with unknown etiology (13 cases). This analysisled to pathogenic mutations in 11 cases (15%), and in 12 other cases (16%) likely pathogenic variants needed follow-up studies. Further analysis of the complete exome data set in 19 patients, revealed novel candidate genes in six cases that are under investigation. In addition, copy number variation analysis of exome data revealed a pathogenic deletion of the *NPHP1* gene in two cases, confirming the clinical diagnosis. We conclude that diagnostic exome sequencing provides a powerful tool for detecting causative mutations in up to31 out of74 patients (40%) with a renal disease.

1981W

Improved Sensitivity and Rapid Confirmation of Variants via Orthogonal Sequencing of Exomes. *J. Thompson[1], N. Chennagiri[1], A. Frieden[1], D. Lieber[1], E. Lopez[1], T. Ross[1], T. Yu[1,2].* 1) Res & Development, Claritas Genomics, Cambridge, MA; 2) Boston Childrens Hospital, 300 Longwood Ave, Boston, MA.

Whole exome and genome sequencing studies have revolutionized the diagnosis of genetic disorders by providing information about pathogenic variants. However, each next-generation sequencing (NGS) platform has limitations in terms of coverage and error profiles. In addition, the large number of variants identified in each patient makes confirmation and interpretation of the data very challenging. ACMG practice guidelines recommend that “all disease-focused and/or diagnostic testing include confirmation of the final result using a companion technology”, a daunting task given the number of variants involved. Thus, the ability to confirm many variants quickly is critical. To maximize sensitivity and to minimize the time needed, Claritas Genomics has developed an approach using complementary and orthogonal sequencing technologies to analyze variants from whole exome sequences. Interpretation is restricted to phenotypically-defined, clinically relevant regions of interest (ROI). Each individual’s exome is sequenced using two separate DNA capture and sequencing methods. DNA is captured using a hybridization-based approach and sequenced using Illumina NGS. In parallel, DNA is captured using an amplification-based approach and sequenced using Ion Torrent NGS. Because independent and complementary technologies are used, the need for Sanger sequencing for confirmation is nearly eliminated. For a 2. 2 Mb neurologically-focused ROI, 91% of the variants are detected on both platforms and thus are orthogonally confirmed. When this analysis is carried out on DNA with an extensive truth set (NA12878), 100% of the orthogonally confirmed variants are true positives, supporting this approach. In addition, 8% of the true variants are found on only one platform and not called on the other due to insufficient coverage. The use of two capture methods and sequencing platforms extends the region over which variants can be detected so the region examined is more than either platform alone is capable. This improves sensitivity of detection, particularly with insertions and deletions because both platforms miss a significant number of such variants. Because the majority of variants are orthogonally confirmed and interpreted for pathogenicity promptly, there is no delay for Sanger confirmation and a rapid report can be provided to the physician. This provides the potential for physicians to act quickly on information with patients benefitting from timely feedback.

1982T

Experiences with the dissemination of secondary findings by diagnostic exome sequencing. *H. G. Yntema[1], M. Nelen[1], M. van Koolwijk[1], F. van Agt[2], N. Steinkamp[3], A. Mensenkamp[1], L. Haer-Wigman[1], L. E. L. M. Vissers[1], F. P. M. Cremers[1], A. van den Wijngaard[4], C. Marcelis[1], I. Feenstra[1].* 1) Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; 2) Ethics Committee 'CMO regio, Arnhem-Nijmegen', Nijmegen, The Netherlands; 3) Department of Ethics, Philosophy and History of Philosophy, Radboud University Medical Center Nijmegen, Nijmegen, The Netherlands; 4) Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands.

In the Genome Diagnostics laboratory at the Radboudumc Nijmegen, whole exome sequencing (WES) has been performed for more than 4,000 patients. The WES diagnostic process is a two-tiered approach, in which tier 1 is limited to the analysis of genes known for the respective disease of the patient (*in silico* gene panel approach) and tier 2, performed if tier 1 is negative and informed consent is given, the analysis of the entire protein coding sequence. This second tier has the intrinsic risk of identifying secondary findings: mutations in disease genes not implicated in the aetiology of the disease the patient is tested for. As part of WES implementation as a diagnostic test, we set up a procedure on how to deal with secondary findings. Upon identification, a multidisciplinary committee of experts can be assembled on an *ad hoc* basis to discuss the dissemination a secondary finding. The committee consists of a clinical laboratory specialist, a clinical geneticist, a molecular geneticist, a social worker, a lawyer, and an ethicist. Other specialists with specific expertise are consulted upon indication. The committee checks informed consent, evaluates the molecular finding in the patient's clinical context, and discusses the potential medical and social impact of the finding. When it is decided (by majority vote) that the finding should be returned, a separate report is issued to the requesting clinician, not only reporting the mutation, but also the arguments for reporting the finding, as well as guidelines on how to further proceed. To date, WES beyond the gene panel of approximately 1,500 cases has been finished. Thirty-five (2. 3%) potential secondary findings have been discussed by the committee. For 21 (1. 4%) of these, mostly involving hereditary cancer and cardiac disease, it was (unanimously) decided to report the finding, based on the clear pathogenicity of the mutation and relevant medical actions that could be taken. The main reason not to report the remaining 14 findings was the unclear pathogenicity of the variant. In our hands, 1. 4% of patients receiving WES is confronted with a secondary finding. The expert committee judging these findings is considered beneficial for all laboratory specialists and clinical geneticists of our department. Furthermore, this committee is now also participating in most research projects within our hospital, suggesting that there is a wider need for discussing secondary findings than in a diagnostic setting alone.

1983F

Laboratorial diagnosis of the Fragile X syndrome: a fast, reliable and effective methodological approach supporting the Genetic Counseling services. *G. Molfetta[1,2,3], A. Marques[2,3], C. Furtado[4], E. Ramos[1], W. Silva Jr[1,2,3].* 1) Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Brazil; 2) Center for Genomic Medicine, University Hospital of Ribeirão Preto Medical School, University of São Paulo, Brazil; 3) National Institute of Science and Technology in Stem Cell and Cell Therapy, Regional Blood Center of Ribeirão Preto, Brazil; 4) Department of Obstetrics and Gynecology, Ribeirão Preto Medical School, University of São Paulo, Brazil.

Fragile X syndrome is the most common form of inherited intellectual disability characterized by loss-of-function of FMR1 caused by an increased number of CGG trinucleotides repeats accompanied by its aberrant methylation. It affects men and women with frequencies of 1:4000 and 1:6000, respectively. Mental disabilities in affected individuals ranges from learning disabilities to autism and severe mental retardation, and may be accompanied by a variety of other physical and behavioral characteristics, making the clinical diagnosis more difficult. Our goal was to offer a fast, reliable, effective and not expensive laboratorial diagnosis of Fragile X syndrome. During one year, we evaluated 186 patients under the clinical suspicion of Fragile X syndrome who have consented for the blood taken. Of these, 160 were male and 26 were female relatives of the probands. The methodological approach applied in the present study was 1) for the male diagnosis we evaluated the methylation pattern of the promoter region of the FMR1 gene by High Resolution Melting (MS-HRM) after DNA bisulfite conversion and 2) for the diagnosis of female carriers we employed the AmplideX FMR1 kit (Asuragen) allowing the exact determination of the number of CGG repeats. The MS-HRM method was able to confirm the suspicion of Fragile X in 10% (16/160) of the cases; twelve of them presented a mosaic pattern showing both methylated and non-methylated peaks. In order to confirm methylation mosaicism, the DNA genotyping was carried out in quantitative fluorescence polymerase chain biplex reactions (QF-PCR) using primers for AR and RP2 genes on X-chromosome. The QF-PCR results showed the presence of only one X-chromosome in the male patients. The AmplideX FMR1 kit enabled the diagnosis in 23% (6/26) of the female carrying more than 200 CGG repeats; five patients (19%) were premutation carries showing an allele characterized by more than 55 CGG repeats, and fifteen women (58%) had normal pattern of CGG repeats. Furthermore, one female patient presented the full-mutation allele in addition to the premutation allele characterizing this patient as an important subject to the Genetic Counseling. The molecular approach applied in the present study was informative to 15% of the cases under clinical suspicion of the Fragile X syndrome enforcing this as very reliable and fast methodology. Given the above results, we suggest the use of this approach for genetic counseling of the inherited mental disability.

1984W

Clinical implementation of *in silico* gene panel testing for clinically heterogeneous disorders using exome sequencing. *R. K. Basran, C. R. Marshall, A. Shlien, M. Eliou, J. Orr, L. Lau, D. J. Stavropoulos, P. N. Ray.* Molecular Genetics Laboratory, Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Canada.

The implementation of next generation sequencing technologies (NGS) in molecular diagnostics has resulted in a major transformation in service delivery. The ability to generate accurate sequence data from thousands of genes in a single experiment enables the development of comprehensive gene panel tests for genetically heterogeneous disorders that were previously not feasible by Sanger sequencing. At The Hospital for Sick Children (Toronto, Canada), we have developed and clinically validated expanded *in silico* gene panels using whole exome sequencing (WES) for several heterogeneous disorders including hereditary spastic paraplegia (HSP), connective tissue and bone disorders (CT), hereditary hearing loss (HL), and autoimmune disorders (AI). Previously for HL and HSP our laboratory offered a smaller subset of genes tested through Sanger sequencing. CT and AI disorders were not tested in our laboratory and were previously sent out of Canada for testing. NGS technologies have enabled our laboratory to expand our testing menu to include genes that comprise part of the differential diagnoses for these disorders in order to improve detection rates. Our laboratory's current strategy of using WES has several advantages including: (1) standardized workflows that facilitate the development of additional gene panels; (2) expansion of current gene panel as new disease associations are discovered; and (3) building of an internal database of allele frequencies to aid in interpretation of variants. Within the WES workflow, we have automated all informatics analysis, built both a database to store sequence data and a user friendly website for display and interpretation of variants. Through our validation studies we have shown high reproducibility and accuracy with >99% sensitivity for detection of single nucleotide variants (SNVs) and >95% for detection of small insertion/deletions (indels). The broadening of molecular testing for genetically heterogeneous disorders will lead to increased genetic diagnosis and ultimately better clinical care. .

1985T

Analytical validation of a saliva collection and DNA extraction protocol for a 25-gene hereditary cancer panel. *J. Cassiano, W. J. Landon, E. Goossen, D. Mancini-DiNardo, B. Leclair, B. B. Roa.* Myriad Genetics, Inc. , Salt Lake City, UT.

Background: Early identification of individuals carrying genetic mutations in cancer pre-disposing genes strongly impacts clinical management, significantly reducing morbidity and mortality from associated cancer syndromes. Analysis of DNA extracted from blood is the gold standard of germline molecular diagnostic testing. However, blood venipuncture can present challenges in some cases due to lack of on-site phlebotomy services, patient non-compliance, or inability for some patients to provide a sample due to poor venous access or needle-anxiety. DNA extracted from saliva can prove to be a viable alternative to venipuncture. We describe validation data for the collection and DNA extraction of saliva for use with our pan-cancer panel (PCP) test, an NGS-based sequence and large rearrangement (LR) detection platform for a panel of 25 genes associated with 8 cancer syndromes. **Methods:** 152 paired blood/saliva Population Control (PC) samples from healthy individuals and 29 saliva samples from previously tested patients who carry LRs in various genes (24 deletions, 3 duplications, 1 insertion, 1 triplication) were obtained from consented individuals. DNA was extracted and subjected to analysis for sequence variants and LRs. **Results:** NGS sequencing results based on analysis of 5,380 sequence variants in PCP genes found among 138 patients were 100% concordant between paired blood and saliva samples. 100% concordance was also observed between NGS and Sanger sequencing performed on a subset of paired samples. In combination with data from our NGS validation of blood DNA, the 9,303 total sequence variants found in PCP genes among 238 individuals yield an analytical sensitivity of 100% (99. 96%-100%, 95% C. I.) and specificity of 100% (99. 99%-100%, 95% C. I.) for the sequencing component of our test. NGS LR detection performed on 119 PC paired samples showed 100% result concordance between paired blood and saliva samples. The NGS results on 29 LR+ saliva samples showed 100% concordance with LR results previously reported based on testing blood samples. 100% concordance was also observed for saliva samples using orthogonal LR confirmation methods (aCGH, MLPA). **Conclusions:** There was 100% concordance in sequence variant and LR detection between blood and saliva samples. The development and validation of this saliva sample type protocol may lead to more accessible PCP testing, especially when blood venipuncture is a challenge.

1986F

Performance of ACMG variant classification guidelines within and across 9 CLIA labs in the Clinical Sequencing Exploratory Research (CSER) Consortium. *G. P. Jarvik[1], L. A. Amendola[1], H. McLaughlin[2], M. C. Leo[3], A. Milosavljevic[4], C. Horton[5], R. Ghosh[4], M. Dorschner[1], S. Punj[6], Y. Akkari[6], J. Salama[1], G. Cooper[7], L. Biesecker[8], L. K. Conlin[9], S. Biswas[9], M. Dulik[9], A. Ghazani[10], N. T. Strande[11], Y. Yang[4], E. Van Allen[12], N. Wagle[12], R. C. Green[13], A. Chinnaiyan[14], J. S. Berg[11], J. P. Evans[11], L. Garraway[12], N. Spinner[9], S. E. Plon[4], S. Richards[6], H. L. Rehm[2].* 1) University of Washington, Seattle WA; 2) Partners Laboratory of Molecular Medicine, Boston MA; 3) Kaiser Permanente, Portland OR; 4) Baylor College of Medicine, Houston TX; 5) Ambry Genetics, Aliso Viejo CA; 6) Oregon Health Sciences University, Portland OR; 7) HudsonAlpha Institute for Biotechnology, Huntsville AL; 8) National Human Genome Research Institute, Bethesda MD; 9) Children's Hospital of Philadelphia, Philadelphia PA; 10) Massachusetts General Hospital, Boston MA; 11) University of North Carolina, Chapel Hill NC; 12) Dana Farber Cancer Institute, Boston MA; 13) Brigham and Women's Hospital, Boston MA; 14) University of Michigan, Ann Arbor MI.

There is a need to standardize classification of genomic variants in medical sequencing. To date genomics laboratories have used non-standard classification systems. We have reported both inconsistency across labs in variant classification and a bias towards overcalling pathogenicity (PMID:25637381). The ACMG published guidelines for variant classification for Mendelian disorders (PMID:25741868) designed to increase consistency among labs. The Clinical Sequencing Exploratory Research (CSER) Consortium evaluated the use of these rules by 9 of the CLIA labs supporting CSER projects. 99 variants were considered, representing all categories (pathogenic, likely pathogenic, uncertain significance (VUS), likely benign and benign). 9 were classified by all 9 labs, using both the lab's own classification system and also the ACMG guidelines. 90 variants were classified by 2-3 labs (mean of 2. 85) again using both the lab and ACMG guidelines. The results were examined to evaluate intra-laboratory differences between variant classifications using the labs own criteria vs. adopting ACMG criteria and inter-laboratory differences using either the lab's own system or the ACMG guidelines. Of 335 paired variant calls, the intra-laboratory comparison of lab vs. ACMG criteria matched for 268 (80%). For those 67 discordant calls, the ACMG calls were closer to VUS in 45 (67%). Considering the inter-laboratory agreement for 98 variants, all reviewing labs (range 3-9) agreed for 36 (37%) using the ACMG and 38 (39%) using laboratory specific criteria. The Table summarizes the counts for the range of differences in classification by criteria, where "1 step" that the largest difference was 1 level of pathogenicity order, for example a range of pathogenic to likely pathogenic. Agreement among labs did not differ whether using the lab specific vs. ACMG criteria (p=0. 9); i. e. the ACMG criteria did not yield more consistent variant classification in this exercise. We further analyzed sources of disagreement in the use of the ACMG criteria and identified causes of variance in classifications that will be presented in more detail. In addition to providing useful analyses of how variant classifications approaches vary among labs, these data should allow clarification and refinement of the ACMG criteria that may increase consistency in variant classification.

Counts for inter-lab differences in classification by criteria					
Criteria	None	1 step	2 step	3 step	4 step
ACMG	36	42	19	1	0
Lab	38	41	17	1	1

1987W

Bias has no place in exome sequencing. *K. B. Pechter[1], A. I. Nesbitt[1], E. H. Denenberg[1], S. W. Baker[1], Z. Yu[1], E. T. Dechene[1], E. C. Bedoukian[2], M. Luo[1], L. K. Conlin[1,3], A. Wilkens[2], L. Medne[2], I. D. Krantz[2,3], M. A. Deardorff[2,3], A. B. Santani[1,3].* 1) Division of Genomic Diagnostics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Exome sequencing provides a 20-25% diagnostic rate in identifying the molecular basis for rare genetic disorders, however, there are challenges associated with the interpretation of such complex tests. In particular, we will discuss the importance of an unbiased whole exome sequencing (WES) analysis workflow that leverages detailed phenotypic capture, inheritance modeling, and comprehensive review of the literature. WES was performed on trios and data were analyzed by an in-house bioinformatics pipeline. Gene-associated variants were clinically correlated by both genetic scientists and clinicians, and variants were classified using ACMG guidelines. Case 1: An infant male proband presenting with seizures and complex brain malformations (including polymicrogyria, underdeveloped sylvian fissures, and hypoplastic pons) was analyzed by WES. We discovered a *de novo* missense mutation in the *GRIN2B* gene; *de novo* heterozygous variants in *GRIN2B* cause an autosomal dominant (AD) epileptic encephalopathy phenotype. However, this finding does not explain the complex brain malformations seen in this patient. Ultimately, it is unclear whether or not this case expands the phenotypic spectrum of *GRIN2B*-associated conditions or if the resolution power of WES was insufficient to identify a causative variant for brain malformations in this patient. This case highlights the importance of using an unbiased approach to analyze all data and explore the possibility of two molecular diagnoses. Case 2: A 7 year-old female proband presenting with microcephaly, spasticity, proportionate short stature, osteopenia with multiple fractures, and global developmental delay was analyzed by WES and a*de novo* variant in the *CTNNB1* gene was identified. Pathogenic variants in *CTNNB1* are associated with an AD mental retardation syndrome that overlaps with the majority of the patient's findings but does not explain the osteopenia and fractures. Investigation of the literature revealed heterozygous *CTNNB1* variants cause osteoporosis in mouse models, suggesting this was the molecular diagnosis for these features as well. This case emphasizes the importance of objective phenotype assessment followed by careful review of the literature to establish the full extent of the molecular diagnosis. Overall, these cases highlight some of the complexities that are inherent to the analysis and interpretation of WES data, and speak to the importance of an unbiased approach in executing this complex test.

1988T

Cytogenomic and molecular refinement of *STRC* and *CATSPER2* deafness-infertility syndrome deletion breakpoints. *L. Shi[1], Y. Kharbutli[1], J. Reiner[1], N. Cohen[1], A. M. Oza[2], S. Amr[2], L. Edelmann[1], L. Mehta[1], S. A. Scott[1].* 1) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Laboratory for Molecular Medicine, Partners HealthCare, Cambridge, MA.

Autosomal recessive deafness-16 (DFNB16) is caused by homozygous or compound heterozygous mutations in the *STRC* gene on chromosome 15q15; however, interrogating the gene is challenging due to a highly homologous ~100 kb segmental duplication distal to *STRC*. Larger homozygous deletions including *STRC* and the neighboring *CATSPER2* result in deafness-infertility syndrome, which is characterized by early-onset deafness in both males and females, and exclusive male infertility. Deletions of this region are typically detectable by clinical microarray-based comparative genomic hybridization (aCGH) testing; however, the lack of specific copy number probes throughout *STRC* can result in uncertainty of deletion size, affected genes, and predicted phenotypes. The need for higher resolution molecular methods for investigating this region was highlighted by our case of a 39 year-old female referred for prenatal genetic testing due to advanced maternal age and a previous pregnancy with trisomy 22. Chromosome analysis of the chorionic villus sample indicated a normal female karyotype; however, parental request for prenatal aCGH testing (180K) revealed a homozygous 33. 74 kb deletion of chromosome 15q15. 3 [arr 15q15. 3(43916972-43950720) x0; hg19] that included *CATSPER2*. The homozygous deletion was confirmed by a higher density array (244K) and droplet digital PCR targeting both *STRC* (introns 25 and 23) and *CATSPER2* (intron 7 and exon 7); however, it was not possible to determine if the homozygous deletion included the 5' region of the *STRC* gene. Parental aCGH testing revealed that *STRC* was deleted in the paternal allele, but loss of the 5' region of *STRC* could not be ruled out in the smaller maternal deletion. To better resolve the maternal deletion breakpoint, a two-tiered *STRC* allele specific-PCR assay was developed that narrowed the breakpoint to a 310 bp region located 440 bp upstream of the *STRC* transcription start site, indicating that the fetus carried one intact copy of *STRC*. The baby was born naturally at 39. 5 weeks and passed newborn hearing screening; however, childhood hearing monitoring is planned given the possibility of hemizygous *STRC* promoter dysregulation and the variable age of onset of DFNB16. Taken together, these data highlight the challenges in prenatal detection of homologous disease gene regions and underscore the need for higher resolution molecular methods for selected disease genes to better inform phenotype prediction and genetic counseling.

1989F

Novel compound heterozygous *LIAS* mutations cause glycine encephalopathy. *Y. Tsurusaki[1,2], M. Nakashima[2], H. Saitsu[2], N. Miyake[2], N. Matsumoto[2].* 1) Kanagawa Children's Medical Center, yokohama, JAPAN; 2) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, JAPAN.

Glycine encephalopathy (GCE) (MIM 605899), also known as nonketotic hyperglycinemia, is a rare autosomal recessive disorder of the glycine cleavage system that results in elevated glycine levels in body fluids. GCE is genetically caused by mutations in *AMT* (MIM 238310), *GLDC* (MIM 238300) or *GCSH* (MIM 238330), which encode T-, P- or H-protein, respectively, and together with L-protein constitute the glycine cleavage enzyme. Most GCE patients have *GLDC* mutations. Recently, *GLRX5* (MIM 609588), *BOLA3* (MIM 613183) and *LIAS* (MIM 607031) were identified as causative mutant genes for GCE. To date, three homozygous mutations have been reported in *LIAS*. All previously reported GCE patients showed elevated level of glycine in the serum. LIAS encodes lipoic acid synthetase, a [4Fe-4S]-type iron–sulfur cluster-dependent enzyme that acts as a prosthetic group on H-protein, which is localized within the mitochondria. *GLRX5* and *BOLA3* are iron–sulfur cluster biosynthesis genes. Here we report a patient with GCE and elevated level of glycine in both the serum and the cerebrospinal fluid. Trio-based whole-exome sequencing identified novel compound heterozygous mutations (c. 738-2A4G and c. 929T4C (p. Met310Thr)) in *LIAS*. Our data further supports *LIAS* mutations as a genetic cause for GCE. Aknowledgments: Drs. R. Tanaka, S. Shimada, K. Shimojima, M. Shiina, K. Ogata, T. Yamamoto are highly appreciated for contributing to this work.

1990W

Case series of colorectal cancer patients with BRCA1/2 mutations: Finding actionable genes in patients with atypical presentations. *K. Vikstrom, S. Yang, E. Esplin.* Invitae Corporation, 475 Brannan Street, San Francisco, CA.

Background: A new paradigm in genetic testing for hereditary colorectal cancer (CRC) risk has emerged. With next-generation sequencing (NGS), clinicians can choose to test 7 high-penetrance genes associated with CRC or more comprehensive panels of 30+ cancer genes for roughly the same cost. In this case series, we describe 6 patients who had unclear or overlapping features of different hereditary cancer syndromes and who did not meet NCCN testing guidelines for a specific syndrome. In each case, a gene panel detected a mutation in BRCA1 or BRCA2. Association of CRC with pathogenic variants in BRCA1/2 has been suggested (but not yet extensively demonstrated). Nevertheless, these findings in BRCA1/2 were actionable and would be missed by gene-specific tests under traditional criteria. **Methods:** Our study included 585 consecutive patients with an indication of CRC and/or gastrointestinal (GI) polyps, referred for panel testing at Invitae. Genomic DNA variants were identified using a 7- to 34-gene NGS-based hereditary cancer panel; the clinician's discretion determined panel size. Germline sequence variants and deletions/duplications were classified using a point-based system that closely adheres to ACMG guidelines. Patients' personal and family histories were obtained from test request forms and were de-identified for this analysis. **Results:** Hereditary cancer panel testing found a Likely Pathogenic (LP) or Pathogenic (P) variant in 92 of 585 (15%) patients. Of the 92 mutation carriers, 69 (75%) had a P/LP variant in a known CRC gene (APC, MUTYH, MLH1, MSH2, MSH6, EPCAM, or PMS2), while 6 (6%) had an P/LP variant in BRCA1/2. None of the patients with BRCA1/2 mutations, 2 females and 4 males, reported Ashkenazi Jewish ancestry. **Conclusions:** In this series of patients with an indication of CRC and/or GI polyps, 24 (25%) of LP/P variants identified were in non-canonical CRC genes and, of those, 6 cases (25%) had mutations in BRCA1/2 in the absence of other LP/P variants. The prevalence of BRCA1/2 findings is alone not strong evidence of causal association, however, it is greater than would be expected by random chance and these patients would have been found negative by traditional genetic tests. While more research is needed to understand the relationship between BRCA1/2 and colon cancer, physicians need to be prepared to deal with actionable BRCA1/2 results which are present in patients with CRC, particularly in those patients with atypical presentations.

.

1991T

Whole genome sequencing as a clinical diagnostic tool for heterogeneous Mendelian disease. *J. M. Ellingford[1,2], S. Barton[1], S. Bhaskar[1], S. G. Williams[1], P. I. Sergouniotis[1,2,3], J. O'Sullivan[1,2], J. A. Lamb[4], R. Perveen[1,2], G. Hall[2], W. G. Newman[1,2], P. N. Bishop[2,3], S. A. Roberts[5], S. Bayliss[1], S. C. Ramsden[1], A. H. Nemeth[6], G. C. M. Black[1,2,3].* 1) Manchester Centre for Genomic Medicine, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Sciences Centre, St Mary's Hospital, Manchester M13 9WL, UK. Manchester, United Kingdom; 2) Institute of Human Development, University of Manchester, Oxford Road, Manchester, M13 9WL, UK; 3) Manchester Royal Eye Hospital, Manchester Academic Health Sciences Centre, Central Manchester University Hospitals NHS Foundation Trust, Oxford Road, Manchester, M13 9WL, UK; 4) Institute of Population Health, University of Manchester, Oxford Road, Manchester, M13 9PT, UK; 5) Centre for Biostatistics, Institute of Population Health, University of Manchester, Oxford Road, Manchester, M13 9PL, UK; 6) Nuffield Department of Clinical Neurosciences, John Radcliffe Hospital, University of Oxford, Oxford, OX3 9DU, UK.

Background: Whole genome sequencing (WGS) enables the analysis of genome-wide genetic variation that can range from single nucleotide alterations to large structural genomic variants. WGS has become more affordable and presents dramatic opportunities for a potential shift in diagnostic services for genetic disease. However, diagnostic WGS pipelines are yet to be validated as a clinical tool in the context of heterogeneous genetic disease.

Methods: We surveyed the diagnostic yield of an accredited, conventional next-generation sequencing (NGS) diagnostic testing method for 537 patients with genetically and clinically heterogeneous single gene disorders (specifically, inherited-retinal-disease). A stratified subset of 47 patients also underwent WGS, and we compare the diagnostic yield of a WGS pipeline to that achieved by the conventional NGS diagnostic test.

Findings: Over the clinical analysed region, the WGS pipeline achieved similar sensitivity and specificity rates to the targeted NGS diagnostic testing method. Importantly, the WGS pipeline detected pathogenic variants that were not identified by current NGS diagnostics and, amongst the 47 patients, facilitated an increase in diagnostic yield.

Interpretation: We demonstrate the capability of a WGS pipeline to detect pathogenic genetic variants missed by current methodologies,and to identify essential modifications for current and future molecular diagnostic practice. We establish the potential utility of diagnostic WGS pipelines for clinically and genetically heterogeneous Mendelian disorders.

1992F

Genome versus exome sequencing: Is WGS the better WES? *K. Oexle[1], J. Meienberg[1], K. Zerjavic[1], B. Roethlisberger[2], R. Bruggmann[3], G. Matyas[1].* 1) Center for Cardiovascular Genetics and Gene Diagnostics, Foundation for People with Rare Diseases, Schlieren-Zurich, Switzerland; 2) Division of Medical Genetics, Center for Laboratory Medicine, Aarau, Switzerland; 3) Interfaculty Bioinformatics Unit and Swiss Institute of Bioinformatics, University of Berne, Berne, Switzerland.

The use of next generation sequencing (NGS) in genetic diagnostics currently involves enrichment tools that capture fractions of the genome such as the set of all coding gene segments (exome). Recently, we compared the performance of several enrichment platforms for whole exome sequencing (WES) showing that all these platforms had problems in sufficiently (≥20×) representing the RefSeq coding region, especially GC-rich exons. We have now compared the performance of WES (Agilent SureSelect Human All Exon kit v5+UTR) with whole genome sequencing (WGS), especially WGS without amplification for library preparation (PCR-free WGS; Illumina HiSeq X Ten). PCR-free WGS turned out to be insensitive to GC content and to be more complete and homogenous in coverage of coding exons. Although the average read depth of PCR-free WGS (mean ± 95% confidence interval), 64 ± 7 reads) was roughly half of the average read dept of WES (146 ± 15 reads), the number of RefSeq genes whose coding parts were both completely and sufficiently covered (≥20×) was 6% higher in PCR-free WGS than in WES (99% versus 93%, p<0. 05). As such, WES did not seem to sufficiently serve its purpose. Additional data indicate that this shortcoming of WES cannot be amended by further increasing read depth since coverage distribution remains highly uneven. Hence, the advantage of WGS does not only include the potential of identifying non-coding pathogenic variation but, in view of its more homogenous and complete exomic coverage, WGS may simply be the better WES. Thus, as the costs of WGS decline steadily and the costs of sequence analysis can be curtailed by selecting *in silico* the WGS parts that are of diagnostic interest (virtual WES and gene panels), WES might soon be replaced by non-enriching and PCR-free WGS. Moreover, as genome-wideness and homogeneity of WGS makes it also suitable for the detection of copy number variation (CNV), we have assessed the potential of WGS to replace array techniques in CNV detection as well.

1993W

Diagnostic Exome Sequencing Identifies Alterations in the Newly Characterized Gene, *COQ4*, Expanding the Phenotypic Spectrum. *M. Tsang[1], D. Shinde[1], W. Alcaraz[1], D. El-Khechen[1], V. Bhambhani[2], N. Mendelsohn[2], B. Crabb[2].* 1) Ambry Genetics, Aliso Viejo, CA; 2) Childrens Hospitals and Clinics of Minnesota, MN.

Recently, biallelic mutations in *COQ4* have been reported in five individuals from four unrelated families with CoQ10 deficiency resulting in a wide phenotypic spectrum of mitochondrial disorders. Here we report an additional non-consanguineous family with two affected siblings with compound heterozygous mutations in *COQ4.* The proband is a 16 year old Caucasian girl with infantile onset seizures, severe intellectual disability, microcephaly, short stature, visual cortical dysfunction, neuromuscular scoliosis with restrictive lung disease, pancytopenia, recurrent infections, and epistaxis. She also has dysmorphic features including a prominent nose, extended nasal columella, small mouth, small hands and feet with significant contractures, and small limbs. Her 18 year old brother is similarly affected. A mitochondrial disorder due to autosomal recessive nuclear gene defect was suspected. Multiple genetic and metabolic tests were uninformative. Family centered diagnostic exome sequencing (DES) on the proband and her healthy parents revealed compound heterozygous c. 469C>A (p. Q157K) and c. 202G>C (p. D68H) missense alterations in *COQ4*. Both alterations are expected to be likely pathogenic, based on *in silico* and co-segregation data. Co-segregation analysis revealed that the affected brother was also compound heterozygous for both alterations and an unaffected brother was a heterozygous carrier of the p. D68H alteration. Consistent with a mitochondrial disorder, the patient's lactate and pyruvate normalized following CoQ10 supplementation. *COQ4* is a newly characterized gene that encodes a mitochondrial protein involved in the organization of a multienzyme complex for the biosynthesis of Coenzyme Q10 (CoQ10). CoQ10 is an essential mitochondrial inner membrane-associated lipid that acts as a carrier for electrons from respiratory complexes I and II to complex III. Previously reported patients with *COQ4* associated CoQ10 deficiency have varied clinical presentations ranging from early neonatal death due to acrocyanosis, bradycardia, hypotonia and respiratory insufficiency to hypertrophic cardiomyopathy, arthrogryposis, cerebellar hypoplasia, severe myoclonic epileptic encephalopathy, progressive ataxia, and scoliosis. Here we demonstrate the utility of family centered DES to diagnose a patient with a complex phenotype and expand the phenotypic spectrum of *COQ4* associated CoQ10 deficiency.

1994T

Exome sequencing of patients with primary immunodeficiencies; translational genomics with direct clinical implications. *P. Arts[1], A. Simons[1], A. R. Mensenkamp[1], W. A. van Zelst-Stams[1], J. A. Veltman[1], F. L. van de Veerdonk[2], A. Hoischen[1], M. G. Netea[2].* 1) Departement of Human Genetics, Radboud University Medical Center, Nijmegen, Netherlands; 2) Department of Internal Medicine, Radboud University Medical Center, Nijmegen, Netherlands.

Primary immunodeficiencies (PIDs) are a largely heterogeneous medical condition caused by defects in the innate immune response. The insufficient protection against pathogens leads to recurrent infections in patients suffering from PIDs. Severely affected patients are selected for diagnostic exome sequencing as we expect involvement of a genetic component besides the obvious environmental factors. In the diagnostic setting exome data are filtered for rare protein altering variants in the ~270 genes known to cause PIDs. Since May 2013 analysing this gene package resulted in genetic diagnosis for 10% of over 60 cases; an additional 15% of cases carry a single mutation in a recessive PID gene. 80% of all patients without a genetic diagnosis gave permission for exome-wide analysis in a research setting. Focusing on genes with a known role in immunological pathways led to identification of candidate variants in a large number of patients. The research setting allowed follow-up of the genetic variants on the genetic and functional level. Further genetic analyses included search for recurrently mutated genes in additional patients with similar phenotypes, and performing co-segregation analysis in families. In addition, *in vitro* stimulations of peripheral blood mononuclear cells obtained from patients and controls, and comparing specific cytokine production capacity provides additional functional evidence for the immunological defect. One example with direct clinical implication of this approach was characterization of the genetic, and functional defect in a patient with autoinflammatory symptoms and *SOCS4* mutation (Arts *et al.* 2015). In another example we identified recessive *BAG3* mutations in a patient with progressive multifocal leukoencephalopathy without apparent immunosuppression, and also identified an underlying interferon-gamma deficiency. For other cases the genetic or functional data provided sufficient evidence to change treatment, *e. g.* in two cases with recurrent viral infections interferon-gamma supplementation therapy led to a dramatic improvement in the respective patients' health. In summary, we believe that the combination of clinical and research exome analysis, together with functional immunophenotyping offers a great opportunity for improved diagnostics and ultimately treatment of patients with PIDs.

1995F

Expanding the phenotypic and mutational spectrum of *DYRK1A* – a rare cause of intellectual disability and microcephaly. *A. Knight Johnson[1], A. Blumenthal[2,3], E. Marsh[4], H. Dubbs[4], F. Kobiernicki[1], L. Ma[1], K. Arndt[1], L. Wysinger[1], G. Alkorta Aranburu[1], S. Das[1].* 1) Human Genetics, The University of Chicago, Chicago, IL; 2) Lakeridge Health Oshawa, Oshawa, ON; 3) University of Toronto, Toronto, ON; 4) Children's Hospital of Philadelphia, Philadelphia, PA.

The *DYRK1A* gene encodes a dual-specificity phosphorylation-regulated kinase, and is located within the Down syndrome critical region on chromosome 21. Mutations in this gene have been previously associated with autosomal dominant intellectual disability. All *DYRK1A* mutations described to date are null mutations, and typical features in affected individuals include congenital microcephaly, growth delay, severe intellectual disability, febrile seizures, and mild dysmorphic features including large ears and underdeveloped ear lobes. Out of 308 patients referred to our laboratory for panel testing for genes associated with either non-specific intellectual disability or microcephaly, we identified 2 patients with pathogenic variants in *DYRK1A* (0. 65%), indicating that pathogenic variants in this gene are a rare cause of intellectual disability and microcephaly. One of the pathogenic variants observed was identified in a patient with features typical of what has previously been described in patients with a mutation in *DYRK1A*, including congenital microcephaly, growth delay, intellectual disability, and a history of febrile seizures. The variant identified in this patient created a premature stop codon, p. Arg437*, consistent with the previously reported mutational spectrum for the *DYRK1A* gene. The second pathogenic variant identified was a missense variant, p. Phe308Val, which was observed in the *de novo* state in a 6 year old patient with relative microcephaly, developmental delays, and no history of growth delay or seizures. The p. Phe308Val variant affects a highly conserved amino acid within the tyrosine-kinase catalytic domain of the DYRK1A protein. This is the first report of a pathogenic missense variant in *DYRK1A*, and therefore expands the known mutational spectrum of this gene. The phenotypic features observed in this patient are less severe than what has been reported to date in patients with null mutations in the *DYRK1A* gene, suggesting a possible genotype-phenotype correlation.

1996W

A family with X-linked creatine transporter deficiency diagnosed by whole exome sequencing. *M. Mori, JL. Goldstein, SP. Young, DS. Millington, DD. Koeberl.* Division of Medical Genetics, Duke University Health System, Durham, NC.

X-linked creatine transporter deficiency (CRTR-D) is characterized by cerebral creatine deficiency, leading to intellectual disability, severe speech impairment, seizures, and behavioral issues. The diagnosis is suggested by marked elevation of urinary creatine/creatinine ratio or deficient cerebral creatine on magnetic resonance spectroscopy, and confirmed by identification of a pathogenic mutation in the *SLC6A8* gene. We report a male proband who was diagnosed with CRTR-D, and two younger affected brothers who were subsequently diagnosed. The proband presented to a Metabolic Clinic at five years of age with global developmental delay, absence of speech, epilepsy, recurrent vomiting and feeding difficulty requiring gastric tube feeding, delayed gastric emptying, and severe constipation. His prenatal and birth history were unremarkable. Previous biochemical and genetic testing including CGH microarray had revealed no cause for his symptoms. Urine creatine was not tested. Whole exome sequencing (WES) was performed on a research basis, and a maternally-inherited hemizygous mutation (c. 321_323delCTT; p. Phe107del) in the *SLC6A8* gene was identified. The diagnosis of CRTR-D was confirmed by a markedly elevated urine creatine/creatinine ratio. Both of his younger affected brothers had a mildly elevated creatine/creatinine ratio as newborns, which increased markedly at a few months to confirm the diagnosis. The older affected brother at 14 months of age has developmental delay, hypotonia, poor weight gain and growth requiring gastric tube feeding, myoclonus, and possible left ventricular compaction cardiomyopathy on echocardiogram. The mother's severe constipation may suggest her manifesting heterozygote status. CRTR-D was ruled out in three older brothers. The case illustrates a family with a complex clinical history in which CRTR-D in the proband was diagnosed by WES, and urine analysis of creatine allowed screening of at-risk male relatives. CRTR-D and other creatine synthesis disorders, including arginine:glycine amidinotransferase and guanidinoacetate methyltransferase deficiencies, can be detected by analysis of creatine and its precursor guanidinoacetate in urine, plasma or dried blood spots and therapy has been effective in a subset of affected patients. CRTR-D screening by analysis of urinary creatine in males with developmental delay is strongly recommended.

1997T

Application of multigene panel sequencing in prolonged QT interval patients with no mutation detected in *KCNQ1*, *KCNH2*, and *SCN5A*. *S. Seo, S. Cho, H. Park, S. Lee, M. W. Seong, S. Park.* Departments of Laboratory Medicine, Seoul National University Hospital, Seoul, Korea.

Introduction: Genetic testing for potentially heritable cardiac diseases has advanced from basic scientific discovery to clinical application. A positive genetic testing allows not only confirmation of diagnosis, but also has impact on risk stratification and, in some cases, genotype-based treatment. It is important that genetic testing to be requested with detailed phenotype provided, or it might easily result in a misdiagnosis. Methods: 36 patients with prolonged QT interval patients but no mutation detected in *KCNQ1*, *KCNH2*, and *SCN5A* were included in this study. Molecular genetic testing was performed by using a multigene panel consisted of in 20 genes (*AKAP9, ANK2, CACNA1C, CACNB2, CAV3, GPD1L, HCN4, KCNE1, KCNE2, KCNE3, KCNH2, KCNJ2, KCNJ5, KCNQ1, SCN1B, SCN3B, SCN4B, SCN5A, SLMAP, SNTA1*) known to be associated with inherited arrhythmia. This multigene panel also included around 80 more genes related to other cardiac diseases such as cardiomyopathies. Sequences were captured by TruSeq Custom Enrichment Kit (Illumina) and sequenced with MiSeq (Illumina). Results: Among the variants detected in 20 arrhythmia genes, there were one previously reported mutation of Brugada syndrome (*SLMAP* c. 805G>A), four likely-pathogenic mutations (*ANK2* c. 4259C>T and c. 7503A>C, *CACNA1*C c. 1552C>T and c. 2579G>A) and six variants with unknown significance. This allowed 5 out of 36 patients (13. 9%) to have positive genetic test results by routine interpretation. In addition, genes related to other cardiac disease were examined in patients who had cardiac manifestations other than prolonged QTc. One patient, who carried one of the *ANK2* variant with unknown significance, showed polymorphic ventricular tachyarrhythmia in his EKG. He also carried a novel variant of *RYR2* c. 11995A>G, indicating this variant may be the cause of catecholaminergic polymorphic ventricular tachycardia (CPVT). Another patient, who developed dilated cardiomyopathy with prolonged QTc, were found to carry *TAZ* c. 718G>A variant, a previously reported mutation known to be involved in X-linked infantile dilated cardiomyopathy. Conclusions: As panel sequencing for cardiac diseases became general, sharing of phenotypic information between the clinician and the laboratory is crucial for interpretation of the genetic analysis. Also, well-described mutations should always be taken into consideration, as this might lead to finding the other aspects of the patient.

1998F

A population-based genomic study of inherited metabolic diseases detected through newborn screening. *K. Park[1], S. Park[2], E. Lee[2], J. Park[1], J. Park[3], H. Park[4], S. Lee[4], J. Kim[4].* 1) Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul, Korea; 2) GreenCross Laboratories, Yongin, Korea; 3) Samsung Biomedical Research Institute, Samsung Medical Center, Seoul, Korea; 4) Department of Laboratory Medicine & Genetics, Samsung Medical Center, Seoul, Korea.

Purpose: We aimed to investigate the impact of next-generation sequencing (NGS) on the detection of inherited metabolic diseases (IMDs) in newborn screening (NBS) and to evaluate the genetic epidemiologic characteristics of IMD in the Korean population. **Methods:** A total of 269 dried blood spots with positive results on the current NBS tests were collected from 120,700 consecutive newborns. We screened 97 genes related to NBS using the Ion PGM platform and detected IMDs using the integrated screening model based on biochemical tests and NGS. Haplotype analysis was conducted to detect founder effects in recurrent mutations. **Results:** The overall positive rate of IMDs was 17%. We identified seven additional newborns with preventable IMDs, which would not have been detected prior to the implementation of NGS. The mutation incidence of IMDs was 1 in 2,568 births. Haplotype analysis demonstrated founder effects in p. Y138X in *DUOXA2*, p. R885Q in *DUOX2*, p. Y439C in *PCCB*, p. R285Pfs*2 in *SLC25A13,* and p. R224Q in *GALT*. **Conclusions:** Through a population-based study in the NBS environment, we highlight the diagnostic and epidemiological implications of NGS. We found that the integrated screening model provided better results compared to the current NBS tests because it detected additional cases of treatable IMDs.

1999W

Comprehensive analysis of genes associated with cobalamin deficiency, methylmalonic aciduria and homocystinuria by clinical next generation sequencing identifies novel mutations. *J. Zhang[1], J. Tang[2], Y. Wang[1], F. Li[1], H. Mei[1], J. Wang[1], W. Zhang[1], L. Wong[1].* 1) Molecular and Human Genetics, Baylor College of Medicine, Baylor Miraca Genetics Laboratories, Houston, TX, USA, 77030; 2) Department of Medical Genetics, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China, 510080.

Introduction: Vitamin B12 (cobalamin, Cbl) which consists of a class of chemically related compounds with element cobalt in the center of a planar tetra-pyrrole ring is a co-factor for enzymes involved in chain fatty acids, cholesterol, and amino acids metabolism. Defects in Cbl dynamics including its absorption, transport, and metabolism result in methylmalonic aciduria with or without homocystinuria. A targeted Next Generation Sequencing (NGS) panel developed in a clinical setting was used to identify mutations in genetically heterogeneous methylmalonic aciduria with or without homocystinuria in a cost-efficient manner. **Methods**: 49 probands with clinical or biochemical findings suggesting Cbl deficiency or methylmalonic aciduria with or without homocystinuria were analyzed by a clinically validated Cbl NGS panel. In this panel, 21 genes involved in Cbl dynamics, methionine-homocysteine cycle, and methylmalonyl- and succinyl-CoA conversion were enriched by a customized capture library targeting mainly for coding exons and adjacent region. In addition, over 200 anonymized DNA samples with negative testing results for one or more of these 21 genes were re-analyzed by this Cbl NGS panel. **Results:** The overall detection rate for this Cbl NGS panel with the identification of two allelic (homozygous or compound heterozygous) pathogenic variant in all samples is 42. 8%. Importantly, we identified novel pathogenic variants including *CD320*: c. 386G>T (p. R129L), *GIF*: c. 67_68delCA (p. Q23Efs*86) and *MMACHC*: c. 182G>C (p. R61P) all of which were present in trans-configuration with another pathogenic allele in multiple patients with proper clinical and/or biochemical phenotypes. We also identified a homozygous pathogenic synonymous change in the *MTHFR* gene affecting its splicing resulting in classic severe MTHFR deficiency. This result highlights the importance to include this gene in the Cbl NGS panel to detect severe MTHFR deficiency despite the common and mild-deficiency variants. **Conclusions:** Comprehensive targeted NGS analysis of a panel of 21 genes provides a rapid and cost-efficient method to diagnose genetically heterogeneous methylmalonic aciduria with or without homocystinuria. This method prevails Sanger single gene sequencing as the first tier genetic testing with proper clinical and biochemical indications for Cbl deficiencies.

2000T

The usefulness of array as a clinical diagnostic tool in patients with congenital malformations and developmental delay: Brazilian experience in 162 children. *E. A. Zanardo[1], G. M. Novo-Filho[1], R. L. Dutra[1], M. M. Montenegro[1], A. T. Dias[1], A. M. Nascimento[1], F. A. R. Madia[1], T. V. M. M. Costa[1], F. B. Piazzon[1], A. F. de Carvalho[4], M. I. Melaragno[3], C. A. Kim[2], L. D. Kulikowski[1].* 1) Laboratório de Citogenômica, Departamento de Patologia, Faculdade de Medicina da USP, São Paulo - SP, Brasil; 2) Unidade de Genética Médica, Departamento de Pediatria, Instituto da Criança-HC/FMUSP, São Paulo - SP, Brasil; 3) Departamento de Morfologia e Genética, Universidade Federal de São Paulo, São Paulo - SP, Brasil; 4) Laboratório de Genética Humana e Mutagêneses, Departamento de Biologia Geral, Universidade Federal da Bahia, Salvador - BA, Brasil.

The genomic imbalances are the most common cause of congenital malformations (CM) and developmental delay (DD), however the etiology of these imbalances are not well understood, making difficult the counseling genetics and the treatment. Currently, the improvement of cytogenomics diagnostic techniques, such as the screening by arrays that allow total coverage of the genome, is fundamental to set an unequivocal clinical diagnosis and a more suitable genotype-phenotype correlation for patients with CM/DD, especially when the clinical suspicion is not well understood, and may improve the evaluation of these patients more efficiently. We report our experience with the implementation of several array platforms (Agilent, Affymetrix, and Illumina) and probe densities in clinical diagnostic and scientific research of patients without genetic conclusive diagnostic. The results were confirmed by MLPA and/or FISH techniques. Thus this study evaluated the genome of 162 children with CM/DD. All patients were previously assessed by conventional cytogenetic analysis. We identified several different genomic alterations in 137/162 (~84. 6%) patients, including deletion, duplication and loss of heterozygosity. Some patients 41/137 (~30%) showed only one copy number variation in the genome, others 41/137 (~30%) presented two abnormalities and 55/137 (~40%) revealed three or more alterations on different chromosomes, maybe due to a complex rearrangements. Our findings showed that the array is a powerful tool to identification and characterization of genomic changes and provides precise diagnosis of unidentified or unexplained diseases suspected to have a genetic cause, contributing to appropriate clinical management of the patients. The interpretation of genotype-phenotype correlations in patients with complex genomic rearrangements is a very difficult task but the results can directly contributes to the elucidation of new syndromes.

2001F

Improved methods for diagnosing mitochondrial disorders using whole exome sequencing. *I. A. Barbosa[1], K. Thompson[2], C. L. Alston[2], E. W. Sommerville[2], C. Fratter[3], J. Poulton[4], R. McFarland[2], R. W. Taylor[2], C. Deshpande[5], M. A. Simpson[1].* 1) Medical and Molecular Genetics, King's College London, London, United Kingdom; 2) Wellcome Trust Centre for Mitochondrial Research, Newcastle University, Newcastle upon Tyne, United Kingdom; 3) Oxford Medical Genetics Laboratories, Oxford University Hospitals NHS Trust, Oxford, United Kingdom; 4) Nuffield Department of Obstetrics and Gynaecology, Women's Centre, University of Oxford, Oxford, United Kingdom; 5) Clinical Genetics Unit, Guys and St Thomas' NHS Foundation Trust, London, United Kingdom.

Mitochondrial diseases are a highly complex group of multisystem disorders characterised by biochemical defects in oxidative phosphorylation. Mutations in the nuclear genome have been identified as the major cause of inherited mitochondrial disorders presenting early in infancy or childhood. We employed whole exome sequencing (WES) for the identification of disease-causing variants in sixty-two individuals with clinical and biochemical presentations of mitochondrial disease in whom mitochondrial DNA rearrangements and point mutations had previously been excluded. We established a pipeline to assist the interpretation of the patients' exome variant profile in the context of nuclear genes previously associated with mitochondrial disease. In addition we evaluated the exome variant profile using publically available databases of genes with known or predicted mitochondrial function and of genes previously associated with related clinical phenotypes. We evaluated rare (MAF < 0. 01) and previously unobserved alleles under both dominant and recessive models in each affected individual. A multidisciplinary team reviewed the identified putative pathogenic variants for each individual, prior to validation of the variant calls with Sanger sequencing. Overall, this approach yielded a diagnostic rate of 42%. This metric reflects both individuals harbouring pathogenic mutations in nuclear-encoded mitochondrial disease genes routinely tested within UK laboratories (18%) and a further 24% with pathogenic variation in genes that are not currently tested as part of a diagnostic work-up. We have identified a number of highly plausible candidate causal variants located in genes with described mitochondrial function and experiments to define the functional consequences of these alleles on mitochondrial function are ongoing. Our results illustrate the efficacy of whole exome sequencing and an appropriate analytical strategy to provide molecular diagnostics in this complex clinical phenotype.

2002W

Novel Mitochondrial Variant Identified in Leber Hereditary Optic Neuropathy Patients. *S. Lee[1], M. Seong[1], S. Seo[1], Y. Yu[2], J. Hwang[3], S. Cho[1], J. Kim[4], S. Park[1,4].* 1) Department of Laboratory Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea; 2) Department of Ophthalmology, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, Korea; 3) Department of Ophthalmology, Seoul National University Bundang Hospital, Seongnam, Korea; 4) Biomedical Research Institute, Seoul National University Hospital, Seoul, Korea.

Leber's hereditary optic neuropathy (LHON) is characterized by acute or subacute bilateral visual loss, primarily caused by mitochondrial gene mutation. The majority of LHON are accounted for four common primary mutations. Here we investigated other genetic cause for LHON cases without four primary mutations by mitochondrial genome sequencing. Twelve LHON patients were included in the study. The sequence analysis of the mitochondrial genome was done using Roche GS sequencing technology and NextGENe software. We selected sequence variants at a frequency ofgreater than 15% in read frequency. A total of 23 missense variants were identified and we selected five likely pathogenic or pathogenic variants through in-silico prediction and Sanger confirmation. Two ND1 variants, m. 3635G>A and m. 3472T>C were previously reported as pathogenic one. We identified one novel pathogenic variant, m. 13259G>A (ND5) in a patient along with two novel likely pathogenic variants, m. 14180A>G (ND6) and m. 8520A>G (ATP8) in two patients, respectively. Multiple sequence alignment and hydrophobicity analysis supports of the pathogenicity of this novel variant. The full mitochondrial genome sequencing may be beneficial to LHON patients without primary mutation.

2003T

Identification of known and novel mutations in Mitochondrial Disorders. *J. Sheth[1], D. Solanki[2], N. Pandya[3], H. Patel[1], M. Mistry[1], R. Bhavsar[1], F. Sheth*. 1) Foundation For Research in Genetics and Endocrinology, Institute of Human Genetics Biochem & Molec Bio, Inst Human Gen, Ahmedabad, Gujarat, India Email:jshethad1@gmail. com; 2) Mantra clinic, Vivekanand Multispeciality Hospital,, Bhavnagar, Gujarat, India; 3) Shelby Hospital, Shelby Hospital, Department of Infertility and IVF, Ahmedabad, Gujarat, India.

Purpose of study: Identification of mitochondrial disorders are a multi organ involvement occuring due to mitochondrial and nuclear gene mutation**Methods**: Present study represents three families having a child with suspected mitochondrial disorder at the age of 6 m to 3 yr. Based on clinical presentation two children were investigated initially for SURF1 and POLG1 gene mutation by Sanger sequencing and one case was investigated by a panel of genes involved in mitochondrial encephalopathy by NGS. **Result and Discussion**: 1: A female child born to non consanguineous parents at full term and at 6 months proband had reduced movement with encephalopathy. On investigation, three common mutations of Leigh Disease (T12706C, A13084T & G13513A) were ruled out. Further analysis of SURF1 gene identified homozygous splice site mutation in intron 8 (IVS+(+1)G-A). Case 2:2 years male child was born to a consanguineous parents presented with refractory epilepsy, motor developmental delay and failure to thrive with right upper limb lower limb paresis. MRI, CSF lactate, urine GCMS and enzymes for GM1, GM2 gangliosidosis, NCL were normal. He was investigated for POLG1 gene mutation and homozygous deleterious mutation c. 911T-G (p. L304R) in exon 4 was detected. Proband was confirmed to have Alper's syndrome. Case 3: A 3 years old female child born to a consanguineous parents. Had an episode of convulsion at 3 months. Proband expired at 3 years after sudden vomitting episode. Aminoacid study showed increased levels of plasma lactate, alanine & tyrosine and was suspected to have died of mitochondrial disease. Mother was investigated for the mitochondrial encephalopathy gene panel using Illumina NGS platform. Heterozygous mutation c. 2095G-A (p. V699M) in exon 15 of *PC* gene and c. 911T-C (p. I304T) mutation in exon 4 of *FASTKD2* gene was detected in the mother. Further investigation demonstrated heterozygous mutation c. 2095G-T (p. V699L) in exon 14 of *PC* gene in the father with no mutation in FASTKD2 gene. Prenatal studies were done for case 1 and 3 based on the mutation associated with mitochondrial disorder. Present study demonstrates the identification of a rare gene mutation and usefulness of gene panel study by NGS for mitochondrial encephalopathy. It also demonstrate nuclear DNA mutation seems to be more common in mitochondrial disorders.

2004F

Paternity Reclaimed: Ancestry Testing Reveals a Unique Case of Congenital Chimerism. *K. M. Sheets[1], B. Kirkpatrick[2], M. Baird[3], J. Heinig[3], D. Davis[3], D. B. Starr[4].* 1) Vibrant Gene Consulting, LLC, Cambridge, MA; 2) Genomic Medicine Institute, Geisinger Health System, Harrisonburg, VA; 3) DNA Diagnostics Center, Inc. , Fairfield OH; 4) Department of Genetics, Stanford University School of Medicine, Stanford, CA, United States of America.

DNA testing is utilized when a biological relationship between individuals requires confirmation such as child custody, child support, or immigration cases. In the US over 382,000 legal relationship tests are ordered annually1, using PCR assays that examine short tandem repeat (STR) regions. Limitations of these assays have been noted1,2. We report a case of congenital chimerism revealed as a result of incompatible blood types between parents and child, negative paternity test results, and an avuncular relationship revealed through ancestry testing. Tetragametic chimerism is a condition that occurs when embryonic cells from dizygotic twins fuse early in development, leading to the birth of a singleton with some or all cells containing alleles from both dizygotic twins2. This condition was suspected after a pregnancy conceived with intrauterine insemination revealed an exclusion of paternity while at-home ancestry testing indicated an uncle/nephew relationship between the father and child. Confirmatory testing was recently performed by a laboratory accredited for legal relationship testing. Gametic tissue from the father was tested and revealed the presence of a mixed DNA profile consistent with a major and minor contributor. Chimerism was confirmed when it was found that one son matched one set of alleles and the other son matched the second set. The STR test results from several different tissues of the father are provided. To our knowledge, this is the first reported case in which paternity was initially excluded by standard DNA testing methods (buccal swab), the results called into question because of ancestry/relationship testing, and later included as the result of alternate genetic (semen) analyses. Although this case is rare, with the increased use of assisted reproductive technology, this outcome could occur with increasing frequency2. Sources:(1) Annual Report Summary For Testing in 2010. AABB Annual Report, pp. 1-9. (2) Boklage, C. (2005). Embryogenesis of chimeras, twins and anterior midline asymmetries. Human Reproduction, 579-591.

2005W

Improved diagnostic yield of neuromuscular disorders applying whole exome sequencing for patients arising from consanguineous population. *Z. Fattahi[1,2], M. Fadaee[1,2], R. Vazehan[2], E. Parsimehr[2], A. Abolhassani[2], SH. Nafissi[3], MR. Akbari[1,4], A. Kariminejad[2], H. Najmabadi[1,2].* 1) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran; 2) Kariminejad-Najmabadi pathology and Genetics Center, Tehran, Iran; 3) Department of Neurology, Tehran University of Medical Sciences, Tehran, Iran; 4) Women's College Research Institute, Women's College Hospital, Toronto, Canada.

Neuromuscular Diseases (NMDs) include a broad range of disorders affecting muscles, nerves and neuromuscular junctions. Approximately 761 different disorders fit in this group which is categorized into 16 different subgroups and 406 known genes. Diagnosis of these disorders is divided into three phases of clinic, pathology and molecular diagnostic laboratory. Today, molecular genetic studies are being considered as a strong laboratory tool in the neuromuscular disorders' workup. In general, overlapping phenotypes and heterogeneous genetic nature of NMDs have created diagnostic challenges in the field. So, there are still lots of undiagnosed patients despite the supportive clinical and pathological documents. Non-specific clinical features, large number and size of known genes, high number of unidentified genes, and multiple disorders associated with a same gene calls the implementation of next-generation sequencing as a new strategy to increase the diagnostic yield. In this study, whole exome sequencing was applied as a fast and accurate technique to investigate genetic causes of referred patients with inherited neuromuscular disorders to our laboratory. These patients were mostly offspring of consanguineous marriages, diagnosed with a kind of muscular dystrophy. During a year, total of 28 patients were referred and examined with the help of whole exome sequencing. Data analysis including alignment, variant calling and filtering redundant variants was performed to achieve a final list of the most probable pathogenic variants in known genes. Variant confirmation and co-segregation analysis as well as genotype-phenotype correlation was performed for detection of the causative variant. Eventually, in 20 out of 28 (71. 4%) patients with muscular symptoms, the pathogenic variant was identified in the following known genes; DMD, CAPN3, FHL1, RYR1, SYNE1, PLEC, TNNT1, SGCB, LMNA and twelve novel pathogenic variants were detected. Today, the advantage of whole exome sequencing in clinical diagnostic strategies of heterogeneous disorders is clarified. In our cohort of patients, the diagnostic yield of 71. 4% was achieved which is quite high compared to the overall diagnostic yield of 25 to 50% in other reports. This could be explained by the consanguineous background of these patients and is another strong indication for effectiveness of offering whole exome sequencing panels in diagnostic laboratories especially in populations with high rate of consanguinity.

2006T

Pitfalls of multiple ligation-dependent probe amplification in *DMD* gene mutation analysis. *M. Kim[1], S. Cho[1], J. Chae[2], B. Lim[2], J. Lee[1], S. Lee[1], S. Seo[1], H. Park[1], A. Cho[3], S. Kim[4], S. Park[1], M. Seong[1].* 1) Departments of Laboratory Medicine, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, South Korea; 2) Departments of Pediatrics, Seoul National University Hospital, Seoul National University College of Medicine, Seoul, South Korea; 3) Department of Pediatrics, Ewha Womans University School of Medicine, Seoul, South Korea; 4) Department of Laboratory Medicine, National Medical Center, Seoul, South Korea.

Background: Multiple ligation-dependent probe amplification (MLPA) is important technology in molecular diagnosis of Duchenne/Becker muscular dystrophy (DMD/BMD. However, little about false positive rate of MLPA in this disease is known. Here, we retrospectively reviewed results of 591 alleged DMD/BMD patients who MLPA analysis had been performed in our laboratory between May 2006 and Dec 2012. **Methods:** Mutational analysis was performed as the following procedures: first, whole coding region was screened using MLPA; second, if normal on MLPA, subsequent direct sequencing was performed to search sequence variations for the clinically suspicious patients. In addition, direct sequencing was also done for single exon deletion or duplication, or indeterminate results on MLPA. **Results:** A total of 290 cases showed exon deletion or duplication results on MLPA. Among them, 75 cases (25. 9%) initially revealed deletion or duplication for single exon. Subsequent direct sequencing revealed that 11 single exon deletion cases (3. 8%) were false positive due to sequence variations involving probe hybridization or ligation site. Abnormal MLPA results were closely related with sequence change type and location within probe hybridization. The C to T change was most frequently observed single nucleotide change (55. 9%, N = 19). CA mismatch caused by C to T or A to G substitutions for forward probes binding sites or T to C, G to A nucleotide changes for reverse probes by C to T substitution and having lower melting point (Tm) than 75oC were significantly associated with abnormal MLPA results. **Conclusions:** Our study reveals that about one seventh of large gene rearrangements involving single exon in *DMD* might be false positive results by sequence variations within probe hybridization or ligation site. Therefore, single exon deletion or duplication has to be validated by other methods, and furthermore, careful design of MLPA probes is needed to avoid false positive results.

2007F

Informative results of exome sequencing in 10 out of 20 cases/sibships. *M. Smith[1], K. E. Singh[1, 2], N. M. Gallant[1,2], E. C. Chao[1], V. E. Kimonis[1].* 1) Dept Pediatrics, Univ California, Irvine, CA; 2) Memorial Care, Miller Children's Hospital, Long Beach CA.

We carried out clinical exome sequencing on 20 patients. In 10 cases/sibships pathogenic mutations were identified in genes that are known to impact phenotype and clinical findings in our patients were consistent with the clinical features described in the literature in cases with defects in those genes. A pathogenic mutation in the X-linked gene *SLC16A2* occurred in a male patient with developmental delay, hypotonia and delayed myelination. A pathogenic splice site mutation in the X-linked *CUL4B* gene occurred in a male patient with megalencephaly, developmental delay, and lack of speech development by 8 years. This mutation led to exome skipping and to intron inclusion. Two pathogenic mutations (phase unknown) were identified in the *POLR3B* gene in a patient with developmental delay and seizures. This gene encodes the second largest subunit of RNA polymerase III and mutations have previously been reported in cases with leukodystrophy. A pathogenic mutation in the *ARID1B* gene was found in a female patient with developmental delay, dysmorphology and congenital malformations; heterozygous deletions of *ARID1B* have been previously described in multiple individuals with developmental delay. Compound heterozygous mutations in the *NUBPL* gene were found in siblings with developmental delay, ataxia and seizures and cerebellar hypoplasia. An 18y old with hypovolemia, osteoporosis and spinal defects was found to be homozygous for a pathogenic mutation in the Barrter syndrome gene *SLC12A1*. Pathogenic homozygous splice-site mutation in the *SCARB2* gene was found in sibs with myoclonic epilepsy; *SCARB2* mutations are reported in action myoclonus and renal insufficiency. In a female with seizures developmental regression and deafness a pathogenic de novo mutation was found in the X-linked *MED12* gene and a known pathogenic mutation also occurred in *TCTN1*. *MED12* mutations have been found mainly in males with developmental delay however there are reports of females with intellectual disability and MED12 mutations. Heterozygous *TCTN1* gene mutations have previously been described in Joubert syndrome. Compound heterozygous mutations in *NGLY1* (N-glycanase 1) were found in siblings. We identified a *CACNA1S* mutation in a case with myopathy similar to a literature reported case. We continue to find exome sequencing to have a significant diagnostic yield for rare disorders leading in some individuals to specific therapies. Exome sequencing may be used as first or second line testing.

2008W

The Diagnostic Yield of a Multi-Gene Panel for Neuromuscular Disorders with Special Consideration to Variants of Unknown Significance. *A. S. Lindy, M. Bradbury, S. F. Suchy.* Neurogenetics, GeneDx, Gaithersburg, MD.

Neuromuscular disorders (NMD) encompass over 200 clinically and genetically heterogeneous disorders that primarily affect the peripheral nervous and musculoskeletal systems. Overlapping features can make a clinical diagnosis difficult, allowing potential health risks associated with the disorder to go unmonitored. Multi-gene panel testing is one tool available for providing an efficient genetic diagnosis. We evaluated the performance of an NMD panel of 73-76 genes. Testing was performed on 150 individuals with NMD using next generation sequencing (NGS) along with exon-level copy number analysis. A definitive positive result was identified in 33% of cases and defined as one or two pathogenic or likely pathogenic variants in a single gene, depending on inheritance. Negative results were reported in 18% of cases and inconclusive results were reported in 49% of cases, due to the identification of a variant of uncertain significance (VUS). Given the relatively recent association of some of these genes with NMD, and the lack of well-established criteria for variant classification within those genes, the likelihood of identifying a VUS is high. Although 33% of cases had no VUS, 52% of cases had 1-3 VUS. Of those cases with inconclusive results, 41% had a VUS in an autosomal dominant (AD) or X-linked (XL) disorder, while 15% had a single pathogenic variant in an autosomal recessive (AR) disorder, potentially representing carriers of AR conditions. Inconclusive results are difficult to relay to patients and create challenges in determining how to proceed with clinical management. However, interpretation of a VUS in the context of clinical information can guide additional testing that can further clarify the molecular results. In many cases, reclassification of the variant was possible based on segregation analysis (60 variants), biochemical studies (2 variants), or immunostaining of a muscle biopsy (26 variants). Muscle biopsies can be especially useful in the reclassification of VUS in disorders associated with both AR and AD inheritance (i. e. COL6A related disorders). Our results demonstrate the effectiveness of NMD multi-gene panels. Using these panels enables not only the identification of common causes of NMD, but also novel genetic variants in rare disorders that might otherwise be overlooked. Furthermore, our results indicated that even the finding of a VUS can lead to a positive diagnosis when molecular testing is combined with targeted follow-up testing.

2009T

Allele-specific digital PCR to differentiate CYP2D6 heterozygous duplication events. *P. A. Hegerich[1], S. Patel[1], J. Forcellini[2], M. P. Borgman[2], T. Hartshorne[1].* 1) Genetic Analysis, Thermo Fisher Scientific, South San Francisco, CA; 2) PGXL Laboratories, Louisville, KY.

The drug metabolizer phenotype of the major P450 enzyme, CYP2D6, can be predicted by genetic analysis. The combination of full, reduced, or no function alleles, in addition to gene copy number, is used to determine ultra-rapid, extensive, intermediate, or poor metabolizer status. Data from quantitative PCR (qPCR) experiments using TaqMan™ SNP Genotyping assays and TaqMan Copy Number assays can be translated to star allele diplotypes associated with metabolizer phenotypes. While qPCR can decipher most of the information necessary, phenotypes cannot be unequivocally assigned for samples that contain a CYP2D6 duplication or multiduplications and are heterozygous for alleles that are known to be duplicated and that have different functional levels. For these samples, allele-specific copy number analysis by digital PCR (dPCR) can be done in order to identify the duplicated allele. Digital PCR enables single sample analysis with high precision and sensitivity by partitioning target molecules into 20,000 individual reactions, thus elucidating the ratio of the heterozygous alleles. For the allele-specific copy number variation (ASCNV) application, TaqMan SNP assays to CYP2D6 variants that are associated with specific duplicated alleles were run in dPCR on samples of known SNP genotype and CNV status. Initial validation was conducted on Coriell cell line gDNA samples. Here we will describe work done on buccal swab samples collected from two different sites. Representative samples carrying CYP2D6 duplications that were heterozygous were selected to run with ASCNV dPCR. Samples were first digested with a restriction enzyme to separate tandem duplicated CYP2D6 alleles then run on the QuantStudio™ 3D Digital PCR system. Reactions positive for each allele, detected by allele-specific VIC™ or FAM™ dye-labeled probes, were visualized, and allele ratios determined using the QuantStudio 3D AnalysisSuite™ software on the Thermo Fisher Cloud. For samples heterozygous for target SNPs, 3-copy samples gave close to 1:2 ratios and the duplicated allele was readily identified. Interestingly, one *2/*4 sample with CN=4 and initial call of EM/IM returned a result of *2/*2/*2/*4 and thus a true phenotype of UM. We have thus shown that ASCNV dPCR and TaqMan SNP assays is a simple and effective method for identifying specific duplicated alleles in heterozygous samples. This method facilitates accurate CYP2D6 allele genotyping and better prediction of drug metabolizer phenotype.

2010F

Resolving false positive *CYP2D6* genotype results: *CYP2D7* variation is the culprit. *A. K. Riffel[1], M. Dehghani[2,3], T. Hartshorne[4], J. S. Leeder[1,5], K. Rosenblatt[2,3], A. Gaedigk[1,5].* 1) Division of Clinical Pharmacology, Toxicology & Therapeutic Innovation, Children's Mercy Kansas City, Kansas City, MO; 2) CompanionDx Reference Laboratory, Houston, TX; 3) Devision of Oncology, Department of Internal Medicine, University of Texas Health Science Center at Houston, TX; 4) Genetic Analysis, Thermo Fisher Scientific, South San Francisco, CA; 5) School of Medicine, University of Missouri-Kansas City, Kansas City.

Background: TaqMan™ genotyping assays are widely used to genotype *CYP2D6,* which encodes a major drug metabolizing enzyme to guide therapy. Assay design for *CYP2D6*, however, is arduous due to the presence of 2 pseudogenes, structural and copy number variation and numerous SNVs, some reflecting the wild-type sequence of the *CYP2D7* pseudogene. We have identified samples testing false-positive for the SNVs defining the *CYP2D6*15* (137Tins; frameshift) and **35* (31G>A, rs769258) allelic variants. The aim of this study was to identify the culprit causing false positives and remediate those by redesigning and validating alternative TaqMan genotype assays. **Methods:** DNA from 13,866 samples were genotyped by the CompanionDx® lab using TaqMan genotyping assays. Twelve samples had true positive calls and 54 samples had false positive calls for *CYP2D6*15* (confirmed by resequencing or the Luminex xTAG® CYP2D6 test). Nineteen samples were selected for follow-up. Long-range PCR was utilized to amplify the entire *CYP2D6* and *CYP2D7* genes for subsequent genotyping and/or Sanger sequencing. A high resolution melt assay was developed for a *CYP2D7* SNV. Alternate TaqMan assays were evaluated using *CYP2D6* and *CYP2D7*-gene specific XL-PCR and gDNA as templates. **Results:** The selected 19 samples were positive for *CYP2D6*15* when genotyped by CompanionDx. However, the 137Tins could not be confirmed for the majority of samples by Sanger resequencing. Genotype analysis of *CYP2D6* and *CYP2D7*-specific XL-PCR templates suggested that the false-positive mutation signal originated from *CYP2D7*. Sequencing of *CYP2D7* revealed a CC>TG SNV in exon 1 that reverts the sequence to *CYP2D6* and allows the TaqMan assay PCR primer to bind. Because *CYP2D7* also carries the Tins, a mutation signal is generated. This *CYP2D7* SNV was also responsible for generating a false-positive signal for the close-by *CYP2D6*35* rs769258 SNP. Alternate assays have been designed and evaluated. **Discussion:** Although TaqMan assays are carefully designed and evaluated before commercially marketed, rare or unknown SNVs underneath primer and/or probe regions can impact the performance of many PCR-based genotype assays, including TaqMan. In the presented case, a rare SNV in a highly similar pseudogene allowed primer binding to the variant site and thereby caused false-positive allele calls. Regardless of the test platform used, it is prudent to confirm rare allele calls by an independent method.

2011W

Characterizing CGG footprint at FMR1 gene by TP-PCR: Implication in diagnosis. *S. Agarwal, S. Muthuswamy, D. D. Dean.* Dept Genetics, Sanjay Gandhi Inst Med Sci, Lucknow, Uttar Pradesh, India.

Poor knowledge about role of CGG repeat at FMR1and lack of availability of proper and cheap molecular diagnostics may be a major barrier to early diagnosis that could improve quality of life and prognosis especially in the developing countries. As the treatment for Fragile X syndrome is not feasible, prevention of FXS by prenatal diagnosis of carrier women early during pregnancy is important. Therefore through present study, we used our in-house standardized TP-PCR method for screening CGG repeat expansion. The results were also reconfirmed with methylation sensitive PCR. We screened a total of 66 samples including 45 suspected FXS male children, 10 premature ovarian failure cases and 2 prenatal samples. TP-PCR method identified 5 (11%) out of 45 suspected FXS cases (intellectually disabled) to be positive for full mutation, 1 (10%) of 10 premature ovarian failure cases as permutation positive and 1 of 2 prenatal sample as permutation positive. Present study demonstrates that a significant number of FXS among intellectually disabled and identification of permutation in premature ovarian failure cases reinforces the need for family screening in early intervention and prenatal option. In prenatal cases it helped the parents to take decision for continuing pregnancy.

2012T

Increased identification of *CFTR* mutations using an expanded panel of validated pathogenic mutations. *W. Sun[1], D. Rabin[2], D. Goos-Root[1], T. Angeloni[1], S. Keiles[2], J. Radcliff[2], C. Strom[1].* 1) Molecular Genetics, Quest Diagnostics Nichols Institute, San Juan Capistrano, CA; 2) Quest Diagnostics, Madison, NJ.

Cystic fibrosis (CF) is caused by mutations in the *CFTR* gene. More than 1,900 CF mutations have been identified, most at frequencies <0. 1%. ACMG/ACOG guidelines recommend screening for 23 common *CFTR* mutations, which together account for the majority of mutant alleles in Ashkenazi Jewish and non-Hispanic Caucasian populations. Adding validated mutations may enhance detection. One obstacle to expanding the panel is the expense and time needed for sequence analysis of large numbers of mutations. Next-generation sequencing (NGS) allows rapid analysis of large numbers of mutations simultaneously, greatly expanding options for *CFTR* mutation screening. In 2014, we introduced an NGS-based expanded CF panel that detects up to 155 *CFTR* mutations. All additional mutations were deemed pathogenic: selection criteria for these mutations were a published association with a clinical diagnosis of CF; in vitro demonstration of disruption of the CFTR pathway or demonstration of a stop codon in the pathway; and an allele frequency ≥0. 01% in North America and Europe. Here we report the frequency of pathogenic mutations in the first series of clinical samples tested with this panel. Pathogenic *CFTR* mutations were detected in total of 340 samples. Of these, 61 (18%) were not part of the 23-gene ACMG/ACOG panel. Overall, 29 of the non-ACMG/ACOG panel mutations were detected in at least 1 sample. The most common was 3876delA (n=10), which was the 4th most frequent mutation overall (after deltaF508, R117H, and G542X). The next-most common non-ACMG/ACOG panel mutations were F311del, L206W, and R117C (detected in 4 samples each); and A559T, R352Q, and Y1092X (detected in 3 samples each). These mutations were detected at higher frequencies than more than half of the mutations in the ACMG/ACOG panel (data not shown). Of the remaining non-ACMG/ACOG panel mutations, 14 were detected in 1 sample each and 8 were detected in 2 each. Our findings indicate that the expanded panel detects pathogenic *CFTR* mutations in a substantial proportion of samples that would have been missed by testing with the 23-mutation panel alone. This analysis underscores the importance of continued evolution of *CFTR* mutation screening.

2013F

Hemoglobin Depletion Increases Sensitivity of Next Generation Sequencing-based Transcriptome Profiling. *D. Munafo, B. Langhorst, C. Sumner, C. Chater, L. Mazzola, J. Bybee, D. Rivizzigno, S. Russello, F. Stewart, E. Dimalanta, T. Davis.* New England Biolabs, Inc. , Ipswich, MA.

Peripheral blood can reflect molecular profile changes occurring in other tissues during pathological events even before clinical symptoms have appeared. Blood is also more accessible and easily collected than other tissues making it ideally suited for biomarker discovery. Whole blood transcriptome analysis by next generation sequencing allows for accurate detection of known as well as unknown transcripts with low- and high-abundance and it is beginning to demonstrate utility in early clinical diagnostics. One limitation of the whole blood RNA sequencing is that, the high content of hemoglobin transcripts (constituting up to 75% of total transcripts population) can mask detection of lower abundant biological meaningful transcripts. In this work, we demonstrate that globin depletion significantly increases sensitivity of transcript detection without altering transcript expression profile of the non-globin transcripts. Here, we present a method to deplete for hemoglobin transcripts. We applied this method to deplete multiple blood-derived samples (peripheral blood, leukocytes, bone marrow and umbilical cord blood) from adult, fetal and embryonic hemoglobin transcripts. Using next generation RNA sequencing, we assess depletion efficiency and we compare library complexity and transcript expression correlation before and after globin depletion. Whole blood RNA sequencing from single male donor reported to have 50% of total sequencing reads being hemoglobin transcripts. At 50 million paired-end read coverage, 23% hg19 annotated genes were detected. On the contrary, after globin depletion, only 0. 1% of sequencing reads mapped to globin transcripts increasing sensitivity of transcript detection by 30% (allowing for reliable detection of 2,787 additional transcripts). We achieved high depletion efficiency (up to 99. 9% globin depletion) with minimal off target effects (high FPKM correlation between depleted and non-depleted libraries; R2= 0. 87486). Globin depletion by this method is not altering the transcript expression levels. Globin reduction increases the coverage of less abundant transcript. This method can potentially be expanded to deplete for other highly express transcripts with minimal biological relevance. Moreover, it is amenable to high-throughput sample preparation and robotic automation to easy implement in a clinical setting.

2014W

Association analysis of genetic variants with type 2 diabetes in a Mongolian population in China. *H. Bai[1], W. Huo[1], H. Liu[1], S. Suyalatu[1], N. Narisu[2], Q. Wu[1].* 1) Inner Mongolia University for the Nationalities, tongliao, China; 2) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA.

T2D is a complex disease hallmarked by insulin resistance and pancreatic beta-cell dysfunction. In China, 9. 7% and 15. 5% of the entire population suffer from T2D and prediabetes, respectively. There are 4 million Mongolians currently living in various regions of China . With the help of recent advances in genotyping and sequencing technology, large scale genome wide association studies (GWAS) have identified approximately 70 single nucleotide polymorphisms (SNPs) conferring susceptibility to type 2 diabetes (T2D). However, most of these loci have not been replicated in a diverse populations and much genetic heterogeneity has been observed across ethnic groups with different genetic backgrounds. We tested 28 SNPs previously found to be associated with T2D by GWAS in a Mongolian sample of northern China (497 diagnosed with T2D and 469 controls) for association with T2D and diabetes related quantitative traits. We replicated a previous T2D association of rs2237897 near KCNQ1(OR=1. 48; P=0. 001)in a different Mongolian sample and replicated T2D association of ten other SNPs, namely,rs7578326 (IRS1), rs1531343 (HMGA2), rs8042680 (PRC1), rs7578597 (THADA), rs1333051 (CDKN2), rs6723108 (TMEM163), rs163182 (KCNQ1), rs1387153 (MTNR1B), rs243021 (BCL11A), and rs10229583 (PAX4) in our sample. Further, wetested four diabetes related lipids traits, namely,total cholesterol (TC), high density lipoprotein cholesterol (HDL-C), low density lipoprotein cholesterol (LDL-C), triglyceride (TG), and showed that risk allele of the strongest T2D associated SNP in our sample, rs757832 (IRS1) to be associated with increased level of TG. We observed substantial difference of T2D risk allele frequency between the Mongolian sample and 1000G Caucasian sample. Further study of genetic architecture of these variants in susceptibility of T2D is needed to understand the role of these variants in heterogeneous populations. Keywords: SNPs, genetic association, Mongolian population, T2D. AcknowledgementsWe thank our study individuals for their generous participation in this study. This study was supported by the National Natural Science Foundation of China(81160101,81060098).

2015T

Identification and characterization of aberrant splicing variants using a generic approach. *A. J. Bergsma[1,2], M. Kroos[1], M. Hoogeveen-Westerveld[1], D. Halley[1], A. T. van der Ploeg[2], W. W. M. Pijnappel[1,2].* 1) Clinical Genetics, Erasmus Medical Center Rotterdam, Rotterdam, Zuid-Holland, Netherlands; 2) Pediatrics, Erasmus Medical Center Rotterdam, Rotterdam, Zuid-Holland, Netherlands.

Diagnosis of monogenic diseases often involves sequence analysis at the level of genomic DNA in which only the protein-coding exons are analyzed. This may result in failure to identify variants affecting pre-mRNA expression and splicing. Here, we present a generic assay that can be used to identify such variants without prior knowledge of the function of the gene product. Al coding exons are analyzed in an unbiased manner using a combination of flanking exon PCR, exon-internal quantitative PCR and sequence analysis. The approach was tested on Pompe disease, a devastating skeletal muscle disorder resulting from variants in the acid alpha glucosidase (GAA) gene. Patients with partial or uncharacterized GAA variants revealed several splicing defects including exon skipping, intron retention, cryptic splice site usage, and mRNA degradation. These events could not be predicted using splice prediction programs. mRNA degradation analysis could be used to determine Cross Reactive Immunological Material (CRIM) status with prognostic value for disease severity and response to Enzyme Replacement Therapy (ERT). The extent of leaky wild type splicing correlated with disease severity. We conclude that this approach is valuable for unbiased identification of pathogenic variants and that the quantitative information obtained can be used for clinical information.

2016F

A comprehensive analysis of ultrastructures of erythrocytes from patients with Glucose-6-Phosphate Dehydrogenase deficiency by atomic force microscopy. *F. Zishui, J. Chengrui, T. Jia, J. Weiying(Corresponding author).* Department of Medical Genetics, Zhongshan School of Medicine, Sun Yat-sen Univer, canton, China.

Acute hemolytic anemia will be triggered by oxidative stress in the patients with Glucose-6-phosphate dehydrogenase (G6PD) deficiency. However, the underlying hemolytic mechanism is unknown. To make clear the hemolysis mechanisms, a systematic study on membrane ultrastructures had been undertaken by atomic force microscopy (AFM) in G6PD deficient red blood cells (RBCs) after tert-butyl-hydroperoxide (t-BHP) oxidation. The RBCs architecture seem to be impaired causing the cells to lose its ability to maintain the typical discoid shape. Erythrocyte shrinkage and the increased membrane roughness were observed in the G6PD deficient cells. Compared with healthy control RBCs, as the concentration of t-BHP up to 0. 1mM, the membrane roughness of G6PD deficient RBCs showed significant difference (P<0. 05). Oxidative damage results in increased roughness of erythrocyte membrane, which may shorten the RBCs lifespan. The results in the present study will give an increased understanding for the hemolysis mechanism of G6PD deficiency.

2017W

Comprehensive Glycomics Analysis by MALDI TOF/MS in Human CSF. *X. Li[2], M. Davids[3,4], M. Kane[3,4], L. Wolfe[3,4], M. Raihan[2], C. Boerkoel[5], W. Gahl[3,4], M. He[1,2].* 1) Department of Pathology and Laboratory of Medicine, University of Pennsylvania, Philadelphia, PA; 2) The Michael J Palmieri Metabolic Laboratory, Children's Hospital of Philadelphia, Philadelphia, PA; 3) NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, NIH, Bethesda, MD; 4) Office of the Clinical Director, NHGRI, National Institutes of Health, Bethesda, MD; 5) University of British Columbia, Vancouver, Canada.

The Congenital Disorders of Glycosylation (CDG) are a group of diseases with highly variable phenotypes and inconsistent clinical features. Over 100 CDG disorders have been identified and most of these are defects in protein glycosylation. Protein glycosylation is highly variable among different human tissues. For example, types of protein glycosylation of the human brain differ from those of peripheral organs and correspond to differing expression patterns for genes encoding glycosylation enzymes. To explore possible defects in protein glycosylation in the brain, we developed comprehensive glycomics analysis of N-glycome, O-glycome and free glycome by MALDI TOF/MS in cerebrospinal fluid (CSF). We used 0. 6 ml CSF that was frozen immediately after the collection for this analysis. The CSF was filtered through a size exclusion membrane to separate free oligosaccharides from glycoprotein. N-glycans were released from total CSF glycoprotein via PNGase F digestion and O-glycans were released separately by reductive beta-elimination reaction. After glycans were purified through SPE columns, they were permethylated and analyzed by MALDI TOF/MS. Thirty eight different N-glycan species, 85 free oligosaccharide and 25 O-glycan species were identified in 10 control CSF samples. The predicted structures for the majority of CSF glycans are found in online databases including UniCarbKB and CFG. Compared with plasma glycomics data, CSF hasmuch more O-glycan species including O-mannosylated glycans and polysialylated O-glycans. Comparing 166 CSF samples from patients with undiagnosed neurological disease from the NIH Undiagnosed Diseases Program (UDP) with 10 control CSF samples from different age groups that were collected by our clinical laboratory, about 20 UDP patients have CSF glycomics profiles significantly deviated from the profiles of control CSF samples. Four of the UDP patients have known genetic disorders that alter or are predicted to alter protein glycosylation in human brain, including one patient with putative CAD deficiency, a defect in uridine biosynthesis. Interestingly, one of the known patients has a completely normal plasma glycomics profile and urine oligosaccharide profile, while both his CSF N and O-glycomics profiles are abnormal. Our results suggest that CSF glycomics analysis could be a useful tool to discover new diseases or disease mechanism.

2018T

Detection of Large Rearrangements in a Pan-cancer Gene Panel using Next Generation Sequencing. *D. Mancini-DiNardo, T. Judkins, C. Daniels, J. Craft, J. Holladay, B. B. Roa.* Myriad Genetics, Inc. , Salt Lake City, UT.

Background: The advent of Next Generation Sequencing (NGS) has facilitated the detection of sequencing variants across multiple genes simultaneously. To date, however, large deletions and duplications are largely detected by alternative technologies such as MLPA and microarray CGH. Here we describe the detection of large rearrangements (LR) in a clinically validated pan cancer panel test using NGS. **Methods:** Dosage analysis was performed across gene coding exons and select promoters using NGS. All LRs were confirmed using one of three orthogonal approaches: targeted microarray CGH, MLPA or confirmatory PCR and sequencing of the mutant product. MLPA was initially used as the primary assay for dosage analysis in *PMS2* and *CHEK2*. *PMS2*-specific sequencing analysis and/or long range PCR were used where appropriate to confirm the presence of LRs in *PMS2*. **Results:** The proportion of clinically significant defects attributable to large genomic rearrangements across 25 genes was determined to be 9. 3% using the methods described above. Collectively, 86% of all LRs detected were deletions and 13% were duplications. We also detected five insertions, including the *BRCA2* Portuguese founder mutation (c. 156_157insAlu), and two triplications. Comprehensive coverage of exonic regions in this NGS assay also facilitated the detection of 9 partial deletions. Importantly, there were a large number of clinically significant LRs detected in *PMS2*, of which 36% occured in exons 11-15, a region that shares significant sequence homology to the pseudogene, PMS2CL. This finding stresses the importance of being able to accurately detect mutations in this region and distinguish between clinically significant LRs that occur in *PMS2*, and benign LRs in PMS2CL. **Conclusions:** NGS technology has been demonstrated as a reliable method to detect sequencing variants and has facilitated simultaneous, multi-gene analysis. The ability to interrogate for the presence of large deletions and duplications using the same technology has further streamlined the process of mutation detection, which increases the efficiency of testing and contributes to reduced turn-around-times. We have shown that dosage analysis by NGS can reliably detect the presence of genomic LRs in our pan cancer panel test. This study shows that LRs constitute a significant proportion of mutations found in individuals with a personal or family history of hereditary cancer, and should be part of a comprehensive genetic testing strategy.

2019F

Prognostic prediction of prostate cancer metastasis and biochemical recurrence. *A. Pearlman[1], K. Upadhyay[1], K. Cole[1], J. Loke[1], C. Campbell[1], S. Freedland[3], K. Sun[1], Y. Shao[2], H. Ostrer[1].* 1) Dept Pathology, Albert Einstein College of Medicine, Bronx, NY; 2) Division of Biostatistics, NYU School of Medicine, New York, NY; 3) Dept Urology, Cedars Sinai Medical Center, Los Angeles, CA.

Prostate cancer is the most commonly diagnosed male cancer and the second leading cause of cancer deaths among men in the United States. If left untreated, the majority of prostate cancers remain indolent for decades, but some progress to metastasis and death. If treated with surgery or radiation therapy, the risk of metastasis is reduced, but erectile dysfunction, urinary incontinence and rectal bleeding may occur in up to 50% of patients. The goal of this study was to validate a previously reported metastatic potential score (MPS) that could predict the likelihood of a prostate cancer to metastasize or recur biochemically (BCR) (Pearlman J Probab Stat 2012:873570). Using arrayCGH, genomic DNA copy number alterations were measured in radical prostatectomy specimens from 76 Duke University (DU - Affymetrix Oncoscan v3) or 104 Memorial Sloan Kettering (MSK - Agilent oligo array) men with prostate cancer. Multivariate logistic regression (logit) and Cox proportional hazards (Cox) models were used to assess the clinical validity of the MPS and other early pre-operative clinical predictors for outcomes of metastasis and BCR. A multivariate logit model comprised of the MPS and preoperative PSA levels resulted in a predictive accuracy of 80%, $p < 0.01$ for the outcome of metastasis in both DU and MSK cohorts. Accordingly, the outcome of metastasis-free survival in a multivariate Cox model of the MPS resulted in hazard ratios = 3. 7 and 7. 3, $p < 0.03$ along with a concordance index of 0. 70 and 0. 74 for the MSK and DU cohorts, respectively. Log rank test of the MPS plotted as Kaplan-Meier survival curves (cut at 50% of MPS range) resulted in a significant separation in both DU and MSK cohorts $p = 0.01$. Furthermore, for the outcome of BCR, the multivariate logit model of MPS/preoperative PSA resulted in a predictive accuracy of 0. 71, $p = 0.003$ (DU) and 0. 76, $p = 0.006$ (MSK). For the DU cohort MPS was the only predictor that achieved statistical significance when applied to a Cox model resulting in hazard ratio = 3. 6, $p = 0.002$ and concordance index of 0. 62, $p = 0.004$. In the MSK cohort the multivariate Cox model of MPS/preoperative PSA resulted in a hazard ratio of 5. 9, $p = 0.00008$ and concordance index = 0. 71, $p = 2.4E-07$. Thus, MPS is an accurate predictor of metastatic risk (and metastasis-free survival) for men with prostate cancer from different cohorts and tested on different arrayCGH platforms.

2020W

Identification of a novel homozygous COL18A1 mutation causing Knobloch syndrome using homozygosity mapping and whole exome sequencing. *A. Haghighi[1], A. Haghighi[2], A. Tiwari[3], N. Piri[4], G. Nurnberg[5], N. Saleh-Gohari[6], J. Neidhardt[7], P. Nurnberg[8], W. Berger[3].* 1) University of Toronto, Canada; 2) Department of Genetics, Harvard Medical School, Boston, MA; 3) Institute of Medical Molecular Genetics, University of Zurich Wagistrasse 12, CH-8952 Schlieren, Switzerland; 4) Kentucky Lions Eye Center, Department of Ophthalmology and Visual Sciences, University of Louisville, Louisville, KY; 5) Cologne Center for Genomics (CCG), University of Cologne, Weyertal 115b, 50931 Cologne, Germany; 6) Genetic Department, Kerman University of Medical Sciences, Kerman, Iran; 7) Toronto General Hospital, University of Toronto, Toronto, Canada; 8) University of Cologne, Joseph-Stelzmann-Str. 26, D-50931 Cologne, Germany.

In this study we investigated the genetic cause of a chorioretinal dystrophy with high myopia of unknown origin in a child. The proband was the result of a consanguineous marriage. Homozygosity mapping was performed using the Human SNP Array 6. 0. We also carried out whole exome sequencing (WES) using the patient's DNA. Results of homozygosity mapping and WES revealed a novel homozygous mutation in the *COL18A1* gene in the patient. This mutation causes a frameshift and premature stop codon in the last exon. The parents were both heterozygous for this variant. Mutations of *COL18A1* can cause Knobloch syndrome (KS). Retrospective analysis of clinical records of the patient revealed surgical removal of a meningocele present at birth. Our patient had all the main features of KS and a rare manifestation, chorioretinal degeneration. This is a first case of KS from Iranian population. Since this mutation is located in endostatin domain, a functional relevance of endostatin in KS is suggested. This study shows that linkage analysis in combination with whole exome sequencing is a proper approach for identifying the disease-causing mutations in conditions with unclear initial clinical diagnosis.

2021T

Screening for mutations and variants in FSHR gene in women with Premature ovarian failure in a South Indian cohort. *M. Sujatha, V. Madan Mohan, G. Jay Prakash, Ch. Sravanthi, A. Jyothy.* Institute of Genetics and Hospital for Genetic Diseases,Osmania Univeirsity, Begumpet,Hyderabad,Telangana State, India.

Purpose of the study : Premature ovarian failure (POF, MIM311360) is defined as a cessation of menses (amenorrhea) accompanied by elevated gonadotrophin levels before the age of 40 years. POF-I is considered as primary amenorrhoea(PA), POF-II is considered as secondary amenorrhoea (SA), and POF-III (POF,MIM311360). POF is an early ovarian malfunction with decreased production of follicles. Normal functioning of the FSHR receptor is crucial for follicular development. The Exon 1 encodes the N-termina part of the extracellular domain and exon 10 is large and encodes the C terminal part of the transmembrane domain and intracellular domain. Allelic variants (p. Thr307Ala and p. Asn680Ser) and mutations in the FSHR gene have been described as possible causes of premature ovarian failure. These findings led us to investigate the presence of mutations in all the exons and allelic variants in FSHR gene. Materials and Methods : A total of 100 patients and 50 controls from Institute of Genetics and Hospital for Genetic Diseases were considered for the present study . Clinical histories and endocrine profiles were obtained. FSH receptor gene exons 1,2,3,4,5,6,8,9,10A,10B were screened for unknown mutation by PCR, SSCP, DHLPC and DNA sequencing methods. Exons 8,10C,10A,Exon E-G for known mutations were screened by PCR, RFLP and DNA sequencing. Allelic Variants by PCR ,RFLP and DNA sequencing . Results :No mutations were detected in exons 1,2,3,4,5,6,8,9,10A, 10C, 10 D, 10E-G. But in exon 10B a novel mutation T1142G in a herozygote state was identified. The frequency of NN, NS and SS genotypes were 20. 8, 62. 5, and 16. 6% for POF-I and 26. 3, 42. 1, and 31. 5% for POF-II and 28. 57, 52. 38, and 19. 04% for POF-III. The control group showed 34. 2% (NN), 44. 7% (NS), and 21. 05% (SS). The frequency distribution of allelic variants TN, AS and, TS and AN were 48. 3, 41. 6, 6. 6, and 3. 3% for POF-I; 47. 5, 50, 2. 5, and 0% for POF-II; 39, 27, 17, and 17% for POF-III, and control group showed 54, 34, 6, and 6% respectively. The FSH levels of Ala307Ala/ Ser680Ser (SS) genotype for POF-I, POF-II, POF-III were 0. 92±0. 44, 10. 65±1. 34 and 150±28. 91 ($p<0.0001$) respectively. Conclusions:We conclude that mutations in all the exons of FSHR gene are rare in our population and allelic variant Ala307Ala/ Ser680Ser (SS) genotype is associated with clinical and endocrine profiles in patients with premature ovarian failure in South Indian population.

2022F

Development and clinical validation of a novel diagnostic test for Mucin1 kidney disease. *B. Blumenstiel[1], M. DeFelice[1], O. Birsoy[1,2], Q. Yu[1], Z. Leber[1], N. Lennon[1].* 1) Broad Institute of Harvard and MIT, Cambridge, MA; 2) Partners Healthcare, Laboratory for Molecular Medicine, Boston, MA.

Mucin -1 Kidney Disease (MKD), also known as medullary cystic kidney disease type 1, is a rare autosomal dominant inherited kidney disease which leads to progressive loss of kidney function and end stage renal disease. Progression of the disease most often requires dialysis or kidney transplantation in the later decades of life. Previous reporting has linked this progressive tubulo-interstitial disease to a difficult to detect mutation markedly underrepresented in massively parallel sequencing. The mutation is identified as a single base insertion in a single copy of a 60bp repeat unit comprising a long (~1. 5-5 kb), GC-rich (~85%) VNTR sequence in a coding region of the MUC1 gene. This single base cytosine duplication extends a 7C homopolymer to an 8C homopolymer, creating a frameshift which introduces an early stop codon and is expected to result in a truncated protein (Kirby *et al*. 2013). The inherent difficulty in detecting this specific mutation using standard commercial methods required the development of a novel assay for MKD diagnosis in a clinical setting. Here we describe the development, clinical validation and implementation of a mass-spectrometry based CLIA assay for accurately diagnosing patients with MKD. Validation of the MKD diagnostic assay tested across 48 samples in six replicate runs demonstrated a conclusive call rate of 95. 5% and accuracy of 100%. .

2023W

NGS-based Carrier Screen for Gaucher's Disease Calls Variants and Detects Large Rearrangements Between *GBA* and *GBAP1*. *K. M. D'Auria, M. R. Theilmann, K. Iori, C. S. Chu, I. S. Haque, E. A. Evans, H. P. Kang, J. R. Maguire, D. Muzzey.* Counsyl, South San Francisco, CA.

INTRODUCTION: With a carrier rate of 1 in 18 in the Ashkenazi Jewish population, Gaucher's Disease is included on most carrier-screening panels, yet 95% homology between the disease-associated gene (*GBA*) and its pseudogene (*GBAP1*) complicates genotyping by NGS alone. Variant L444P, which is present in 30% of the affected population, is particularly difficult to assess because the mutant base in *GBA* is the normal base in the pseudogene. As such, low-throughput and costly assays—including allele-specific qPCR or long-range PCR followed by Sanger sequencing—have been used to resolve this challenging variant and others, such as D409H, that are also pseudogene-derived. Here we present an entirely NGS-based method that detects ten of the most common deleterious *GBA* variants, giving a 97% carrier detection rate.

METHODS: Hybrid-capture probes were designed to enrich for DNA fragments spanning the variants of interest. To ensure that only *GBA*-derived reads contributed to variant calling in highly homologous regions, reads were discarded if they did not contain a sufficient number of *GBA*-derived bases for unambiguous mapping by the alignment software. Rearrangements in the locus were detected by combining a depth-based copy-number analysis and an assessment of allele balance at each variant position. NGS results for variants L444P and D409H were validated using a combination of long-range PCR and allele-specific TaqMan PCR on 60,000 samples.

RESULTS: Relative to the PCR assay, our NGS test correctly genotyped the L444P and D409H variants in all 60,000 samples. Remarkably, 26% of L444P carriers resulted from fusion genes and would have been false negatives using either a standard SNP caller on NGS data or a qPCR assay with *GBA*-derived primer sequences, underscoring the importance of our custom rearrangement-detection algorithm. Our assay additionally greatly improved the specificity of calling L444P, identifying 84% of heterozygous calls from a standard SNP caller as false positives. The allele frequencies of all other variants are in line with previous studies and sequencing projects.

CONCLUSIONS: NGS sequencing alone, with proper probe design and analysis, can accurately genotype all common *GBA* variants, even those that are pseudogene-derived and nested in highly homologous regions.

2024T

Rapid Capture Methods for Comprehensive Carrier Screening. *D. Muzny[1], J. Hu[1], C. Buhay[1], F. Xia[2], H. Doddapaneni[1], H. Cui[2], Y. Ding[1,3], A. Pourpak[3], M. Wang[1], E. Boerwinkle[1,4], P. Fang[2], N. Verraraghavan[1], Y. Yang[2], C. Eng[2], R. Gibbs[1], A. Beaudet[2].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX,; 3) Baylor Miraca Genetics Laboratories, 2450 Holcombe Blvd, Houston, TX; 4) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX.

Advances in sequencing technologies and resulting improvements in gene discovery for Mendelian diseases now make accurate and rapid large scale carrier testing for recessive disease an imminent reality. We developed a novel 500kb carrier mutation gene capture panel that includes 168 complete genes that contain at least 850 known common genetic variants of clinical relevance. This capture based panel utilizes a '*lightning capture'* process to deliver variant calls in 5-7 days after sample intake. The '*lighting capture*' process includes: quick enrichment library preparation, (5-6 hours) and capture enrichment (approximately 8 hours); rapid sequencing (Illumina HiSeq2500) and data analysis via the HGSC-developed Mercury pipeline. We also employ the SNPTraceTm assay by Fluidigm in order to ensure reliable sample identification and to test for sample cross-contamination. These same loci form the SNPTrace test have been incorporated into the carrier capture design (500kb). Carrier testing for SMA, fragile X, deletions and duplications in DMD and recurrent clinically relevant CNVs in other genes is currently performed separately, using the Fluidigm platform. This method has been validated utilizing >1,100 control samples. Current methods employ a cost-effective 47plex co-capture format for hybridization with the regional capture reagent followed by sequencing of 94 samples (2 capture pools) per HiSeq 2500 lane for data generation. High enrichment efficiency was observed (72% reads on target and buffer) and superior coverage metrics (99. 1% of target bases ≥20x coverage, and 326x median coverage) across the design with ~400 Mb sequencing yield. A detailed analysis of the design performance using 140 de-identified samples found that known carrier mutations were correctly identified with high confidence (98. 5%), including large/complex indel mutations. Access to universal carrier testing for hundreds of medically relevant (severe) autosomal recessive and X-linked disorders with low cost and quick TAT will be highly impactful, particularly with the addition of conditions such as DMD and deletions and duplications of MECP2. Pre-conception or early post-conception testing of partners is the optimal form of screening to aid prediction of a couple's risk of offspring with an inherited genetic disorder.

2025F

Comparison between HPV Oncotect and Nuclisens EasyQ assay and its potential role in detecting preneoplastic lesions of the cervix. *F. Papa, C. Vaccarella, M. B. Majolini, M. Belli, A. Luciano, V. Mazzucchi, M. CA. Rongioletti.* Clinical Pathology Department, San Giovanni Calibita Hospital, Rome, Italy.

Objectives: This study compare the performance of NucliSENS EasyQ test (bioMerieux), a RT-PCR based method, or HPV OncoTect (InCellDx) a flow cytometry-FISH based method, in the detection of the E6/E7 mRNA expression of hrHPV**.** Moreover, we investigated the potential role, for the detection of high grade lesions, of the HPV OncoTect compared to pap smears to improve cervical cancer screening. **Methods:** We enrolled 173 patients positive for HR-HPV DNA and/or pap smear and referred for evaluation by means of the NucliSENS EasyQ. All patients were tested with HPV OncoTect and 87/173 underwent a colposcopy and histological evaluation according to clinical protocol. **Results:** HPV OncoTect displayed a reactivity of 25% in negative samples, 40% in atypical squamous cells of undetermined significance ASCUS, 48% in low-grade squamous intraepithelial lesion LSIL and 80% in high-grade squamous intraepithelial lesion HSIL. The corresponding results for NucliSENS EasyQ assay were 42, 60, 74, 90%, respectively. Histology analysis revealed that the sensitivity of HPV OncoTect compared to NucliSENS EasyQ was 20% versus 54% in normal specimens, 45% versus 69% in CIN1 and 87% versus 74% in CIN2+, respectively. Higher specificity was observed analysing CIN2+ samples with HPV OncoTect (69%) versus NucliSENS EasyQ (36%). **Conclusions:** HPV E6/E7 mRNA Oncotect test is more specific than NucliSENS EasyQ in identifying women with CIN2+ but has a lower sensitivity. HPV OncoTect is a promising tool for early prediction of persistent HPV infection and seems to be an interesting method to evaluate the preneoplastic lesions of the cervix and improve cancer screening.

2026W

Dried blood spot RNA sequencing (DBS-RNA-Seq): A novel approach for the identification of circulating biomarkers. *A. Wolfe, A. Siniard, I. Schrauwen, M. De Both, R. Richholt, M. Huentelman.* TGen, Neurogenomics Division, Phoenix, AZ.

Circulating RNA biomarkers can be used as indicators of the presence of disease, response to treatment, or potentially as a test of an individual's health status. A significant issue facing the field of biomarker discovery right now is the general lack of cohort size for study as well as longitudinal sampling of the recruited cohort. Much of this is likely due to the fact that most biospecimens are collected using relatively invasive approaches. We sought to optimize RNA biomarker assessment in an easy to collect biospecimen, the dried blood spot (DBS) To collect biospecimens with higher frequency and use them during longitudinal analysis, we developed a DBS-based next generation RNA sequencing approach (DBS-RNA-Seq). Drops of blood (~30ul) were collected onto RNA-stabilizing filter papers and allowed to dry. On average we isolated 10ng of total RNA from one DBS. This RNA was of sufficient quality and quantity to generate Illumina transcriptome and miRNA-Seq libraries, resulting in the detection of 7,000 genes and ~500 miRNAs when sequencing a sample to a depth of ~2 million or 650,000 counts, respectively. RNA-Seq analysis of DBS that were stored for 1, 7, 30, and 60 days demonstrated the presence of no significantly differentially expressed transcripts, suggesting that RNA species collected in this fashion are stable for long periods of time. This work therefore demonstrated the feasibility and appropriate performance of the DBS-RNA-Seq approach. Next, we attempted to utilize DBS-RNA-Seq to identify circulating RNAs associated with 1 hour of aerobic exercise. Each Monday for 4 consecutive weeks, DBS were collected from one study participant at the following time points: 5am (first wake), 9:30am, every 10 minutes during exercise for a total of 6 collections, and then 1, 2, 3, and 4 hours following exercise. The aerobic exercise consisted of a 17-mile road cycling ride of 165 Watts average power. After 24 hours of drying, total RNA was isolated from each DBS and utilized for transcriptome analysis. Several transcripts known to be induced by exercise, and many new transcripts, were identified during exercise as well as during the hours immediately following exercise. This work demonstrated the utility of DBS-RNA-Seq to identify biomarkers associated with a "condition", in this case, an aerobic exercise bout. Therefore, we propose that this approach could be utilized to characterize longitudinal RNA biomarkers in larger cohorts of disease cases and controls.

2027T

Using a new multiplex CNV analysis technology detecting *SLC26A4* deletions and duplication in 84 Chinese subjects with enlarged vestibular aqueduct. *H. Yuan[1,2], Y. Lu[1,2], J. Cheng[1,2], Z. Jiang[3], P. Dai[2].* 1) Medical Genetics Center, Southwest Hospital, Third Military Medical University, Chongqing, China; 2) Dept. of Otolaryngology, Chinese PLA General Hospital, Beijing, 100853, China; 3) 3. Center for Genetic and Genomic Analysis, Genesky Biotechnologies Inc. , Shanghai, 201315, China.

About 90%subjects with enlarged vestibular aqueduct (EVA) carry biallelic mutations in*SLC26A4*gene in clinic. However, only one allelic mutation or none mutation are detected in 10% EVA subjects by Sanger sequencing. We explore to detect possible copy number variations (CNV) in Chinese EVA subjects using multiplex CNVplex® technology. Fifty-one subjects with only one identified mutant and 33 lacked any detectable mutation in *SLC26A4* gene by Sanger sequencing were recruited in this study. We apply a new developed high-throughput multiplex CNV analysis technology to detect *SLC26A4* deletions and duplications in these Chinese EVA subjects. Sanger sequencing confirmed the breakpoints and segregation in these EVA families. Five different deletions and one duplication were identified in 23 of 84 Chinese EVA subjects. Three different exon 1 to 3 deletions were identified in 10cases, and exon 5 to 6 deletion was identified in 11 cases. Exon 7-10 deletion and exon 18-21 duplication were identified in one case. Our findings indicated that CNVs in *SLC26A4* play an important role in EVA etiology. Exon 1 to 3 deletion and exon 5 to 6 deletion are two pathogenic hotspots in Chinese EVA subjects. CNV detection with CNVplex® is necessary to increase diagnostic rate in genetic testing for hearing loss by identifying CNVs in *GJB2*and other genes underlying hearing loss other than *SLC26A4*.

2028F

TINA modified primers allow faster PCR. *S. M. Echwald, N. D. Mikkelsen.* Anapa Biotech A/S, Hvidovre, Denmark.

Rapid testing is important for the clinical relevance of many diagnostic assays. As PCR has become the leading technology for molecular diagnostics, increasing PCR speed has become important. Often, the efficiency of PCR-based clinical diagnostic assays is limited by the PCR primers. Here, we report the use of ortho-Twisted Intercalating Nucleic Acid (TINA) in PCR primers which results in increased efficiency and allows significantly shorter PCR cycling time without changing other PCR conditions. By incorporating TINA in the 5' region of PCR primers, assay time for routine assays for infectious agents such as MRSA, can be reduced significantly while retaining PCR specificity and efficiency, where DNA-based assays fail. TINA is a novel intercalator-type oligo modification which stabilize Watson-Crick based antiparallel DNA duplex helixes. A single oTINA molecule increased the annealing temperature (Tm) of an antiparallel DNA duplex by 3-4 °C. TINA is developed as a standard amidite and is available as a standard oligo addition from Eurofins Genomics and TriLink Biotechnologies.

2029W

Exome sequencing in children with life-threatening community-acquired *Pseudomonas aeruginosa* infection. *J. Fellay[1], S. Asgari[1], I. Bartha[1], LJ. Schlapbach[2].* 1) School of Life Science, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, VD, Switzerland; 2) Paediatric Critical Care Research Group (PCCRG), Mater Research, University of Queensland, Brisbane, Australia.

Background: *Pseudomonas aeruginosa (P. aeruginosa)* is a very rare cause of sepsis in children without comorbidities, but carries an extremely high mortality. Fatal *P. aeruginosa* sepsis in previously healthy children may be the first manifestation of a pathogen-specific primary immunodeficiency (PID). We used exome sequencing of carefully selected children and their parents to search for rare genetic variants that confer susceptibility to this infection. **Methods**: Seven family trios were included in the study. The probands were previously healthy children, age >1 month and <17 years, who developed sepsis due to community-acquired *Pseudomonas aeruginosa* bacteremia. Six of them died from the infection. The parents were apparently healthy individuals. We obtained high-coverage exome sequencing data for all study participants. After short-read alignment, we used GATK haplotypecaller to call single nucleotide variants (SNVs) and small insertions/deletions (indels), and SnpEff for variant annotation. We searched for rare loss-of-function (LoF) variants and assumed various inheritance models. **Results**: We found a novel frameshift indel in exon 7 of the BTK gene in one of the patients. BTK mutations are known to cause X-linked agammaglobulinemia, a PID that had already been associated with severe Pseudomonas sepsis. Using a taqman assay, we confirmed the presence of the variant in hemizygous form in the child and in heterozygous form in the mother. We also demonstrated the absence of immunoglobulins in a plasma sample of the affected child. We did not identify any obvious causal variants in the 6 other cases. **Conclusions**: Our results demonstrate that *P. aeruginosa* sepsis can be caused by rare, deleterious human genetic variants, and confirm that Mendelian susceptibility to common infections can have a dramatic impact in apparently healthy children.

2030T

Molecular inversion probe based re-sequencing in a clinical setting – highlighting *BRCA1* and *BRCA2*. *A. Hoischen[1], A. Mensenkamp[1], M. Kwint[1], R. Derks[1], B. van Lier[1], E. Bosgoed[1], A. Rikken[1], M. Tychon[1], H. Ouchene[1], E. Boyle[2], B. Martin[2], K. Neveling[1], J. Shendure[2], M. Nelen[1], M. Ligtenberg[1].* 1) Department of Human Genetics, Radboud University Medical Center, Nijmegen, Netherlands; 2) Department of Genome Sciences, University of Washington, Seattle, Washington, USA.

Next generation sequencing (NGS) has revolutionized genome- or exome-wide sequencing, but there is still a need for improved NGS-based individual gene testing concerning reliability, accuracy, throughput and costs. Those values are largely determined by enrichment procedures. Molecular inversion probes (MIPs) have shown to be a cost-effective enrichment used in multiplex, particularly if used in large cohorts1. Here we present a workflow of re-sequencing *BRCA1* and *BRCA2* using single molecule molecular inversion probes (smMIP) in combination with NexSeq500 sequencing. The strategy involves the analysis of all coding exons and their flanking intronic sequences, deep intronic SNPs and a genome-wide backbone of 55 genotyping SNP markers. We designed 527 overlapping smMIPs on both strands avoiding known SNPs in probe binding sites2. Every base was targeted by at least two independent smMIPs. Next 152 *BRCA1* and *BRCA2* mutation positive cases were sequenced for both genes with an average coverage of >250-fold. This yielded 100% coverage of all targeted bases, with all coding bases to be covered at least 40x by at least one independent smMIP, and 97. 6% of coding bases covered at least 40x by at least two independent smMIPs. All diseases causing mutations and all known polymorphisms in both genes were detected, including one SNP that was missed using capillary sequencing due to allelic dropout. No false positives were reported. In addition we performed dilution series of known mutations. These provide evidence that smMIP based assays can also detect mosaic mutations with <1% mutation alleles. In conclusion, targeted *BRCA1* and *BRCA2* re-sequencing using smMIPs shows high sensitivity and specificity and can replace sequencing workflows in a clinical setting. In addition smMIPs allow low enrichment costs per sample and highly scalable workflows. References:1) O'Roak et al. Science 2012; Hiatt et al. Gen Res 20132) Boyle et al. Bioinformatics 2014.

2031F

Achieving high-sensitivity for clinical applications using augmented exome sequencing. *A. Patwardhan[1], J. Harris[1], N. Leng[1], G. Bartha[1], D. M. Church[1], S. Luo[1], C. Haudenschild[1], M. Pratt[1], J. Zook[2], M. Salit[2], J. Tirch[1], M. Morra[1], S. Chervitz[1], M. Clark[1], S. Garcia[1], G. Chandratillake[1], S. Kirk[1], E. Ashley[1,6], M. Snyder[1,5], R. Altman[1,4], C. Bustamante[5], A. J. Butte[1,3], J. West[1], R. Chen[1].* 1) Personalis, Inc. 1330 O'Brien Drive, Menlo Park, California 94025, USA; 2) Biosystems and Biomaterials Division, National Institute of Standards and Technology, Gaithersburg, Maryland, USA; 3) Institute for Computational Health Sciences, University of California, San Francisco, California 94158, USA; 4) Departments of Bioengineering & Genetics, Stanford University, Stanford, California 94305, USA; 5) Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA; 6) Center for Inherited Cardiovascular Disease, Stanford University School of Medicine, Stanford, California 94305, USA.

Whole exome sequencing is increasingly used for the clinical evaluation of genetic disease, yet the variation of coverage and sensitivity over medically relevant parts of the genome remains poorly understood. Several sequencing-based assays have base-level coverage that is inadequate for clinical assessment. Using sequence data obtained from the NA12878 reference sample and pre-defined lists of medically-relevant protein-coding and noncoding sequences, we compared the breadth- and depth-of-coverage obtained among four commercial exome capture platforms and PCR-free whole genome sequencing. In addition, we evaluated the performance of an augmented exome strategy, ACE, that extends coverage in medically relevant regions and enhances coverage in high-GC rich regions that are challenging to sequence. Leveraging reference call-sets, we examined how improved coverage effects variant detection sensitivity. We observed insufficient coverage with each of the conventional exome-capture and whole-genome platforms across several medically interpretable regions. These gaps included known disease-associated loci and areas of the genome required for reporting secondary findings (ACMG). At clinically-relevant coverage levels (100% bases covered at >20x, with map Q20 reads), ACE improved coverage among genes in the medically interpretable genome (>90% covered compared to 10-78% with other platforms), the set of ACMG secondary finding genes (91% covered compared to 4-75% with other platforms) and a subset of variants known to be associated with human disease (99% covered compared to 52-95% with other platforms). Improved coverage translated into improved accuracy. ACE variant detection sensitivities exceeded those observed with conventional whole-exome and whole-genome platforms. Clinicians should consider analytical performance when making clinical assessments, given that even a few missed variants can lead to false negative results. We show that by comprehensively targeting medically interpretable areas, including those areas that are challenging to sequence, we improve sensitivity in regions of the genome relevant for clinical interpretation. In clinical applications where comprehensive coverage of medically interpretable areas of the genome requires higher localized sequencing depth, an augmented exome approach offers both cost and performance advantages over other sequencing-based tests.

2032W

Rapid screening of severely ill newborns and infants using whole genome sequencing. *R. J. Sinke[1], C. C. van Diemen[1], W. S. Kerstjens-Frederikse[1], T. J. de Koning[1,2], B. Sikkema-Raddatz[1], J. D. H. Jongbloed[1], K. M. Abbott[1], J. C. Herkert[1], P. B. T. Neerincx[1], G. de Vries[1], M. Meems-Veldhuis[1], M. Viel[1], A. J. Scheper[1], K. de Lange[1], J. Dijkhuis[1], J. van der Velde[1], M. de Haan[1], M. A. Swertz[1], K. A. Bergman[2], C. M. A. van Ravenswaaij-Arts[1], I. M. van Langen[1], R. H. Sijmons[1], C. Wijmenga[1].* 1) Dept. of Genetics, UMC Groningen, University of Groningen, Groningen, The Netherlands; 2) Beatrix Children's Hospital, University Medical Center Groningen, Groningen, The Netherlands.

For severely ill newborns quick molecular diagnoses are of utmost importance for clinical decision-making and can prevent unnecessary and sometimes invasive diagnostics. To date, immediate molecular testing is not a routine procedure for all patients since this is available only for few diseases. Here we describe a procedure and present the first results to analyze 2800 genetic disorders in severely ill newborns and infants by rapid whole-genome sequencing (WGS). WGS is carried out in parallel to standard diagnostic procedures (including imaging, SNP-arrays, biochemical testing, mutation analysis, etc.). The entire procedure from inclusion to reporting is completed in approximately 2 weeks. The final evaluation of the results is done by a multidisciplinary team of pediatricians, clinical geneticists, technicians, clinical genetic laboratory specialists, researchers and bio-informaticians. Thus far we have included 14 patients in the study and have provided a diagnosis of a monogenic disease for three patients. These patients presented with different clinical characteristics and that could be explained by mutations in the *EPG5, KLHL41* and *RMND1* genes. One patient was diagnosed with a 1p36 microdeletion upon routine diagnostic testing. We also tested 5 patients, who died within the first year of life, and their parents with rapid clinical exome sequencing, focusing on the same set of 2800 genes. We found compound heterozygous mutations in *BRAT1* in one child with an unexplained severe seizure and rigidity disorder. Currently we are evaluating the procedure, comparing our results with those from studies of the Kingsmore's group, who published diagnostic yields of 80% (in a prospective study, Saunders et al. Sci Transl Med. 2012;4:154ra135) and 57% (in a retrospective study, Willig et al. Lancet Respir Med. 2015 35:377-387). We will emphasize on phenotype selection, technical aspects of coverage and filtering methods. One preliminary conclusion explaining the lower yield in our studies, may be that our patient inclusion was biased towards patients with an unclear, complex phenotype.

2033T

Initial evaluation in constitutional cytogenomics of CNVkit, an algorithm for genome-wide copy number determination using on- and off-target reads in whole exome sequencing data. *A. Wiita[1], E. Talevich[2], L. Cheng[1], D. Wheeler[1], Z. Qi[1], J. Yu[1].* 1) Laboratory Medicine, University of California, San Francisco, CA; 2) Pathology, University of California, San Francisco, CA.

Background: Copy-number analysis is a routine clinical test for the diagnosis of autism, developmental delay, and congenital anomalies. The current gold standard methodology is cytogenomic microarray analysis. However, next generation sequencing, particularly whole exome sequencing, is becoming a routine clinical test for genomic analysis. There remains a need for improved bioinformatic strategies to extract copy number information from these datasets, potentially eliminating the need for separate microarray testing. **Methods:** We present the initial evaluation of a newly-described bioinformatic algorithm, CNVkit, in the setting of constitutional cytogenomics. CNVkit uses both on- and off-target reads as generated by targeted next generation sequencing data to determine genome-wide copy number variation. CNVkit accurately captures complex copy number variation in cancer samples (Talevich et al, submitted). However, its performance in constitutional genetics, with infrequent single-copy gains and losses, is unknown. In addition, the majority of CNV analysis software for exome sequencing is based on Illumina platform data. The Ion Proton technology holds promise for clinical laboratories based on shorter run times, though these data may present additional analytical challenges. Here, we compare CNVkit analysis on 11 DNA samples, analyzed by both whole exome capture chip (BGI custom design) on Ion Proton and SNP array (Illumina Infinium CytoSNP-850K). **Results:** We found high concordance between CNVkit analysis of exome data and SNP array for CNVs >400 kb (19/19 variants detected). This finding resulted in an analytical sensitivity of 100% compared to array. For variants below this size threshold, however, we found minimal concordance, with only 5 of 53 variants identified by array also called by CNVkit. These results suggest that CNVkit may be able to identify the majority of clinically significant copy number variants in a constitutional genetics setting. However, the non-exonic resolution of CNVkit, ~100 kb, may limit its ability to detect smaller variants. We will perform qPCR to confirm discordant CNV calls found by SNP array. We also aim to further refine the CNVkit algorithm for use with Ion Proton sequencing data. **Conclusion:** CNVkit shows initial promise in the analysis of copy number variants in a constitutional genetics setting and on Ion Proton data, though more refinement is required to equal SNP array for diagnostic testing.

2034F

Clinical whole exome sequencing reveals previously unreported likely pathogenic variants in the SCN8A gene in patients with early onset epileptic encephalopathy with intellectual disability. *F. Vetrini[1], J. Zhang[1], W. He[1], Z. Niu[1], M. Walkievicz[1], R. E. Person[1], A. L. Beaudet[1], D. M. Muzny[2], R. A. Gibbs[1,2], J. Lupski[1], C. M. Eng[1], Y. Yang[1].* 1) Department of Molecular And Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX.

Voltage-gated sodium channels (NaV) are essential for initiating and propagating neuronal cell action potentials in the brain. Pathogenic variantsin *SCN8A*, which encodes the sodium channel voltage-gated alpha8-subunit (Nav1. 6), one of the most abundant sodium channels in the brain, have recently been recognized as one of themechanisms leading to epilepsy and intellectual/developmental disabilities (IDDs). *SCN8A* related encephalopathy presents in infancy with multiple seizure types, in some cases associatedwith sudden unexpected death in epilepsy (SUDEP), motor manifestations including hypotonia, ataxia, dystonia, hyperreflexia, and psychiatric manifestation including behavioral disturbance and autism. We have used Whole Exome Sequencing (WES) to identify *SCN8A* pathogenic variants in seven patients with epilepsy and IDDs of unknown etiology. The patients presented with a variable clinical picture including early onset epileptic encephalopathy (n=7), structural brain abnormalities (n=3), ID (n=7), hypotonia (n=2), movement disorders (n=4), seizure–associated cardiac conduction defects (n=1). A total of five *SCN8A* novel missense mutation (p. D58V, p. 216G, p. A874S, p. Q1470E and p. G1475R) were found in this patient cohort and were predicted to be damaging to the structure of NaV1. 6 channel by Polyphen2 and SIFT *in silico* prediction programs. The mutations are scattered throughout the gene and potentially affect conserved domains of the channel: one in the cytoplasmic N-terminal domain (p. D58V), one in the linker region between S3 and S4 of domain I (p. V216G), one in the linker between S4 and S5 of domain II (p. A874S), and two in the inactivation gate between domains III and IV (p. Q1470E, p. G1475R). Sanger sequencing did not identify these variants in the biological parents, however, affected siblings from two different families carried the same variants, c. 2620G>T (p. 874S) and c. 4408C>G (p. Q1470E) respectively, strongly suggesting mosaicism in a parent in the two families. For the other three families, the variants appeared to arise *de novo* in the patients. However, the possibility of low level mosaicism should still be considered when counseling parentsand estimating recurrence risk for the family. In conclusion, the reporting of five novel *SCN8A* likely pathogenic variants may help further determine genotype-phenotype relationships for patients with early onset epileptic encephalopathy with IDDs due to defects in the voltage-gated Nav1. 6 channel.

2035W

Development of a high throughput workflow for CFTR mutation screening. *T. Hartshorne, J. Lang, E. Swartzman, P. Brzoska, D. de Castro, L. Tracy.* Genetic Analysis, Thermo Fisher Scientific, South San Francisco, CA.

Cystic fibrosis (CF) is an autosomal recessive genetic disease that is caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. It is the most common genetic disease of Caucasians, affecting 1 in 3000 newborns and having a carrier frequency of 1 in 25. CF impacts mucin production in lungs and other organs, and is a progressive incurable disease. To facilitate rapid, low cost, high throughput screening of individuals that carry CFTR mutations, we set out to develop a panel of over 200 TaqMan™ assays that detect specific CF-causing mutations. Target mutations were selected from the CFTR1 and CFTR2 databases, and prioritized based on frequency and consequence. CF-causing mutations occur primarily within coding and regulatory regions and include single nucleotide base substitutions, small insertion deletions, large deletions, repetitive sequences, and triallelic polymorphisms. The diverse nature of the mutations presented challenges to developing assays that used a single chemistry and that could be run together on one platform, but for the most part, these were overcome by using various design strategies. TaqMan SNP genotyping and endpoint assays were initially tested on both 384-well plates and OpenArray™ plates (= 3072-well), run on a real-time PCR system, with Coriell cell line gDNA samples that carry CFTR mutations. Given that most CFTR mutations are rare and control gDNAs are not available for them all, assays were also tested with synthetic DNA controls representing 3 genotypes for each CFTR variant, to ensure robust discrimination of alleles and genotypes. If an assay failed to perform well, the assay was redesigned and tested until a sufficiently robust assay was produced. The testing of DNA isolated from blood and buccal cell samples is underway, as are assay accuracy and concordance studies. We will present our progress towards development of a complete sample-to-data analysis workflow for high throughput CFTR mutation detection. Example assay data and analysis methods will be shown, in particular for difficult targets such as the 5T/7T/9T polymorphism. Our CFTR genotyping workflow will enable selection of assays based on CFTR mutation occurrence in populations of interest, screening of small or large numbers of samples using customizable panels of 60, 120 or 180 assays, and highly accurate and reproducible acquisition of CFTR genotypes from blood and buccal cell samples.

2036T

Validation of a high resolution NGS method for detecting spinal muscular atrophy carriers among phase 3 participants in the 1000 Genomes Project. *J. L. Larson[1], A. J. Silver[1], D. Chan[1], C. Borroto[2], B. Spurrier[2], L. M. Silver[2,3].* 1) GenePeeks, Inc. , Cambridge, MA; 2) GenePeeks, Inc. , New York, NY; 3) Department of Molecular Biology and the Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ, USA.

Spinal muscular atrophy (SMA) is the most common pan-ethnic cause of early childhood death due to mutations in a single gene, *SMN1*. Most chromosome 5 homologs have a functional gene and dysfunctional copy, *SMN2*, with a single synonymous base substitution that results in faulty RNA splicing. However, the copy number of *SMN1* and *SMN2* is highly variable, and one in 60 adults worldwide are SMA carriers. Although population-wide screening is recommended, current SMA carrier tests have not been incorporated into targeted gene panels. Here we describe a novel computational protocol for determining SMA carrier status based solely on individual exome data. We find complete concordance with results obtained with the current qPCR-based testing standard. We applied our protocol to the phase 3 cohort of the 1,000 Genomes Project and found carrier frequencies in multiple populations consistent with the present literature. Our process is a convenient, robust alternative to qPCR, which can easily be integrated into the analysis of large multigene NGS carrier screens.

2037F

Standardizing the quality QC of any DNA isolation to safeguard the success of genetic testing. *T. Martens, T. Boonefaes, E. De Raeymaecker, T. Montoye.* Trinean, Dulle grietlaan 33, 9050 Gentbrugge, Belgium.

As genetic tests become more powerful yet also more complex and costly, it is prudent to include quality control steps throughout the entire workflow to ensure the quality of the downstream data obtained. To avoid financial and time costs due to analytical failures downstream, it is essential to include at the start an informed go/no go decision based on a good assessment of the DNA/RNA isolation yield and purity, preferably including the detection of carry-over constituents that may interfere with the tests. Here, we validate a simple approach for DNA quantification and purity QC, using a large DNA sample set derived from a variety of human tissues combined with a wide scale of extraction methods. This new QC tool combines the micro-volume spectrophotometry on the Xpose reader by Trinean with its spectral content profiling app for Mammalian gDNA to specifically quantify the isolated DNA as well as the amount of contaminating constituents in the sample contributing to the measured UV/Vis spectra. In this study, a large collection of DNA samples was used to compare the Xpose reader with recommended absorbance and fluorescence-based DNA quantification and purity analysis tools to verify its specificity for DNA quantification and impurity detection.

2038W

Genetic testing supply and demand: Tracking growth and competition in the genetic testing marketplace. *T. A. Murphy, J. D. Schneider, G. W. Hooker.* Bioinformatics, NextGxDx, Franklin, TN.

Introduction: The genetic testing landscape is growing and changing at a rapid pace, due in large part to innovations in sequencing capability. As the sequencing cost per base pair goes down, laboratories are able to expand their test offerings by introducing both new options for clinical use and, in many cases, options that compete with tests offered by other laboratories in the market. However, with broader scope testing comes the need for greater analytic infrastructure, which can be costly. Drawing from our data resource of more than 38,000 genetic tests we aimed to describe the growth in the market from June 1st, 2014 to June 1st, 2015. We also aimed to test the hypothesis that, corresponding to growth and increased competition in the market, average test price would decrease in this time period. **Methods:** We curate and maintain a database of clinically available genetic tests offered by US CLIA-certified laboratories. We collect information from laboratory websites on the tests they offer, and other available test attributes including methodologies, turn-around-times, test descriptions and test list prices. The mean time interval between updates of test information is 14 days. We tracked new tests entering the market on a monthly basis and compared average costs per test from June 2014 to June 2105. **Results:** Across this time period, we saw a net of 2,815 new tests introduced to the market, at a rate of about 8 new tests/day. 888 of these tests were multi-gene panel tests. As the market grew, the average price per test dropped by 14. 9% overall, with a 16. 8% decrease in the average price per single gene test and a 6. 5% decrease in the average price per multi-gene panel test. Of all tests in our database, clinical exome sequencing prices dropped the most, going from an average price of $8426 per exome test product to $6074 per exome test product. **Conclusions**: With expanding options for clinical genetic and genomic testing have come decreasing list price costs associated with these tests. The dynamic nature of the market suggests opportunities for utilization management and cost savings. which have the potential to increase patient access to genetic tests.

2039T

"Ion Ampliseq Custom Arrhythmia Panel" comprising 68 cardiac channelopathy genes is a gold standard for the rapid and sensitive detection of novel genes and variations in Long QT (LQT) syndrome.

B. Turkgenc[1,2], S. G. Temel[3,4], H. H. Aykan[5], A. Sulu[6], M. G. Ramoglu[7], F. Uysal[8], O. M. Bostan[8], T. Karagoz[5], F. Akalin[9], O. Baspinar[6], Y. Alanay[10,11], A. Celiker[12], S. A. Ozer[2], M. C. Yakicier[13]. 1) Acibadem Genetic Diagnostic Center, Istanbul, Turkey; 2) Marmara University, Department of Medical Biology and Genetic, Istanbul, Turkey; 3) Near East University, Faculty of Medicine, Department of Histology and Embryology, Lefkosa, North Cyprus; 4) Uludag University, Faculty of Medicine, Department of Histology and Embryology, Bursa, Turkey; 5) Hacettepe University, Faculty of Medicine, Department of Pediatric Cardiology, Ankara, Turkey; 6) Gaziantep University, Faculty of Medicine, Department of Pediatric Cardiology, Gaziantep, Turkey; 7) Ankara University, Faculty of Medicine, Department of Pediatric Cardiology, Ankara, Turkey; 8) Uludag University, Faculty of Medicine, Department of Pediatric Cardiology, Bursa, Turkey; 9) Marmara University, Faculty of Medicine, Department of Pediatric Cardiology, Istanbul, Turkey; 10) Acibadem Maslak Hospital, Department of Pediatric Genetic, Istanbul, Turkey; 11) Acibadem University, Faculty of Medicine, Department of Pediatric Genetic, Istanbul, Turkey; 12) VKV Amerikan Hastanesi, Istanbul, Turkey; 13) Acibadem University, Faculty of Science, Melocular Biology and Genetic, Istanbul, Turkey.

Background: Long QT (LQT) syndrome is a genetically heterogeneous disease with high frequence predicted between 1/2000 and 1/5000 and associated with a high risk of sudden death. It has been informed 14 different genes for LQTs until now. Approximately 25% of LQT families don't have detectable mutations in reported 14 LQT genes. The identification of the other genes responsible of LQT has a big importance on specification of molecular and diagnostic approaches. **Purpose:** Aim of our study is to investigate new cardiac channelopathy gene mutations underlying LQT's by using extremely sensitive and specific detection method "custom AmpliSeq™ panel". **Methods:** Whole blood was obtained from the proband and family members with informed consent. Genomic DNA was isolated by usingWizard® Genomic DNA Purification Kit, Promega. Targeted sequencing was performed on ion Personel Genome Machine using Custom Ion Ampliseq Panel comprising 68 genes known to be associated with cardiological arrhythmias. Data analysis was performed using Torrent Suite Software. The missense mutations were confirmed by GenomeLab™ GeXP Genetic Analysis System, Beckman Coulter. **Results:**We identified 28 different mutations (26 missense, 1 insertion and 1inframe deletion) among 12 unrelated Turkish families with LQTs. 8 of these were novel, 7 of these were reported as pathogenic alleles, the remaining 13 were reported as ''unknown" allele with MAF<0. 01%. We identified 1 novel and 3 reported KCNQ1 mutations, 1 ANK2 mutation and 2 AKAP9 variations as expected for LQTs. KCNQ1, ANK2 and AKAP9 variants were together with at least one other channelopathy variants (RYR1, CACNA1S, LDB3, SCN1B, CACNA2D1, CACNB2, BAG3, TRPM4, RBM20, MYH7, CASQ2, PKP2, DSG2 and RYR2). TRPM4, HCN4 and TNII3 variants were identified in other LQT families, separately. One case had a de-novo RYR2 and germline DSP mutation. Last patient had MYH6 and TCAP gene variations together. **Conclusion:** Our 68 gene panel for inherited *arrhythmia*sidentifies novel candidate genes responsible for LQTS and it's a gold standard on molecular diagnosis of the disease. Our study highlights that LQT panel should not be limited with only 14 gene, but also should be investigated other cardiac channelopathy genes. For future directions functional analysis of variations are under estimated. **Acknowledgements:** Inherited *arrhythmia panel is* supported by grant from SANTEZ Project (0253. STZ. 2013-2), Turkey.

2040F

Genetic and epigenetic analysis of amyotrophic lateral sclerosis (ALS) patients in the Slovenian population.

M. Ravnik-Glavac[1], K. Vrabec[1], D. Glavañ, B. Koritnik[2], L. Leonardis[2], L. Dolenc-Grošelj[2], J. Zidar[2], B. Rogelj[3]. 1) Department of Molecular Genetics, Institute of Pathology, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia; 2) Institute of Clinical Neurophysiology, Division of Neurology, University Medical Centre Ljubljana, Ljubljana, Slovenia; 3) Department of Biotechnology, Jožef Štefan Institute, Ljubljana, Slovenia; Biomedical Research Institute, Ljubljana, Slovenia.

Amyotrophic lateral sclerosis (ALS) is the most common motor neuron disease. It is typically fatal within 2-5 years of symptom onset. A complex interaction between genes, environmental exposure and impaired molecular pathways contributes to pathology in patients with ALS. Epigenetic mechanisms control the hereditary and reversible regulation of gene expression without altering the basic genetic code. Although recent genetic studies have substantially improved our understanding of the causes of ALS, especially familial ALS, an important role of non-genetic factors in ALS is also recognized and needs further study. We first performed genetic analysis of *SOD1*, *TARDBP*, *FUS* in *C9ORF72* genes in 84 Slovenain patience with sporadic form of ALS and revealed mutations p. Val14Met and p. Gly93Cys in *SOD1* gene and 2 synonymous substitutions, p. Arg522Arg in *FUS* gene and a novel change p. Leu330Leu in *TARDBP*. Screening of *C9ORF72* revealed HREM in 5 patients. In concordance with other studies *C9ORF72* is the most common detected disease causing change. The research was continued in the field of methylation. Based on the results of the whole genome methylation profiling performed on 4 samples 14 genes were selected and the results were validated in 101 patients by MS-HRM. Methylated promoters were detected in 8 genes but the promoters of 6 genes were found unmethylated. Results of MS-HRM were confirmed by Sanger sequencing. In the field of miRNA, qPCR was used to examine the expression of 10 miRNA. Increased expression of let-7b, miR-132, miR-206 and miR-638 was detected. Reduced expression of miR-338 and its host gene *AATK* was found. MiR-9 was differently expressed in ALS patients, but its host gene *C1ORF61* was not expressed in blood. Two intergenic miRNAs, miR-663a which had variable expression across samples and miR-124a whose expression was reduced, were analyzed by MS-HRM. Results showed that methylation of their promoters are not the cause for reduced expression. MiR-143 and miR-451a both showed decreased expression but no CpG islands were found in their promoters. While some progress has been made in recent years, the majority of the genetic and epigenetic variations involved in the ALS are yet to be identified. .

2041W

The most variable region of the genome: The next generation of HLA-typing platforms. *J. Kaur[1], J. Sims[2], E. Earley[2], K. Robasky[2], E. Lai[1].* 1) Takeda Pharmaceuticals International, Inc, Deerfield, IL; 2) EA|Quintiles, Durham, NC.

The Human Leukocyte Antigen (HLA) region of the genome encodes self-recognition proteins, thereby blocking autoimmunity. For decades, data have been collected on this region in support of transplantation tissue-matching, but HLA-typing also has many other important applications. In addition to informing treatment for autoimmune dysfunction, recent advances in oncotherapy exploit HLA gene expression in cancer. Furthermore, HLA-typing has been used to predict adverse drug responses such as hepatotoxicity. However, these genes also represent the most variable region of the human genome, and typing them by traditional methods is both manual and time consuming. Fortunately, variants in the HLA genes display high linkage disequilibrium (LD) and the most common haplotypes are well documented and limited in number. Emerging HLA-typing technologies leverage these characteristics through multiplexed primer design and phasing of the HLA region to yield less costly, highly sensitive and specific HLA allele recognition. Here, we contrast and compare these nascent NGS-based and array-based platforms against the polymerase chain reaction (PCR) and sequence-specific oligonucleotide probes protocol from Luminex (PCR-SSOP-Luminex). We performed these assays on patient-derived samples which we population-stratified using hapmap data for controls. Using a reference panel consistent with the sample population, we found that HLA-types could be imputed on arrays with platform-agnostic phasing software at better than 90% concordance for 2-digit accuracy, and less concordance for 4-digit accuracy. When comparing the Omni5 to the Affymetrix Axiom (used for the UK Biobank), the Axiom array has significantly fewer probesets than the Omni5, and also fewer loci over-lapping with markers in the reference panel. This difference affects the concordance estimates, with the Omni showing 89. 33% concordance with the gold-standard on Class I alleles and the Axiom 83. 33% concordance. This is contrasted with the NGS-based assay which shows 99. 33% concordance. Notably, the price point for array technologies are an order of magnitude lower than NGS-based assays, which are more affordable still than the gold-standard. Informed by platform-specific cost data, we offer additional evidence-based guidance for contextual choice of HLA-typing technology.

2042T

DIRECT targeted library preparation for Illumina Sequencing. *C. Hendrickson[1], K. Patel[1], N. Henig[1], T. Davis[2], S. Russello[2], B. Galvin[1].* 1) Directed Genomics, Ipswich, MA; 2) New England Biolabs, Ipswich, MA.

Next-generation sequencing, coupled with targeted capture of disease-associated genes, has made the transition from research and translational studies to patient screening and diagnostic testing. However, current target enrichment methods, developed initially for research, lack the speed, simplicity, and specificity required for effective clinical execution. To meet these requirements, we have developed the DIRECT target capture method that offers significant advantages over traditional in-solution hybridization and multiplex PCR protocols. The DIRECT method for enrichment enables rapid capture of desired DNA fragments followed by conversion of these fragments into a highly specific, Illumina-compatible library within 7 hours, enabling sample-to-sequencing in one day. To validate the efficacy of this methodology with both intact DNA and fragmented circulating tumor DNA, we present data generated from our CancerDIRECT panel designed to capture genes commonly mutated in cancer. Our data demonstrates that this approach captures targeted sequence with high specificity and uniformity, and detects variants with great accuracy and sensitivity.

2043F

PCR based target enrichment for NGS panels and Sanger variant confirmation. *S. Lefever[1,2], J. Vandesompele[1,2], F. Coppieters[1,2].* 1) Ghent University, Center for Medical Genetics, Ghent, Oost-Vlaanderen, Belgium; 2) pxlence, Dendermonde, Oost-Vlaanderen, Belgium.

Targeted resequencing is an important application in clinical diagnostics. A wide range of target enrichment strategies have been developed, enabling the customer to focus on genomic regions of interest. Enrichment does not only significantly reduces sequencing costs per sample but also facilitates downstream data analysis considerably and imposes less ethical concerns. Due to its flexibility in design, high sensitivity and specificity, the polymerase chain reaction (PCR) is ideally suited as enrichment strategy. We developed and validated a primer design tool called primerXL and have generated almost one million assays for both fresh frozen and formalin-fixed paraffin-embedded (FFPE) samples, covering over 98. 7% of the human exome. Assays were designed to limit single nucleotide polymorphism (SNP) presence in primer annealing sites and minimize off-target amplification, thus generating assays achieving equimolar and maximal specificity under uniform PCR conditions. We have validated 2200 assays sampled from 200 disease causing genes. Wet-lab success-rate was determined to be over 96. 5 % upon testing by means of quantiative PCR without any optimization. Amplicons for two commercial reference DNA samples were pooled and sequenced on a MiSeq instrument. Sequencing results showed unprecedented equimolar coverage across the different amplicons with over 2/3rds of the amplicons having a coverage within 2-fold of the mean (and almost 90% within 5-fold of the mean). Specificity was very high with almost 95% of the assays having less than 2% of off-target mapping. Using our assays, NGS gene panels have already been developed for congenital blindness (16 genes), deafness (15 genes) and various cancer types (16 genes) achieving uniform sequencing coverage using different library preparation methods and sequencing instruments. To date, the Center for Medical Genetics in Ghent incorporates the primerXL assays in a high-throughput singleplex enrichment workflow to replace Sanger sequencing-based diagnostic tests with NGS (under ISO15189 accreditation). Due to the excellent performance the assays, a spin-off called “pxlence” was recently founded. Its short-term goal is to provide customers easy access to the predesigned assays, enabling them to enrich any exonic region or confirm any variant of interest through either NGS or Sanger sequencing. More information is available at www. pxlence. com.

2044W

Comprehensive Genetic Exploration of Skeletal Dysplasia Using Targeted Exome Sequencing. *J. Bae[1,2], N. Kim[1], W. Park[1,2,4], T. Cho[3].* 1) Samsung Genome Institute, Seoul, South Korea; 2) Department of Health Sciences and Technology, Samsung Advanced Institute for Health Sciences and Technology, Sungkyunkwan University, Seoul, Korea; 3) Division of Pediatric Orthopaedics, Seoul National University Children's Hospital, Seoul, Korea; 4) Department of Molecular Cell Biology, Sungkyunkwan University School of Medicine, Suwon, Korea.

Skeletal dysplasias are a group of genetic disorders of the skeletal system caused by mutations of genes involving bone and cartilage metabolism. A recent update of the nosology and classification of genetic skeletal diseases included 40 groups and 456 disease entities, listing 226 causative genes that have been discovered for 316 diseases. Clinical diagnosis of the specific disease entity causing skeletal dysplasia can be significantly difficult for several reasons. The purpose of this study was to evaluate the clinical utility of targeted exome sequencing (TES) as a molecular diagnosis tool for patients with skeletal dysplasias. A total of 186 patients diagnosed with or suspected to have skeletal dysplasia were recruited over a period of 3 years. TES for 255 genes associated with the pathogenesis of skeletal dysplasia was performed, and candidate variants were selected using bioinformatics analysis. All candidate variants were confirmed by Sanger sequencing, correlation with the phenotype and co-segregation study in the family. TES detected 'confirmed' or 'highly likely' pathogenic sequence variants in 74% (71 cases) of the assured clinical diagnosis group, as well as pathogenic sequence variants in 11 cases (16. 9%) of the uncertain clinical diagnosis group. TES successfully detected pathogenic variants in all 25 cases of previously known genotypes. TES data also suggested a copy number variation that led to a molecular diagnosis. TES data also suggested a copy number variation that leads to a molecular diagnosis. We demonstrated the feasibility of TES for molecular diagnosis of skeletal dysplasia. Further confirmation is needed for a final molecular diagnosis, including Sanger sequencing of candidate variants and suspected but poorly captured exons.

2045T

Discovery of small RNA biomarkers of hypertension through next-generation sequencing of urinary exosomes. *A. R. Davis[1], S. K. Sen[1], R. Quarells[2], C. Oguz[1], J. Vargas[1], S. Davis[1], G. H. Gibbons[1], MH-GRID.* 1) Cardiovascular Disease Section, National Human Genome Research Institute, 10 Center Drive, RM 7N321, Bethesda, MD. 20892; 2) Cardiovascular Research Institute, Morehouse School of Medicine, 720 Westview Drive Atlanta, GA 30310.

BackgroundGiven the fundamental role of the kidneys in the etiology of hypertension, biological signals of this disease may be found in urine. Discovery of urine biomarkers could improve diagnostics for essential hypertension. Exosomes are microscopic vesicles, 50-90 nm in diameter, excreted from renal tubular epithelial cells into urine. Recent advances in massively parallel sequencing allow for sub-nanogram amounts of nucleic acids (on the scale of regular exosomal miRNA content) to be analyzed using miRNA-Seq. However, no studies exist that have attempted this with small volumes of urine that are in the range of regular specimen collection in clinical studies. Hence, we set out to do a proof-of-principle study for the use of massively parallel sequencing on urinary exosome miRNA for discovery of hypertension biomarkers. ResultsSixteen 1ml frozen urine samples (eight each with and without Norgen preservative, respectively) were compared using the Norgen Urine Exosome RNA Isolation Kit to determine whether RNA yield would support massively parallel sequencing. We found substantially higher miRNA yields from urine with Norgen preservative (mean yield 62pg vs. 387pg). After determining the preferred biomaterial (urine with preservative), six low-value 1ml samples were chosen for a miRNA-Seq pilot project. Since the efficacy of miRNA-Seq library preparation with picogram amounts of RNA is not yet feasible, extracted RNA samples were amplified with the NCode™ miRNA Amplification System after adding a non-human synthetic spike-in miRNA control. Sequencing libraries were constructed using the Illumina TruSeq Small RNA kit, multiplexed and sequenced on a HiSeq2500 sequencer. Five of the six libraries passed QC and gave data close to the targeted 10M reads (average 9. 9M reads) while one library had lower yield and produced only 1. 6M reads. Average alignment rate across the six libraries was 60. 15 percent. Interestingly, along with the miRNAs, sequence data contained snoRNA, snRNAs, tRNAs and piRNAs, highlighting the diversity of exosomal small RNA content. ConclusionsWe demonstrate the feasibility of using limited amounts of urine from clinical subjects as substrates for biomarker discovery using massively parallel sequencing. To the best of our knowledge, our project represents the first examination of the urinary exosomal miRNome as it relates to hypertension and opens the door to exciting new discoveries of diagnostic biomarkers.

2046F

Maternal capillary blood: A new source of circulating cell-free fetal DNA for noninvasive prenatal testing. *R. Primacio, C. Jacob, D. Potter.* Gateway Genomics, San Diego, CA.

Noninvasive prenatal testing (NIPT) is currently limited to maternal venous blood collection by trained phlebotomists, which can restrict accessibility to prenatal screening. We aimed to identify a new source of circulating cell-free fetal DNA that would be convenient for broad population screening and reliable enough for noninvasive prenatal diagnosis. Maternal capillary blood samples were obtained from pregnant women (6-24 weeks gestation) by finger stick. Plasma was separated from whole blood by centrifugation, and circulating cell-free DNA was isolated using a commercial DNA extraction kit. Real-time quantitative PCR was performed to detect fetal DNA using a multi-copy sequence on the Y chromosome. An endogenous control gene was used to measure total cell-free DNA (maternal and fetal). We detected cell-free DNA in all maternal capillary blood samples. Y-chromosome specific sequences were detected in all pregnancies confirmed to have a male fetus. All gender results were in concordance with known fetal sex, without false-positive or false-negative results. The overall diagnostic accuracy was 100%, and provided 100% sensitivity and specificity. This new method for prenatal diagnosis of fetal gender from maternal capillary blood is simple, accurate, and reliable. The results of this study demonstrate that fetal DNA detection using maternal capillary blood is highly feasible and easily adaptable for population screening. This method simplifies collection of maternal blood and should increase the accessibility of NIPT.

2047W

Capturing phenotype data and standardising recruitment for rare diseases in the Genomics England 100,000 genome project. *A. D. Devereau[1,2], M. Ryten[1,3,4], E. Thomas[1,3], C. Turnbull[1,5], A. Milward[1], T. Fowler[1,6], M. J. Caulfield[1,7].* 1) Genomics England, Queen Mary University of London, Dawson Hall, Charterhouse Square, London, EC1M 6BQ, UK; 2) Manchester Centre for Genomic Medicine, 6th Floor, St Mary's Hospital, Oxford Road, Manchester, M13 9WL, UK; 3) Department of Medical & Molecular Genetics, King's College London, Guy's Hospital, London, UK; 4) Reta Lila Weston Research Laboratories, Department of Molecular Neuroscience, University College London (UCL) Institute of Neurology, London, UK; 5) Section of Cancer Genetics, Institute of Cancer Research, Sutton, UK; 6) UK Public Health, Epidemiology and Biostatistics, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK; 7) William Harvey Research Institute and NIHR Biomedical Research Unit in Cardiovascular Disease at Barts, Queen Mary University of London, London, UK.

The Genomics England 100,000 genomes project will recruit and whole genome sequence (WGS) 50,000 English National Health Service (NHS) patients within its Rare Diseases Programme. It aims to transform diagnostic services and improve the understanding of rare human diseases. The reliable identification of genuine disease-associated genetic variants from the broader background of variants present in all human genomes is challenging. Successful implementation of WGS therefore depends on the selection of patients with a high probability of having a disease with an underlying monogenic cause. Equally, detailed and standardised phenotypic description of each patient is essential for the successful analysis of their genomic data. Therefore the implementation of the Rare Diseases Programme has required the development of clear guidelines regarding patient recruitment and detailed models for the collection of phenotypic information. Given the complexity of rare diseases (over 120 are included within the Programme) it has been necessary to provide guidance in a disease-specific manner. This has been achieved through the creation of disease-specific eligibility statements and phenotypic capture models. The eligibility statements provide clear inclusion and exclusion criteria and specify prior testing that should have been completed before recruitment. The data models encompass qualitative phenotypic description using the Human Phenotype Ontology (HPO) and capture of quantitative clinical and laboratory tests. These models have been built within an online catalogue which enables generation of disease-specific data collection forms within recruitment centers, generation of XML data models for data upload, and version control of the complex modelling process. We report on our experiences of developing these resources and of secure data within existing standards. We also report on the value of the process in itself as a means of defining best practice in the assessment and diagnosis of rare diseases, and clarifying relationships between independent diseases, as evidenced by the frequent re-use of models. The creation and publication of these resources will be one of the legacies of the Genomics England programme.

2048T

HLA variant identification techniques in African Populations. *M. O. Pollard[1], S. Peacock[3], N. Park[2], C. Pomila[1], D. Gurdasani[1], M. Quail[2], M. S. Sandhu[1].* 1) Genetic Epidemiology Group, Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 2) Sequencing R&D, Wellcome Trust Sanger Institute, Cambridge, United Kingdom; 3) Histocompatibility and Immunogenetics Laboratory, Cambridge University Teaching Hospitals NHS Foundation Trust, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 0QQ, UK.

The Human Leukocyte Antigen (HLA) genes located on chromosome 6 are responsible for regulating immune function via antigen presentation and are one of the determining factors for stem cell and organ transplantation compatibility. Additionally various alleles within this region have been implicated in autoimmune disorders, cancer, vaccine response and both non-infectious and infectious disease risk. The HLA region is highly variable; containing repetitive regions; and co-dominantly expressed genes. This complicates short read mapping and means that assessing the effect of variation within a gene requires full phase information to resolve haplotypes.

One solution to the problem of HLA identification is the use of statistical inference to suggest the most likely diploid alleles given the genotypes observed. The assumption of this approach is the availability of an extensive reference panel. Whilst there exists good population genetics data for imputing European populations, there remains a paucity of information about variation in African populations. Filling this gap is one of the aims of the Genome Diversity in Africa Project and as a first step we are performing a pilot study to identify the optimal method for determining HLA type information for large numbers of samples from African populations.

To that end we have obtained samples from 125 consented African participants selected from 5 populations across Africa (Morrocan, Ashanti, Igbo, Kalenjin, and Zulu). The methods included in our pilot study are Sanger sequencing (ABI), NGS on HiSeqX Ten platform (Illumina); long-range PCR combined with single molecule real-time (SMRT) sequencing (PacBio); and for a subset of samples library preparation on GemCode Platform (10x Genomics), which delivers valuable long range contextual information, combined with Illumina NGS sequencing.

Results from capillary sequencing suggests the presence of a minimumof twonovel alleles. Long Range PCR have been performed initially on a subset of samples using both primers sourced from GenDX and designed as described in Shiina et al (2012). Initial results from both primer sets were promising on Promega DNA test samples but only the GenDX primers proved effective on the African samples, producing consistently PCR products of the expected size in the Igbo, Ashanti, Morrocan and Zulu samples. We will present early results from our evaluation of the different sequencing technologies.

2049F

Impact of reasonable genetic testing in prevention of rare genetic disorders. *F. Hashemi-Gorji[1], V. R. Yassaee[1,2], A. Khojasteh[3], P. Toossi[4], Z. Ravesh[1].* 1) Genomic Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 2) Dept. of Medical Genetics, Faculty of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 3) Dept. of Oral and Maxillofacial Surgery, Dental School, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 4) Dept. of Dermatology, Skin Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Mandibuloacral dysplasia with type A lipodystrophy (MADA) is a rare genetic disorder inherited in an autosomal recessive fashion characterized by hypoplasia of the mandible and clavicles, acro-osteolysis and lipodystrophy due to mutations in the *LMNA* or *ZMPSTE24* genes. In the current study, we have investigated a consanguineous family clinically diagnosed with Mandibuloacral Dysplasia type A having an affected child for the *LMNA* gene alteration(s). Mother is now pregnant (12 weeks), seeking an advice for the fetus condition, ensuring to have a healthy baby. Peripheral blood was obtained from all family members after informed consent was achieved. Genomic DNA was isolated. The sequence of the *LMNA* gene, including all exons and exon-intron boundaries were analyzed by PCR and Sanger sequencing. Molecular analysis ascertained a homozygous mutation c. 1620G>A (p. M540I) in the affected child and heterozygous alteration in the rest of the family. Genomic DNA was isolated from CVS sample received, amplified using specific primers for identified deleterious mutation and analyzed by Sanger sequencing. Two pathogenic mutations c. 1620G>A and c. 1698C>T was identified in the fetus in heterozygous state. Genetic counseling as well as justified rapid and sensitive genetic testing can provide reassurance for the families to prevent the high burden of genetic disorders.

2050W

Application of Serum miRNA Signature for Minimization of Immunosuppression and Diagnosis ofRejection Following Liver Transplantation. *B. Keating[1], B. Chang[1], T. Guettouche[2], S. Asare[3], D. Phippard[3], M. DesMarais[4], A. Shaked[1].* 1) University of Pennsylvania, Philadelphia, PA; 2) Children's Hospital of Phialdelphia, PA; 3) Immune Tolerance Network, San Francisco, United States; 4) UCSF, San Francisco, CA.

Over 14,000 liver transplants are performed worldwide yearly. Even with advances in graft management, and immunosuppression therapy (IST), rates of acute cellular rejection (ACR) remain significant, and co-morbidities from ISTs such as nephrotoxicity and liver damage are still major clinical issues. The ability to sub clinically identify liver graft recipients on ACR trajectories, and those that can successfully withdraw from IST using minimally invasive, robust, reproducible biomarkers with high specificity and sensitivity would represent a major advancement in personalized patient care. We performed miRNA profiling of 752 transcripts on 318 serum samples from 90 liver recipients transplanted from the NIH Immune Tolerance Network immunosuppression withdrawal-030 and the NIH Clinical Trials in Transplantation03 studies. 48 recipients were randomized to supervised withdrawal of IST, prior to a clinically indicated biopsy event, and were analyzed for prediction of rejection and to identify those who may tolerate lower IST. Results: Serum miRNA profiles at time of biopsy from 104 samples with and without biopsy proven ACR were compared in a two-stage study. 15 miRNAs were observed to be significantly associated with ACR diagnosis after multiple testing corrections (FDR-adjusted p-value < 0. 05). A logistic regression model consisting of a 3 miRNA panel was identified that could differentiate ACR from non-ACR with an AUC of 0. 90 (95%CI=0. 84-0. 95), 92. 6% sensitivity & 84. 2% specificity [p=0. 0001]). This 3 miRNA ACR signature was tested in an independent validation set (sera from 19 ACR & 16 non-ACR patients), confirming the performance of the model to differentiate ACR vs non-ACR (AUC of 0. 89 [95%CI: 0. 83 - 0. 94], 84% sensitivity & 75% specificity, p = 0. 01). Statistically significant alterations in this three miRNA ACR panel preceded the rejection event by up to 40 days. The composite score of another distinct 3 miRNA panel early after initiation of IST minimization (at 75% pre-withdrawal dose) identified recipients able to withstand significant IST reduction (<25% of pre-withdrawal dose, AUC=0. 88 [95%CI: 0. 80 - 0. 95], sensitivity=0. 82, specificity=0. 90, p = 0. 02). Conclusion: The findings in these prospective clinical trials demonstrate that two distinct multi-marker miRNAs signatures from sera can be used to: diagnose ACR up to 40 days before the manifestation of clinical symptoms; predict rejection trajectories; and guide personalized minimization of IST.

2051T

A protein glycosylation screen to diagnose rare genetic disorders and unravel disease mechanisms. *M. Davids[1,2], M. Kane[1,2], X. Li[3,4], M. Mohd[3,4], M. He[3,4], W. Gahl[1,2], C. Boerkoel[5].* 1) NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, MD; 2) Office of the Clinical Director, NHGRI, National Institutes of Health, Bethesda, MD; 3) Department of Pathology and Laboratory of Medicine, University of Pennsylvania, Philadelphia, PA; 4) The Michael J Palmieri Metabolic Laboratory, Children's Hospital of Philadelphia, Philadelphia, PA; 5) University of British Columbia, Vancouver, Canada.

The NIH Undiagnosed Diseases Program(UDP) concentrates its efforts on the most puzzling medical cases, focusing on disease discovery and patient diagnosis using both genomic data and basic research. Some UDP patients could have Congenital Disorders of Glycosylation (CDGs), whose phenotypes lack consistent clinical features or biomarkers, and are difficult to diagnose. Therefore, we undertook an agnostic screen of plasma and urine to identify abnormalities of protein glycosylation in UDP patients. Plasma and urine screens were completed for 110 patients, and included profiling of plasma N-linked glycans by MALDI-TOF, plasma O-linked glycans by LC-MS/MS and MALDI-TOF, and urine free oligosaccharides by MALDI-TOF/TOF. Over 75% of UDP patients had a glycosylation profile that deviated from a healthy control population in at least one specimen. Additional evaluation of the N-linked and O-linked glycosylation profiles in primary dermal fibroblasts of 56 patients with glycome abnormalities resulted in a tractable glycome phenotype in ~50% of these cases. Using this approach, we identified different CDGs, lysosomal storage disorders (affecting glycoprotein degradation), and rare disorders in which glycosylation is affected indirectly. We also added abnormal glycosylation findings to the phenotypical spectrum of a known rare disease, shedding light on a potential factor of the underlying pathogenic mechanism. In summary, we described a comprehensive screening approach to identify primary protein glycosylation abnormalities and provide a useful tool for biochemically phenotyping patients with unknown diseases and unraveling the underlying mechanisms of diseases.

2052F

Identification of *SPINK* 5 mutations in consanguineous families confirms the diagnosis of Netherton syndrome. *A. Mirzaei[1], L. Youssefian[1,2], H. Vahidnezhad[1,2,3], S. Raftari[1], S. Sotoudeh[2], Q. Li[1], J. Uitto[1].* 1) Thomas Jefferson University, Philadelphia, PA; 2) Tehran University of Medical Sciences, Tehran, Iran; 3) Pasteur Institute of Iran, Tehran, Iran.

Netherton syndrome (MIM 256500) is a rare autosomal recessive genetic disease characterized by ichthyosiform erythroderma, ichthyosis linearis circumflexa, hair shaft abnormality ["bamboo hair"], atopy and increased risk of recurrent infections. This disease is caused by *SPINK5* gene mutations that lead to absent or decreased expression of the lymphoepithelial Kazal-type-related protease inhibitor (LEKTI) protein in all stratified epithelia. In this study, we have examined 3 patients from consanguineous families diagnosed by characteristic clinical features. Our patients were 2, 4 and 31 year old females who showed generalized skin lesions since birth. Microscopic analysis of the hair shaft revealed in all of our patients typical "bamboo hairs"of trichorrhexis invaginata. Mutation analysis by sequencing of all 34 exons of *SPINK5* in these patients revealed homozygous recurrent loss-of-function mutations, consisting of a nonsense mutation or a nucleotide insertion resulting in frameshift of translation: previously published g. 31358dupT (an insertion of T in exon 9) in one family, and recurrent c. 2368C>T (Arg790Ter) in exon25 in two families. Identification of these mutations in *SPINK5* confirms the clinical diagnosis of Netherton Syndrome and forms the basis for genetic counseling, prenatal testing and preimplantation genetic diagnosis.

2053W

The spectrum of pathogenic and likely pathogenic variants observed in individuals with melanoma undergoing inherited cancer testing. *D. Farengo-Clark, S. Hiraki, M. Marshall, L. Susswein, K. Hruska, R. T. Klein.* GeneDx, Gaithersburg, MD.

About 5-10% of malignant melanoma cases (MM) occur in familial clusters, and pathogenic and likely pathogenic variants (PV/LPV) in *CDKN2A* are identified in 35-40% of familial cases. PV in *CDKN2A* increase the risk of pancreatic (PC), breast, and brain cancer. Other MM susceptibility genes including *CDK4, PTEN, BRCA2* and *BAP1* have been reported. The aim of this study was to identify the spectrum of PV/LPV observed in individuals with a personal and/or family history of MM, multiple primary MM and/or PC. The clinical histories of all individuals reporting a personal history of MM who underwent multi-gene panel testing (not including *BAP1*), *CDKN2A*/*CDK4* or *BRCA1/2* sequencing and deletion duplication analysis at GeneDx were reviewed. Two-tailed Fisher's exact test was used to determine statistical significance. A total of 301individuals reported a personal history of MM, 9% (27/301) of whom were found to harbor a PV/LPV in one of the genes tested. PV/LPV were identified in *CHEK2* (7), *ATM* (6), *BRCA2* (4), *BRCA1* (2), *FANCC* (2), and one each in *CDKN2A, PMS2, PTEN, RAD51C, RAD51D* and *VHL*. Nine of 301(3%) individuals reported >3 primary MM, 3 of whom were found to carry a PV/LPV (*CHEK2, CDKN2A* and *BRCA2*). The presence of >3 MM was significantly associated with carrying a PV/LPV compared to those with <3 MM (p=0. 037). Seventy-seven patients (25%) reported a family history of MM, 7 of whom (9. 1%) carried a PV/LPV (ATM (3), CHEK2, CDKN2A, FANCC and PTEN). This yield was not significantly different from those without a family history of MM (p=0. 25). Nine individuals had a diagnosis of both MM and PC and 22% (2/9) carried a PV/LPV (ATM, VHL). The only carrier of a *CDKN2A* PV was diagnosed with three MM before age 40 and had a family history of both MM and PC. While familial MM is associated with PV in *CDKN2A*, we found that 81% of PV/LPV identified in patients with MM were in other genes, including a large proportion in *CHEK2* (26%) and *ATM* (22%). As MM is a multi-factorial disease, it will be important to further investigate whether these genes contribute to MM risk and the level of that risk. The overall frequency of PV/LPV was significantly higher in individuals with >3 primary MM and was not associated with a family history of MM. PV/LPV in CDKN2A were identified in a small fraction of patients with MM or familial MM, suggesting that other factors may account for the majority of MM.

2054T

Identification of pathogenic/likely pathogenic variants in individuals reporting a personal history of both breast and colorectal cancer. *M. Marshall, M. Roberts, L. Yackowski, S. Solomon, D. Farengo-Clark, Z. Xu, R. T. Klein, K. Hruska.* GeneDx, Gaithersburg, MD.

Breast (BC) and colorectal (CRC) cancers are both individually common in the general population, but knowledge about a single gene predisposing an individual to an increased risk for both cancers is limited. The focus of this analysis was to determine the yield of pathogenic/likely pathogenic variants (PV/LPV), in any gene offered as part of a multi-gene cancer panel, in individuals with a personal history of both BC and CRC. We searched our Inherited Cancer panel tests performed at GeneDx through February 2015 for individuals reporting a personal history of both BC and CRC. Clinical histories were reviewed to evaluate which patients met the National Comprehensive Cancer Network (NCCN) Hereditary Breast and Ovarian Cancer (HBOC) genetic testing criteria, Amsterdam II Criteria (ACII), and Bethesda Guidelines (BG). Two-tailed Fisher's exact test was used to determine statistical significance compared to those reporting a personal history of only BC or CRC. A total of 176 individuals with a personal history of both BC and CRC were identified; 23 PV/LPV were identified in 21 individuals (*ATM* (1), *BRCA1* (4), *CHEK2* (7)*, MLH1* (2)*, MSH2* (1)*, MSH6* (5), *PALB2* (1), *PMS2* (1), and *PTEN* (1)). *CHEK2* PV/LPV were observed in 5. 0% (7/139) of individuals undergoing testing with a panel that included this gene. This rate is higher than that observed in individuals reporting a personal history of BC (2. 7%, 218/8168, p =0. 10) or CRC (2. 0%, 20/991, p = <0. 05) alone. Of the 17 individuals for whom sufficient clinical information was available, 54. 7% (11/17) met NCCN HBOC criteria, 11. 8% (2/17) met ACII, 54. 7% (11/17) met BG, and 41. 2% (7/17) met both BG and NCCN HBOC criteria. If only BRCA testing was performed for those meeting NCCN HBOC criteria, 11/12 mutations would have been missed. Forty-six% (5/11) of mutations would have been missed if those meeting BG were only tested for Lynch syndrome-associated genes**.** PV/LPV were most common in *CHEK2* and were observed more frequently in individuals reporting both cancers as compared to those reporting a history of only one. Not only would the majority of these PV/LPV have been missed with single syndrome testing, but PV/LPV were predominantly identified in a gene associated with a lifetime risk for breast cancer satisfying the American Cancer Society recommendation for annual breast MRI, thus changing medical management recommendations based on the genetic test results and making an argument for multi-gene panels.

2055F

Massive analysis of cancer genes in high cancer risk patients by next generation sequencing. *J. Puig-Butille[1,3,4], L. Jimenez[1], P. Jimenez[3,4], M. Potrony[2,3], J. Malvehy[2,3,4], C. Carrera[2,3,4], G. Tell[3,4], C. Badenas[1,3,4], J. Oriola[1,5], M. Milà[1,4], S. Puig[2,3,4,5].* 1) Biochemical and Molecular Genetics Service, Hospital Clinic & IDIBAPS (Institut d'Investigacions Biomèdiques August Pi i Sunyer). Barcelona, Spain; 2) Dermatology Department, Hospital Clinic & IDIBAPS (Institut d'Investigacions Biomèdiques August Pi i Sunyer). Barcelona, Spain; 3) Melanoma Unit, Hospital Clinic & IDIBAPS (Institut d'Investigacions Biomèdiques August Pi i Sunyer). Barcelona, Spain; 4) Centro Investigación Biomédica en Enfermedades Raras (CIBERER), ISCIII. Barcelona, Spain; 5) Faculty of Medicine. University of Barcelona. Spain.

Introduction: Several germline alterations increase the susceptiblity to a wide spectrum of malignances, such as mutations in *CHEK2, STK11, MUTYH, BRCA1/2* or *CDKN2A* gene. Thus, in cancer patients with family history of different neoplasias the study of cancer related genes by a "naive" approach could be useful to identify high or medium penetrance cancer risk alleles. Aim of the study: To asses the utility to molecular screening of a cancer related gene panel in families with a wide range of neoplasias. Methods: The study includes 12 patients carrying high-penetrant mutations previously identifed (Validation Set) and 19 wild-type *CDKN2A* familial melanoma or multiple primary melanoma patients (MPM) (Analyses Set) from 19 families with affected individuals by other neoplasias (1st or 2nd relatives). Analyses was conducted using the TruSight Sequencing Cancer Panel Kit (Illumina, USA) and sequencing the library in a MySeq platfform (Illumina, USA). Raw sequecing data was analyzed by an standard pipeline using the DNAnexus cloud-based data analysis and management platform (https://www. dnanexus. com). Variants were filtering based on quality values and effect (synomimous variants and those non-synonimous variants predicted as bening by both SIFT and PolyPhen-2 tools were removed), and were classified into a five-tiered system based on multiple databases. Results: All mutations in patients from Validation Set were detected (mutations were located in BRCA1/2, XPC, MSH2, MLH1, APC, CDKN2A or MITF gene). Overall, 319 variants were detected among the high risk melanoma patients: 277 were classifed as benign/likely benign, 34 as variant of uncertain singificance (VUS) and 8 as pathogenic/likely pathogenic variant. A pathogenic variant was detected in 21% (4/19) of patients, likely pathogenic in 15,8% (3/19) and VUS in 31,5% (6/19). Pathogenic mutations were detected in BRCA1 (p. Lys654SerfsTer47), MUTYH (p. Gly396Asp, p. Asn335ThrfsTer) and ERCC3 (p. Gln586ArgfsTer25). Conclusion: In cancer patients reporting a familial history of a broad spectrum of malignances, a massive analysis of cancer related genes is a useful approach to detect high/medium penetrant mutations. However, co-segregation anaylses in the other cancer affected relatives within the family should be performed to clarify the role of these variants.

2056W

Association between single nucleotide polymorphism of BMP5 gene and risk of knee osteoarthritis. *RN. Srivastava[1], AC. Sharma[1], A. Mishra[1], SR. Srivastava[1], K. Baghel[1], S. Raj[2].* 1) Orthopaedic Surgery, King George's Medical University, Lucknow, Uttar Pradesh, India; 2) M S Ramaiah Medical College, Bangalore.

Introduction: Osteoarthritis (OA) is a chronic degenerative disorder of multifactorial etiology characterized by loss of articular cartilage, hypertrophy of bone at the margins, subchondral sclerosis and range of biochemical and morphological alterations of the synovial membrane and joint capsule. OA is probably affected by both genetics and environmental causes. Bone morphogenetic proteins (BMPs) are bone-derived factors that can induce new bone formation. Previous study identified few single nucleotide polymorphisms SNP's that altered transcriptional activity of the BMP5 promoter, which implies that polymorphism of BMP5 gene is involved in OA susceptibility. In this study, we investigated the possible correlation between the SNPs rs1470527 and rs9382564 in BMP5 gene and susceptibility to knee OA. **Material & Methods:** : This study was conducted in the department of Orthopaedic Surgery, King George's Medical University (KGMU), Lucknow. In this study cases consisted of men and women ≥40 years that fulfilled American College of Rheumatology (ACR) clinical and radiographic criteria for knee OA. Venous blood samples were obtained from all cases as well as controls for genetic analysis. Polymerase chain reactions were performed for SNP analysis using sequenced specific primer. **Results:** A total of 499 cases that confirmed radiographic knee OA and 458 age and sex matched healthy controls were enrolled. There was no significant difference in demographic characteristics between the cases and controls. The genotype distribution for rs1470527 and rs9382564 SNP was significantly different between the cases and controls ($p<0.0001$). Within both the SNPs of BMP5 gene, genotype CT and TT was found to be significantly ($p<0.0001$) associated with knee OA as compared with the CC genotype. In addition when alleles were compared, T allele of both the studied SNP was observed to be significantly associated with knee OA ($p<0.0001$). Further in relation with clinical symptom of OA, we observed signification association of TT genotype with both visual analogue scale (VAS) ($p<0.0001$) and WOMAC (Western Ontario and McMaster Universities) score ($p<0.05$). **Conclusion:** Our results indicate significant association of rs1470527 and rs9382564 polymorphisms of BMP5 with knee OA. If these data are supported and confirmed, it has major implications for the pathogenesis of OA and for identifying patients at risk for developing knee OA.

2057T

Are signs of obesity and hyperphagia still relevant for the clinical suspicion of Prader-Willi Syndrome? Evidence from a novel quantitative clinical diagnostic model. *L. Cordeiro[1], C. Rocha[2], C. Paiva[3].* 1) School of Medicine and Surgery, Federal University of Rio de Janeiro State, Rio de Janeiro, Brazil; 2) Neurology Graduate Program, Federal University of Rio de Janeiro State, Rio de Janeiro, Brazil; 3) Genetics and Molecular Biology Dept, Federal University of Rio de Janeiro State, Rio de Janeiro, Brazil.

Prader-Willi Syndrome (PWS) is a rare genetic disorder that results from the lack of paternal expression of the 15q11–q13 region. The definitive diagnosis is made by molecular testing, access to which may be limited by cost or geographical constraints. Though clinical diagnostic and screening criteria have been proposed, they were based on consensus of specialists without statistical validation. There are no clinical criteria developed with quantitative or statistical methods. This study bridges this gap by developing a clinical diagnostic/screening model for PWS with data mining and machine learning techniques. Evolutionary algorithms are a distinct class of optimization algorithms that try and mimic biological phenomena. They are typically employed when traditional methods fail, due to their ability of handling complex problems. We assessed the performance of several evolutionary algorithms in the maximization of the AUC of the ROC curve of a combination of 18 clinical signs and symptoms commonly observed in PWS patients. The modelscan be easily implemented in a mobile application, allowing them to be used by clinicians in the ambulatory setting. Our sample consisted of patients from the genetics clinic of a university hospital in Rio de Janeiro with clinical suspicion of PWS. All participants underwent proper molecular testing as per current ASHG guidelines. Confidence intervals for the AUC's were estimated by bootstrapping. The classification performance of all models derived from the evolutionary algorithms was statistically superior to the one from the current clinical model for PWS. The clinical diagnostic model achieved sensitivity of 87% and specificity of 93%, while our screening model reached 100% sensitivity with 71% specificity. Based on these results, we identified six clinical signs and symptoms – hypotonia, poor sucking, hypogonadism, global developmental delay, behavioral problems and typical facies – that seem to be key to the accurate clinical suspicion of PWS. Moreover, two major clinical signs and symptoms that were previously deemed important in the clinical diagnosis of PWS – rapid weight gain and hyperphagia – were not relevant in our models. Since PWS is frequently associated with childhood obesity, our results suggest that the increasing incidence of obesity in the pediatric population may be incorrectly inducing the suspicion of PWS.

2058F

Exon targeted array CGH for identification of clinically relevant small sized, intragenic CNVs. *A. Patel[1], T. Gambin[1], P. Liu[1], W. Bi[1], A. Breman[1], J. Smith[1], S. Lalani[1,2,3], C. Bacino[1,2,3], A. L. Beaudet[1,2,3], J. R. Lupski[1,2,3], C. A. Shaw[1], S. W. Cheung[1], P. Stankiewicz[1].* 1) Dept Molecular & Human Gen, Baylor Col Medicine, Houston, TX; 2) Dept. of Pediatrics, Baylor College of Medicine, Houston, TX; 3) Texas Children's Hospital, Houston, TX.

Copy-number variants (CNVs) smaller than 400 kb in size are often challenging for clinical interpretation and are usually not reported in routine chromosomal microarray analysis (CMA). Currently, CNVs involving single exons remain also beyond the detection limits of whole exome sequencing (WES). In 2008, we designed and clinically implemented array CGH with exonic coverage of over 1,700 disease and candidate genes and demonstrated its efficacy for detection of pathogenic losses or gains of at least one exon as small as a few hundred base pairs in size. Recently, we have expanded our custom-designed oligo array to include greater than 4800 genes with exon coverage and 60K SNP oligos to detect regions of AOH. Weanalyzed with single exon resolution the personal genomes of39,170 patients referred for CMA. To date, we have identified non-polymorphic CNVs in 17,367 (44%) subjects, 3,522 (9. 0%) of which were interpreted as abnormal. Among CNVs involving only a single gene reported in 5,861 patients (15%), we found 132 *de novo* (91 losses and 41 gains) and 1338 inherited (764 losses and 574 gains) events. In addition to 65 known disease genes (40 AD, 10 AR, and 15 X-linked), we have identified 36 non disease associated genes disrupted by 21 *de novo* single gene deletions or encompassed in 16 *de novo* single gene duplications. The most commonly affected single genes include *CHRNA7* (13 dels/277 dups), *IMMP2L* (135 dels/1 dup), *TMLHE* (118 dels/16 dups), *RBFOX1* (53 dels/54 dups), *DMD* (68 dels/31 dups), and *NRXN1* (79 dels/10 dups). By extending this analysis to CNVs harboring 5 or less genes, we identified *de novo* deletions affecting 335 distinct genes. Comparison with 137 autism genes from the SFARI database (categories S, 1, 2, and 3) revealed 59 (43%) genes disrupted by 537 single gene deletions. Out of 141 epilepsy related genes, we detected 26 (18%) genes interrupted by 85 single gene deletions. Additionally, our exon targeted array enabled detection of several pathogenic heterozygous and homozygous single exon CNVs missed in WES analyses. In aggregate, our data further demonstrateutility of an exon-targeted oligo array CGH for detection of small sized CNV, complementing WES in clinical diagnostics, as well as its potential for identification of novel diseasegenes.

2059W

Deficiency of Interleukin-1 Receptor Antagonist (DIRA): Report of the First Indian Patient and a Novel Deletion Affecting *IL1RN*. *A. Almeida de Jesus[1], L. Malle[1], L. Oliveira Mendonca[2], F. X. Donovan[3], S. C. Chandrasekharappa[3], G. Montealegre[1], D. Chapelle[1], D. Suri[4], R. Goldbach-Mansky[1].* 1) Translational Autoinflammatory Diseases Section, NIAMS, NIH, Bethesda, MD; 2) Hospital das Clínicas da Faculdade de Medicina da Universidade de Sao Paulo (HC-FMUSP), Sao Paulo, Brazil; 3) Genomics Core, NHGRI, NIH, Bethesda, MD; 4) Department of Pediatrics, Advanced Pediatric PGIMER, Chandigarh, India.

Statement of Purpose. Deficiency of interleukin-1-receptor antagonist (DIRA) is a rare autoinflammatory disease clinically characterized by early-onset generalized pustulosis, multifocal osteomyelitis and elevation of acute-phase reactants. DIRA is caused by autosomal recessive loss-of-function mutations in *IL1RN.* Seven DIRA causing mutations have been described. We report herein a novel disease-causing deletion affecting *IL1RN* gene and a PCR assay for the screening of Indian patients with a clinical suspicion of DIRA. *Methods.* Peripheral blood genomic DNA was obtained and patient was genotyped for the detection of copy number variations (CNVs) by a SNP array technique using the HumanOmniExpressExome-8v1. 2_A Illumina BeadChip, containing 964,193 SNPs. The detected aberrations were visually evaluated using Nexus Copy Number v7. 5. Primers flanking the deletion ends were designed using Primer3Plus and the breakpoint area was amplified by PCR and sequenced by Sanger technique. *Results.* **Clinical description:** A 5 month-old Indian girl, born to healthy non-consanguineous parents presented with pain on manipulation and irritability at the 3rd week of life and developed a mild pustular rash. Bone scintigraphy suggested osteomyelitis of multiple bones, and cultures from bone biopsy material were negative. Due to a clinical suspicion of DIRA, the patient was enrolled into a NIH protocol and started on recombinant interleukin-1 receptor antagonist (anakinra). Anakinra initiation resulted in marked and sustained clinical and laboratory improvement. **Genetic analysis:** SNP array analysis of patient's genomic DNA showed loss-of-heterozygosity for all SNPs in a region of approximately 21. 4 to 23. 7kb, indicating a homozygous deletion. PCR and sequencing of the breakpoint area allowed us to identify the breakpoints of the deletion, at chr2_hg19_113,865,011 and chr2_hg19_113,887,227, confirming a homozygous 22,216bp deletion that spans the first four exons of *IL1RN* (NM_173843. 2). This deletion has not been previously reported in patients with DIRA. Using a multiplex PCR with primers for wildtype *IL1RN*, the deletion breakpoint, and *GAPDH*, we confirmed that patient's parents are both heterozygous for this deletion. *Conclusion.* We describe the first Indian patient with DIRA and a novel homozygous 22Kb deletion spanning 2 coding exons of *IL1RN*. The primers designed to detect the novel deletion will be useful to screen Indian patients with a clinical suspicion of DIRA.

2060T

How to translate next generation sequencing (NGS) from research to clinical practice? *G. Barcia[1,2], S. Hanein[2], V. Morinière[1], C. Fourrage[1,2], C. Bole[2], P. Nitschke[2], G. Royer[1], Z. Assouline[1], I. Perrault[2], A. Elmorjani[2], R. Gesny[1], C. Haudry[1], M. Henry[1], S. Dos Santos[1], M. Bernardelli[1], S. Poisson[1], M. Rio[1,2], T. Attié-Bitach[1,2], C. Antignac[1,2], S. Saunier[2], V. Cormier-Daire[2], S. Rondeau[1,2], S. Monnot[1,2], J. Steffann[1,2], A. Rotig[1,2], L. Colleaux[1,2], JM. Rozet[1,2], S. Lyonnet[1,2], JP. Bonnefont[1,2], A. Munnich[1,2].* 1) Genetics Department, Necker Hospital, Paris, France, France; 2) Imagine Institute for Genetic Diseases, Paris, France.

Purpose: Whole exome and genome sequencing (WES, WGS) are powerful approaches to decipher the genetic bases of inherited genetic disorders, but they remains costly and time consuming, and are a source of uncertain and incidental findings. For these reasons WES/WGS are less relevant in the frame of diagnostic setting and genetic counselling. The *Imagine* Institute for Genetic Diseases and the Genetics Unit of the Necker Hospital, joined forced to translate a custom targeted next generation sequencing (TNGS) strategy into routine clinical practice. **Methods:** For each core disorder, custom TNGS panels were designed including genes i) whose mutations have already been associated to human diseases in at least2 patients, ii) belonging to cellular pathways specific for the studied disease. **Results:** So far 18 custom TNGS panels have been developed, dedicated to molecular diagnosis of ciliopathies (n=1221 genes, diagnostic yield: 63%), Leber congenital amaurosis (n=55; 57%), intellectual disability (n=253, 23%), mitochondrial diseases (n=215, 40%), Alport syndrome (n=3, 84%), chondrodysplasia and growth delay (n=70, 50%), osteogenesis imperfecta (n=20, 60%), genetic deafness (n=35, 15%), Cornelia de Lange syndrome (n=5, 40%), focal glomerular sclerosis (n=33, 15%), polycystic kidney (n=2, 50%), neural crest disorders (n=66), childhood epilepsy (n=151), cerebellar malformations and microcephaly (n=55), skin diseases (n=90), corpus callosum defects (n=423), Hirshprung disease (n=15), cortical brain malformations (n=55). An overall diagnostic yield (definitely- or likely-pathogenic variants) of 50% was achieved (range:15-85%). Owing to high horizontal and vertical coverage (depth coverage at 30X>99%), this NGS strategy enabled the detection of several disease-causing variants that initially escaped WES. No significant incidental finding was identified. Exonic copy-number analysis using TNGS data was validated. **Conclusions:** TNGS is a robust strategy for the first-step screening of genetic disorders in a diagnostic setting since it speeds up molecular diagnosis avoiding incidental findings and at a lower cost as compared to WES. A dynamic and reciprocal interaction between physicians and molecular biologists is necessary to develop TNGS panel, to select patients to screen, to interpret TNGS data and to draw genotype-phenotype correlations. TNGS offers a rational selection of patients available for the identification of novel genes by WES/WGS in the research setting.

2061F

Secondary Finding Preferences in Whole Genome Sequencing: experiences with a large developmental delay cohort. *K. East[1], K. Brothers[2], K. Bowling[1], M. Bebin[3,4], E. Lose[3,4], S. Simmons[3,4], W. Kelley[1], G. Barsh[1], R. Myers[1], G. Cooper[1].* 1) HudsonAlpha Institute for Biotechnology, Huntsville, AL; 2) University of Louisville, Louisville, KY; 3) University of Alabama at Birmingham, Birmingham, AL; 4) North Alabama Childrens Specialists, Huntsville, AL.

Developmental delay, intellectual disability, and related congenital defects (DD/ID) affect 1-2% of children born and have a large impact on affected individuals and their families. Though many cases are attributed to an underlying genetic cause, identification of the specific variation often proves elusive. As part of the Clinical Sequencing Exploratory Research Consortium, we are conducting whole exome/genome sequencing in 450 affected probands and their parents to establish a genetic diagnosis for DD/ID and better understand the utility and impact of large scale sequencing in a pediatric neurology clinic. To date, we have enrolled 210 families, and have completed sequencing and analysis for 123. We have successfully identified DD/ID pathogenic or likely pathogenic variants in approximately 25% of our probands and have identified potentially pathogenic secondary findings in approximately 10% of parent participants. Participants are randomized to provide preferences for secondary finding disclosure either at time of study enrollment or at result return. Preferences are gathered through a novel instrument developed to illicit preferences on the types of risk information an individual would like to receive with categories defined by disease type (diabetes, cancer, alzheimer's disease). When pathogenic variants are identified, these preferences are used to determine whether a result is returned. Analysis of preferences solicited to date show that the overwhelming majority (>80%) of parents in this study elect to receive all types of secondary findings. However a minority of parents elects not to receive one or more categories of results. The categories declined vary with the most common category declined being risk of developing obesity. Reasons for decline include a lack of interest, perception of low risk status, or anxiety surrounding a specific category or type of disease. Reported secondary findings to date include genetic variation leading to an increased risk for malignant hyperthermia, cancer, and heart disease as well as carrier status for select recessive diseases.

2062W

ClinGen Actionability Working Group: Clinical Actionability in the context of Secondary Findings in Adults and Application to the ACMG 56. *J. E. Hunter[1], L. G. Biesecker[2], A. H. Buchanan[3], S. A. Irving[1], K. Lee[4], C. L. Martin[5], L. Milko[4], A. Niehaus[6], R. Nussbaum[7], J. O'Daniel[4], M. A. Piper[1], E. M. Ramos[6], S. Schully[8], A. F. Scott[9], A. Slavotinek[10], A. Sobreira[9], N. T. Strande[4], M. Weaver[11], E. M. Webber[1], M. S. Williams[3], J. S. Berg[4], J. P. Evans[4], K. A. B. Goddard[1], ClinGen Resource.* 1) Center for Health Research, Kaiser Permanente Northwest, Portland, Oregon; 2) Genetic Disease Research Branch, National Human Genome Research Institute (NHGRI), National Institutes of Health, Bethesda, Maryland; 3) Genomic Medicine Institute, Geisinger Health System, Danville, Pennsylvania; 4) Department of Genetics, University of North Carolina, Chapel Hill, North Carolina; 5) Autism & Developmental Medicine Institute, Geisinger Health System, Danville, Pennsylvania; 6) Division of Genomic Medicine, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland; 7) Department of Medicine and Division of Medical Genetics, University of California, San Francisco, California; 8) Epidemiology and Genomics Research Program, National Cancer Institute, National Institutes of Health, Bethesda, Maryland; 9) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland; 10) Department of Pediatrics, University of California, San Francisco, California; 11) American College of Medical Genetics and Genomics, Bethesda, Maryland.

Genome-scale sequencing provides the opportunity to discover "secondary findings", or variants unrelated to the primary indication for sequencing. Identification of pathogenic variants associated with an increased risk of a significant but preventable medical outcome allows the opportunity to intervene and potentially mitigate or prevent future clinical manifestations in patients and their family members. The Clinical Genome Resource (ClinGen) is a consortium working to build an open-access and centralized resource to assess the clinical relevance of genomic variants. The ClinGen Actionability Working Group (AWG) aims to identify genetic disorders with potential for clinical intervention when pathogenic variants are reported as secondary findings in previously undiagnosed adults. The AWG has developed a standardized protocol to identify and synthesize available evidence of clinical actionability for gene-disorder pairs. The AWG appliesa semi-quantitative scoring metric to the available evidenceto assess four domains of clinical actionability: 1) nature of the health threat; 2) likelihood the threat will materialize; 3) effectiveness of the intervention to prevent harm; and 4) nature of the intervention in terms of risk/burden to the individual. To date, the AWG has applied this protocol to the 56 genes and associated disorders recommended by the American College of Medical Genetics and Genomics (ACMG) to return as secondary findings. The clinical actionability scores for the ACMG 56 will be presented, along with relevant case examples illustrating the complexity of applying a standardized framework for evidence synthesis and scoring of actionability. The assessment of secondary findings in adults does not fully encompass the spectrum of actionable information from the human genome. Thus, future directions for the ClinGen AWG may include broadening the methods to address additional clinical scenarios such as variants actionable in childhood, somatic variation, pharmacogenomics applications, ending the diagnostic odyssey, and reproductive decision making. The ClinGen AWG framework will provide a structure to enable the research and clinical communities to make clear, streamlined, and consistent determinations of clinical actionability based upon transparent criteria to guide analysis and reporting of genomic variation. *ClinGen is funded by NHGRI, NICHD and NCI: 1U41HG006834-01A1, 1U01HG007437-01, 1U01HG007436-01, HHSN261200800001E.* .

2063T

Implementation of Genetic Sequencing into Clinical Practice: The Personalized Diabetes Medicine Program. *J. W. Kleinberger[1], T. J. Mathias[1], N. Ambulos[1], L. J. B. Jeng[1], D. Sewell[1], T. D. O'Connor[1], K. Tanner[1], C. M. Damcott[1], Y. Mohtasebi[1], K. A. Maloney[1], M. J. Nicholson[1], K. Palmer[1], S. A. Stein[1], K. D. Silver[1], R. B. Horenstein[1], E. A. Streeten[1], A. R. Shuldiner[1,2], T. I. Pollin[1].* 1) School of Medicine, University of Maryland, Baltimore, MD; 2) Regeneron Genetics Center, Regeneron, Tarrytown, NY.

Approximately 1-5% of diabetes mellitus, or at least 300,000 cases in the U. S. , is not polygenic but monogenic, most commonly maturity-onset diabetes of the young (MODY), but also neonatal diabetes mellitus (NDM) and syndromic forms. Monogenic diabetes is often misdiagnosed as type 1 or type 2 diabetes due to factors such as clinical overlap or lack of awareness by providers. Correct diagnosis is crucial because some forms of MODY and NDM are treated effectively with inexpensive oral sulfonylureas rather than insulin injections, while *GCK*-MODY does not usually require pharmacologic treatment. The Personalized Diabetes Medicine Program (PDMP) is a genomic medicine demonstration project funded through the NHGRI IGNITE (Implementing Genomics in Practice) Network to implement and evaluate a sustainable model for screening, diagnosis, and individualized treatment of patients with monogenic diabetes in diverse settings including an academic medical center (University of Maryland Center for Diabetes and Endocrinology), Baltimore Veterans Administration Medical Center, an integrated healthcare system (Geisinger), and a private practice (Bay West Endocrinology). Patients are identified by provider referral and by screening using a simple 7-item questionnaire followed by routine bloodwork and family/medical history collection. Eligible patients undergo next-generation sequencing of exonic/flanking regions of 40 genes. To date, 12 referred patients fulfilling eligibility criteria have undergone sequencing. Individuals carried a mean of 155±26 (SD) variants. Protein-altering variants with <5% population minor allele frequency (mean 3. 8±2. 9 per person) were assessed for pathogenicity using 2015 ACMG standards and guidelines. Pathogenic or likely pathogenic variants, including three mutations in two actionable MODY genes (*HNF1A* and *GCK*), are currently being confirmed through Sanger validation in our CAP/CLIA-certified Translational Genomics Laboratory before returning the results and interpretation to the clinician and patient through the electronic health record to guide diagnosis and therapy. An additional 11 samples (6 referrals and 5 identified by screening 272 patients) are being sequenced. This project, which includes engagement of a payer advisory panel, will generate an evidence base for the value of including an actionable highly penetrant genetic cause in the differential diagnosis that will inform standards of care and be applicable to a variety of diseases.

2064F

The diagnosis of neuronal ceroid lipofuscinosis by next generation sequencing eye and seizure panels. *E. C. Lisi[1], R. Sanchez[1], S. Richards[1], J. J. Alexander[2], L. Bean[2], C. Collins[2], M. J. Gambello[1], S. Shankar[1].* 1) Division of Medical Genetics, Dept of Human Genetics, Emory University, Atlanta, GA; 2) Emory Genetics Laboratory, Atlanta, GA.

Next-generation sequencing has revolutionized the practice of medical genetics. Multi-gene panels have provided diagnoses in patients with non-classical presentations for many disorders who would have otherwise likely gone undiagnosed. We report on a series of patients diagnosed with varying types of Neuronal Ceroid Lipofuscinosis (NCL) by identifying pathogenic variants via the Eye or Seizure Disorders Panels at Emory Genetics Lab. Patient 1 is a 5 year old female with a history of speech delay, apraxia and seizures with onset at 3. 5 years of age. A seizure panel revealed two pathogenic variants in *TPP1* (c. 509-1G>C and c. 1016G>A), and subsequent TPP1 activity in leukocytes was deficient, confirming the diagnosis of NCL2. Ophthalmology evaluation was normal. Patient 2 is a 15 year old female evaluated for a two year history of central vision loss, bilateral cystoid macular edema, and childhood onset seizures. The eye disorders panel identified a deletion encompassing the entire *CLN8* gene as well as a c. 200C>T (p. A67V) variant of uncertain significance (VOUS) in *CLN8*. Electron microscopy of lymphocytes showed inclusion bodies, supporting a diagnosis of late infantile NCL. Her brother, Patient 3, is a 22 year old with retinitis pigmentosa and seizures who was found to have the same genotype. Patient 4 is a 4. 5 year old male (consanguineous parents) with a history of refractory epilepsy, profound DD, global hypotonia, and microcephaly. The seizure panel revealed a homozygous c. 523C>T VOUS in *TPP1*. TPP1 enzyme activity in leukocytes was deficient, confirming the diagnosis of NCL2. NCL is a heterogeneous group of disorders; there are currently 13 disease loci. Given the phenotypic variability of NCL, we suggest that seizure or eye disorders gene panels should be considered for any patient with a combination of progressive vision loss, seizures, and/or regression.

2065W

Genetic Diagnosis for Deafness in a Large Clinical Cohort. *C. M. Sloan-Heggen[1,2], A. O. Bierer[1,], A. E. Shearer[1], H. Azaiez[1], E. A. Black-Ziegelbein[1], K. T. Booth[1], K. L. Frees[1], D. L. Kolbe[1,3], C. J. Nishimura[1,3], R. J. H. Smith[1,2,3,4].* 1) Molecular Otolaryngology & Renal Research Labs, Department of Otolaryngology-Head and Neck Surgery, University of Iowa, Iowa City, IA; 2) Department of Molecular Physiology & Biophysics, University of Iowa, Iowa City, IA; 3) Iowa Institute of Human Genetics, University of Iowa, Iowa City, IA; 4) Interdepartmental PhD Program in Genetics, University of Iowa, Iowa City, IA.

Deafness is the most common sensory deficit in humans, affecting 360 million people worldwide. Historically, the extreme heterogeneity of hereditary hearing loss made genetic diagnosis difficult and time consuming. With the advent of Targeted Genomic Enrichment (TGE) and Massively Parallel Sequencing, comprehensive testing for all hearing loss genes has become the most practical method of genetic diagnosing. However, with the rapidly changing knowledge of non-syndromic hearing loss and its mimics, platform maintenance is essential for sustaining the most comprehensive diagnostic testing available. Since January of 2012 over 1,500 unique probands have been sequenced through our clinical diagnostic services using a custom-designed TGE and analysis pipeline covering all known non-syndromic hearing loss genes and with the ability to cover both single nucleotide and copy number variations for a comprehensive hearing loss diagnosis resulting in an overall diagnostic rate of 40%. Each patient's clinical and genetic results were discussed at an interdisciplinary meeting to produce the most inclusive and integrated diagnosis possible. Evaluating such a cohort of all-comers allows several unique analyses. It shows that diagnostic rate is heavily impacted by inheritance, laterality, age of onset, syndromic features, and ethnicity of the patient. Individuals with bilateral, symmetric hearing loss have an increased diagnostic success of 44%, compared to only 1% of patients with unilateral hearing loss. Autosomal dominant hearing loss can be diagnosed 50% of the time, autosomal recessive—41%, and sporadic—37%. Non-syndromic hearing loss was diagnosed in 42% of cases, compared to 27% of syndromic hearing loss. Individuals of a Middle Eastern ethnicity have a high diagnostic success of 72%, versus only 25% in African Americans. Pre-evaluation of a patient's clinical characteristics can recommend use of such a panel for particular cohorts of individuals and allow for pretest counseling to guide patient expectations. Positive diagnosis can preclude additional unneeded testing which would otherwise be used to rule out additional syndromic features.

2066T

Detection of 3 novel mutations in the *GAA* gene in 3 iranian families with glycogen storage disease type II. *S. Tabei, S. Kohan Mozaffari, M. Fardaei.* Medical Genetics, Shiraz University of Medical Sciences, Shiraz, Fars, Iran.

Introduction: Glycogen storage disease type II or Pompe disease is the first recognized lysosomal storage disorder in which an extreme amount of glycogen accumulates in lysosomes of almost all types of tissues especially skeletal muscle and heart due to mutation in acid α glucosidase (*GAA*) gene. Pompe disease comes at different ages and with different degrees of severity. The severity of the clinical presentations, the tissue involvement, and the age of onset generally correlate with the nature of the mutation and the degree of residual enzyme activity. In this study, mutation analysis of the *GAA* gene in 7 families suspected to have Pompe disease was carried out. **Methods**: In order for mutation analysis in these 7 families PCR, Sanger sequencing, ARMS-PCR and RFLP-PCR techniques were used. **Results:** Among these 7 studied families, we determined 3 families with 4 mutations, in which 3 on them include NM_000152. 3:c. 1004G>A, NM_000152. 3:c. 2078dupA and NM_000152. 3:c. 896T>C didn't submit in any database and we successfully submit them in ClinVar with RCV000156939 ,RCV000156940, RCV000156941 ID and in SNPs databases with rs730880022, rs730880372 and rs121907940 ID respectively. In 1 family submitted mutation (NM_000152. 3:c. 1927G>A) with ClinVar and SNPs ID RCV000004238 and rs28937909 was found. **Conclusion:** Despite the importance of measuring glucose oligosaccharides or acid maltase enzyme concentration in urine, plasma, and blood spot as a diagnostic procedure mutation analysis is the best way to confirm the clinical suspicion for Pompe patients and provide valuable information for family screening and genetic counseling.

2067F

Pilot data from the Veritas Genetics myGenome Project. *J. V. Thakuria[1, 2, 3], D. Martinez[1], B. B. Simen[1], H. Fortin[1], M. Callahan[1], P. Estep[1], H. Russell[1], G. T. Berry[4], G. M. Church[1,3].* 1) Veritas Genetics, Somerville, MA; 2) Massachusetts General Hospital, Boston, MA 02115; 3) Harvard Medical School, Boston, MA 02115; 4) Boston Children's Hospital, Boston, MA 02115.

While the number of published reports demonstrating medical utility of exome sequencing in patients suspected of monogenic disease continues to grow, data are still limited on the medical utility of whole genome sequencing in healthy, asymptomatic individuals hindering broader access for the general population. Outdated ACMG recommendations limiting routine return of results to a panel of only 56 genes highlights the growing need for population-based data. Initial reports on the number needed to screen (NNS) in the general population before a suspected pathogenic mutation is identified varies widely from <1% to >20%, but is usually reported in the 1. 5%-5% range (excluding potential reproductive utility from knowledge of carrier status for recessive disorders) - and is highly dependent on variant filtering stringencies. Preliminary pilot phase data from over 50 whole genomes with >100X average coverage from the Veritas Genetics myGenome Project parsed against over 20,000 Mendelian genes, thousands of positive controls (including >350 deeply phenotyped Personal Genome Project (PGP) samples with whole genome data), and undergoing secondary individualized analyses based on voluntarily provided personal and family medical histories, is reported. The pilot cohort was recruited across 5 continents. Study consent, pre- and post-test genomic counseling, as well as return of results, are generally performed remotely via tele- or videoconferencing. All study participants are given the option to receive an easily portable, compressed copy of their whole genome data on thumbnail drive and may opt in for indefinite biobanking of their samples for ongoing research and clinically focused reanalysis of data on a yearly basis on average and guided by evolving medical genomics knowledge. Alternatively, participants may opt to have their samples discarded and backup data deleted after receiving their compressed data and initial medical interpretation of results. The Veritas Genetics myGenome Project represents a viable internationally executed workflow for high throughput, high quality, clinical grade delivery of whole genome sequence and interpretation at population-scale.

2068W

Diabetic Ketoacidosis in Vanishing White Matter. *H. Alamri[1], F. Mutairi[1], A. Alothaim[1], M. Alfadhel[1], A. Alfares[1,2].* 1) king Abdul Aziz Medical City for National Guard, Riyadh, Saudi Arabia; 2) Qassim University, Qassim, Saudi Arabia.

Introduction: Vanishing white matter disease (VWM) is an autosomal recessive chronic and progressive white matter disorder. The symptoms of the disease can begin shortly after birth or in adulthood, and the most common features include neurological deterioration, optic atrophy, seizures, coma, and death. However, hyperglycemia and diabetic ketoacidosis (DKA) have never been reported before in patients with VWM. Case presentation: A male infant, product of a full-term uncomplicated pregnancy with developmental delay, hypotonia, and seizures. MRI of the brain revealed features of white matter diseases. At the age of 8 months, he presented to the emergency department with irritability and was found to have a very high blood glucose (25 mmol/l; normal, 6-10 mmol), HbA1c (11. 4%; normal 4%-5. 6%) levels, as well as ketosis and metabolic acidosis (pH 7. 2). Consequently, he was diagnosed and treated as a case of DKA and discharged on insulin injections. However, further investigations for type 1 and 2 diabetes mellitus, genes associated with neonatal and infantile diabetic syndromes and muscle biopsy for mitochondrial disorders were all negative. Eventually the child died at the age of 2 years while he was on insulin injections due to sever hypotonia and recurrent chest infections. The patient is product of a double cousin marriage, two of his brothers had died at 1 year of age with similar neurological, diabetic and brain MRI manifestations. Whole Exome Sequencing (WES) revealed a homozygous pathogenic variant in *EIF2B1* c. 146T>G p. (Leu49Arg) in the proband and both parents were heterozygous. Discussion: The patient met the consensus statement from the International Society for Pediatric and Adolescent Diabetes (ISPAD) in 2014 for the biochemical diagnostic criteria for DKA. WES revealed the diagnosis of VWM, which is a heterogeneous disease with five genes involved. The eIF2B protein complex plays a critical role in the initiation of protein synthesis and the cellular response to stress by activating the initiation factor eIF2, which interacts with the glucose response stimulus. However, the relation between the eIF2B protein complex and glucose level is unclear and requires further research. Conclusion: We are reporting on a case with both VWM disease and DKA. To the best of our knowledge, this is the first case of a similar presentation, and yields additional observations regarding the role of *EIF2B1* in glucose regulation and expands the phenotype of VWM disease.

2069T

Clinical Rapid Whole Exome Sequencing Experience. *D. Copenheaver[1], J. Juusola[1], K. Retterer[1], P. Vitazka[1], F. Gibellini[1], A. Marth[1], E. Butler[1], M. Deardorff[2], E. Zackai[2], E. Bedoukian[2], A. Wilkens[2], L. Medne[2], A. Lehman[3], D. Niyazov[4], J. Graham[5], Z. Y. Kucuk[6], G. Nalepa[7], C. Griffin[7], M. Willing[8], J. Bernat[9], D. Vats[10], L. Massingham[11], A. Poduri[12], W. Chung[13], G. Richard[1], S. Bale[1].* 1) GeneDx, Gaithersburg, MD; 2) Children's Hospital of Philadelphia, Philadelphia, PA; 3) University of British Columbia, Vancouver, BC; 4) Ochsner Clinic, New Orleans, LA; 5) Cedars Sinai Medical Center, Los Angeles, CA; 6) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 7) Riley Hospital for Children, Indianapolis, Indiana; 8) Washington University in St. Louis, St. Louis, MO; 9) University of Michigan, Ann Arbor, MI; 10) Kaiser Permanente- Southern California, Los Angeles, CA; 11) Rhode Island Hospital, Providence, RI; 12) Boston Children's Hospital, Boston, MA; 13) Columbia, New York, NY.

Whole exome sequencing (WES) is increasingly being used to identify the underlying cause of genetic disease and has a reported diagnostic yield of 25-30%. Recently the use of WES has been reported for undiagnosed patients in the NICU or PICU in whom a rapid genetic diagnosis may significantly alter medical management. To determine the feasibility, yield, and usefulness of WES with a fast turn-around time, we report on results of the first 19 consecutive cases submitted to our clinical laboratory for exome sequencing providing results within 7 days or less. Express WES was performed on proband-parent trios. WES analysis was completed and in all cases a verbal result with presence or absence of diagnostic findings in known disease (HGMD) genes was provided within 3-7 days. Confirmation of all reportable variants and secondary findings was performed, and a written report was issued within 5-13 days from start of testing. A definitive diagnosis was made in 8 of the cases (42%). In one case, two distinct genetic diagnoses were made. Of the positive cases, four (44%) probands were diagnosed with an autosomal recessive disorder, three probands (33%) had de novo and two (22%) had inherited variants associated with autosomal dominant disorders, and one (11%) proband had an X-linked disorder. No reportable ACMG secondary findings were identified in this cohort. The probands' ages ranged from 5 days to 14 years. Eight were infants <11 months, and 13 were <2 years old. The majority of probands were referred for neurological indications (13/19) including seizures (7/19), two children had multiple congenital anomalies, and the remainder had other clinical indications. In three cases, express WES was chosen to aid with decision making regarding a stem cell transplant, lung transplant, or family planning of an on-going pregnancy. Other benefits of a rapid diagnosis were changes in medical management (n=7), the ability to make prognostic predictions and discuss palliative care with the family (n=4), to facilitate carrier testing and informed family planning, and to avoid further invasive or other diagnostic tests (n=9). Although our experience is still limited, our data demonstrate that express WES testing can be accomplished in a diagnostic laboratory setting, that this testing has good diagnostic yield of 42% and, if positive, provides value for determining prognosis, making management and family decisions much sooner than with traditional genetic testing.

2070F

Next-Generation Sequencing test within a neurologic region of interest leads to diagnosis of *RYR1*-related disorder for 36-year-old female after three decades. *P. Gerrol[1], N. Vena[1], A. Rutkowski[2, 3].* 1) Claritas Genomics, Inc, Cambridge, MA; 2) Cure Congenital Muscular Dystrophy, cureCMD. org; 3) Kaiser SCPMG, Los Angeles, CA.

We report on a 36yo female with a progressive neuromuscular disorder who has remained without a definitive molecular diagnosis, thus lacking the benefit of prognostic and familial risk information. She initially presented to Neurology at age 3 with motor delay and lumbar lordosis. Exam showed facial, neck flexor, and proximal weakness and use of a Gower maneuver. After a relatively stable disease course in childhood, there was gradual progression of muscle weakness, proximal greater than distal, in upper and lower extremities. She developed ophthalmoplegia (bilateral restricted upward gaze) and facial diplegia in the 2nd decade and is now wheelchair-dependent. Muscle biopsy showed type 1 fiber predominance and centronuclear myopathy but no cores/ultrastructural changes; however, *MTM1* (MIM 300415) sequencing and deletion/duplication studies were negative. We performed the Claritas Pediatric Neurology Exome, which employs two next-generation sequencing (NGS) technologies to detect and orthogonally confirm variants in 614 neurology-related genes. This approach focuses on genes with known clinical relevance and reduces the number of variants requiring confirmation, thus reducing turnaround time. NGS identified two variants in *RYR1* (MIM 180901): a novel frameshift variant and a previously reported missense variant. The maternally-inherited c. 14662_14663delAT (p. Met4888Valfs*20) frameshift variant predicted to result in a premature stop codon was interpreted as likely pathogenic. The c. 10622C>T (p. Ala3541Val) missense variant had been previously reported, but inconclusive evidence was provided to determine pathogenicity; it was classified as a variant of uncertain significance. The majority of pathogenic *RYR1* variants are private heterozygous missense changes leading to autosomal dominant central core disease and/or increased susceptibility to malignant hyperthermia. In contrast, individuals with congenital centronuclear myopathy tend to be compound heterozygotes for *RYR1* missense and truncating variants. Asymptomatic heterozygous carriers of the latter have been reported, supporting autosomal recessive inheritance. The molecular results identified, suggestive of autosomal recessive *RYR1*-centronuclear myopathy, are consistent with the patient's clinical phenotype, thus providing her with a molecular diagnosis after 30 years of uncertainty.

2071W

Clinical results from a pediatric neurological region of interest using an orthogonal NGS approach to identifying variants. *E. J. White, N. Bachman, S. Gruszka, S. Hansen, V. Lip, P. Milos, N. Napolitano, T. Restrepo, Y. Shen, N. Vena.* Claritas Genomics, Inc. , Cambridge, MA.

Statement of Purpose: There are many pediatric neurologic disorders that present with unique or overlapping phenotypes that often make identifying the underlying genetic variant a multi-year journey when using traditional molecular approaches to identify them. Through the use of next-generation sequencing (NGS) technologies, it is possible to sequence all the protein coding exons in the affected patient's genome simultaneously to quickly identify potential causative variants. In a clinical setting, it remains necessary to verify putative NGS results with additional molecular methods. In the case of whole exome sequencing, the number of variants needing confirmation may be in the hundreds. Methods: At Claritas Genomics, we have developed an approach to clinical testing which limits the number of potential variants needing Sanger confirmation. This approach both focuses on a neurological region of interest (ROI) encompassing 614 genes that have been implicated in causing seven pediatric disorder categories as well as employing two NGS technologies to orthogonally detect and confirm variants in the neurological genes. Any variants that are not confirmed using this approach can be resolved by Sanger sequencing. Results: We have compared results obtained using our assay with results from another provider's whole exome assay for 9 independent samples chosen for their pediatric neurological indications. These analyses were carried out in a blinded fashion. Of the nine samples, four samples produced clinically inconclusive findings in both assays. Three samples produced the same results in both assays. One sample produced a result that was interpreted as clinically relevant in the other provider's analysis but our analysis indicated this gene was not phenotypically relevant to the patient's disorder. The final sample produced a result that was clinically relevant based on our analysis but it was not identified in the other provider's whole exome. The use of orthogonal sequencing platforms provides immediate confirmation of most variants and also yields better sensitivity than either platform alone can generate, providing better results and more timely information for patients seeking answers to their diagnostic questions.

2072T

***De novo* mutations in prenatal trios with abnormal ultrasound findings detected by high coverage NGS panel analysis.** *H. Dai[1], H. H. C Lee[1], J. Zhang[1,2], H. Mei[1,2], S. Chen[1], Y. Feng[1], J. Wang[1,2], V. W. Zhang[1,2], LJ. Wong[1,2].* 1) Baylor Miraca Genetics Laboratories. Houston, TX; 2) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Introduction: *De novo* mutations are the major genetic cause for sporadic autosomal dominant inherited diseases. Trio analysis can provide prompt molecular diagnosis for diseases with fetal abnormalities. Overall *de novo* mutation rate has been evaluated at the genome/exome level in normal population. High coverage capture-based Next Generation Sequencing (NGS) panel trio analysis now is available to interrogate de novo mutations as one time and cost efficient method. **Methods**: Fifty trio samples submitted to us for molecular analysis of the abnormal fetus suspected of Noonan syndrome by target capture NGS using a custom designed capture library. Variants from 183 genes involved broadly in varies biological processes, including the Ras signaling pathway, were analyzed. The average depth of coverage is >600X, sufficient for the detection of low level mosaicism. *De novo* changes were further confirmed by Sanger sequencing in prenatal and parental samples. **Results:** A total of seven *de novo* changes were identified from 50 trios. Corrected with capture library size ~1Mbp, our data suggest an expected *de novo* mutation rate of 438/genome, 4/exome and 1 in 1255 protein coding genes, which is significantly higher than reported previously, presumably due to the biased patient population of abnormal fetal ultrasound suggesting Noonan spectrum disorder. Four missense changes are categorized as pathogenic/likely pathogenic for Noonan syndrome: *RIT1*:c. 170C>G (p. Ala57Gly); *RIT1:* c. 268A>G (p. Met90Val); *PTPN11*:c. 215C>T (p. Ala72Val); *SOS1:c. 508A*>G (p. Lys170Glu). Three *de novo* variants are novel intronic/UTR changes with unknown significance in genes not directly associated with fetal abnormalities (*HBD*: c. -61G>A; *GATA2*:c. *1574A>G; *FH*:c. 1236+28T>G). The allelic ratio of those changes range from 49% to 53%, suggesting pre-zygotic events, rather than somatic mosaicism. No low level mosaic changes in the blood samples of parents were detected with the deep NGS approach. **Conclusions:** Capture based trio NGS analysis provides accurate and prompt assessment for *de novo* mutations in cases with prenatal ultrasound abnormalities. High coverage NGS data also offers the ability power to detect possible mosaicism. Significantly higher *de novo* mutation rate might be expected due to the dominant phenotype and bias of sample selection. In addition to missense changes that are associated with pathogenesis, changes in introns or URT regions are of interest for further investigation.

2073F

Rare Genomics Institute research initiative provides non-profit patient services enabling deep insights into rare disease conditions through genome sequencing and cloud collaboration. *C. J. Lin, R. R. Haraksingh.* Rare Genomics Institute, Hanover, MD. , U. S. A.

The Rare Genomics Institute (RG) is an international non-profit which connects rare disease patients to cutting-edge genomic solutions for diagnosis and treatment. RG has partnered with 8 sequencing sites to provide our patients with whole exome sequencing (WES). An estimated 75% of patients who undergo WES do not receive a genetic diagnosis due to the rarity of their condition and our incomplete understanding of the functional significance of genomic variants. The RG research initiative addresses the need for combined in-depth analysis of the genetic and medical data for each patient which is not provided by physicians, genetic counselors, or sequencing services. Our research model involves crowd-sourcing ideas from our global network of scientists, cloud collaboration, adoption of state of the art computational systems, and utilizing the unique expertise of our in house scientific team to provide actionable insights into each case. Specifically, patient medical and genetic data is stored and shared via a partnership with online medical record system Patients Know Best. In-depth genetic analysis informed by comprehensive and well-defined medical phenotypes for each patient is carried out by our in house team of PhD-trained volunteer analysts using Omicia Opal variant filtering software. These analyses along with an extensive compilation of the literature relating to symptoms, genes of interest, treatments, or similar cases as well as opportunities to join relevant clinical trials and connect with external experts, are assimilated into scientific reports including recommended next steps. These reports are returned to the patients and published on our online forum for further research and collaboration. To date, we have enrolled ~20 families in our pilot program and completed analysis for 12. Of these patients, we found genetic diagnoses for 6, defined the syndrome for 2, connected several to external research experts for follow up studies, recommended clinical trials for a few, and a second family was found for one case. In all completed cases the families received satisfactory biological insight into their condition. Our research model is helping to meet a crucial need for rare disease patients and demonstrates that exome sequencing coupled with in depth analysis leveraging medical history and global collaboration can be deeply insightful. Initial results show that our approach can increase diagnostic yield for rare diseases from 25 to 50%.

2074W

Diagnosis of autosomal dominant polycystic kidney disease (ADPKD) by full gene sequencing and MLPA. *M. Procter[1], E. Paul[1], A. Morris[1], R. Mao[1,2].* 1) Res & Development, ARUP Lab, Salt Lake City, UT; 2) University of Utah Health Sciences Center, Salt Lake City, UT.

Autosomal Dominant Polycystic Kidney Disease (ADPKD) is usually an adult onset, multisystem disease. Fifty percent of individuals with ADPKD will develop end stage renal disease (ESRD) by age 60, however renal cysts may be present from birth. Clinical presentation of ADPKD can have a wide spectrum including the following symptoms: headache, high blood pressure, back or side pain, frequent urination, urinary tract or kidney infections, kidney stones, increase in the size of that abdomen as the kidneys become overgrown with cysts, end stage renal disease (ESRD), and kidney failure. Mutations in *PKD1* or *PKD2* cause abnormal or absent expression of the proteins Polycystin-1 (PKD1) and Polycystin-2 (PKD2). Altered/abnormal expression of the polycystins during development leads to dilation in the renal tubules, ultimately leading to the formation of fluid-filled cysts that are characteristic for this condition. These cysts are detected through renal imaging. Sequence variants in*PKD1* account for disease in 90% of individuals with ADPKD and variants in *PKD2* are causative in 10% of individuals with ADPKD. Individuals with *PKD2* mutations are less likely to develop ESRD than individuals with a *PKD1* mutation. *PKD1* is comprised of 46 exons on chromosome 16 and *PKD2* is comprised of 15 exons on chromosome 4. Chromosome 16 contains a region adjacent to the *PKD1* gene having 97% homology to exons 1 through 33 of *PKD1*. Sanger sequencing to avoid this pseudogene was achieved using a combination of long range PCR, nested PCR, and standard touchdown PCR for a total of 78 PCR amplifications. Deletion and duplication testing was done using MLPA with an extended denaturation time to accommodate the high G-C content of *PKD1*. Accuracy and precision of these assays were evaluated using samples from three normal individuals and two samples with known *PKD2* mutations. Samples with *PKD1* and *PKD2* deletions and duplications were unavailable so PCR amplicon was used as template for evaluation. These *PKD1* and *PKD2* mutation and deletion/duplication assays are robust, accurate, and reproducible. Genetic testing allows for diagnosis of affected individuals, testing of potential kidney donors, and testing of adult relatives of affected individuals. Early detection and treatment can reduce complications and disorders associated with ADPKD.

2075T

Review of *PRPH2 (RDS)* mutations found in patients enrolled in the National Ophthalmic Disease Genotyping and Phenotyping Network (eyeGENE®). *M. J. Reeves[1], K. E. Goetz[1], D. Blain[1], A. V. Garafalo[2], V. NDifor[2], R. S. Parrish[2], S. Vitez[2], X. Wang[1], S. J. Tumminia[3].* 1) Ophthalmic Genetics and Visual Function Branch, National Eye Institute/NIH, Bethesda, MD; 2) Astrix Technology Group, Bethesda MD; 3) Office of the Director, National Eye Institute/NIH, Bethesda, MD.

Purpose: The National Ophthalmic Disease Genotyping and Phenotyping Network, eyeGENE®, is a multicenter genomic medicine initiative that recruits individuals with inherited eye diseases from the United States and Canada for genetic testing and research into these diseases. *PRPH2*/*RDS* is a gene associated with various autosomal dominant conditions including retinitis pigmentosa, pattern dystrophy, macular dystrophy, and others. Seventy-one eyeGENE® participants of 52 familieswith a *PRPH2*/*RDS* mutation were reviewed to assess potential genotype-phenotype correlations. **Methods:** A blood sample was obtained from each participant, and genomic DNA was extracted. Clinical information submitted by the referring clinician was reviewed. A portion of the DNA was shipped to either the Laboratory for the Molecular Diagnosis of Inherited Eye Disease (LMDIED) at the University of Texas Health Science Center or GeneDx. *PRPH2*/*RDS* was screened by conventional Sanger or next generation sequencing, and the results were reported back to the eyeGENE® Coordinating Center and shared with the referring clinician and participant. **Results:** Mutations were distributed across exons 1, 2 and 3, with 39% of mutations in Exon 1, 40% in Exon 2, 4% in Exon 3, and the rest intronic. The most prevalent mutation type was an IVS2 mutation, c. 828+3A>T, which affected 9 of 52 families (17%). The c. 422A>G, p. Y141C mutation in Exon 1 was found at the second highest frequency of 4/52 (7. 7%). The c. 715C>T, p. Q239X mutation found in Exon 2 had a frequency of 3/52 (5. 8%). Patients were enrolled with a variety of diagnoses; e. g. , 20 patients as Stargardt disease, 26 as pattern dystrophy, 5 as cone rod dystrophy, 5 as retinitis pigmentosa, 3 as Best disease, 1 as choroideremia and 11 patients with 2 or 3 of these classifications. While the overall age of onset for this cohort was generally at around 40 years of age, it was difficult to determine whether or not there were true correlations between genotype and phenotype. **Conclusions:** The *PRPH2*/*RDS* gene has been shown to have high phenotypic variability, even within the same family. Because the disease presentation may be so variable and fit more than one diagnosis, it may be beneficial for eyeGENE® and others to develop algorithms and responsive diagnostic questionnaires so that the reported phenotype will be more standard between clinicians and a broader approach to genotype-phenotype correlations might be taken with larger patient datasets.

2076F

Development and implementation of a targeted next generation sequencing (TNGS) strategy for the molecular diagnosis of intellectual disability. *S. Hanein[2], C. Fourrage[1,2], G. Royer[1], A. Elmorjani[2], M. Rio[1,2], J. Amiel[1,2], V. Cormier-Daire[1,2], G. Baujat[1,2], C. Bole[2], P. Nitschke[2], J. Steffann[1,2], L. Colleaux[2], S. Lyonnet[1,2], JP. Bonnefont[1,2], A. Munnich[1,2], G. Barcia[1,2].* 1) Medical Genetics, Necker Hospital, Paris, France, France; 2) Imagine Institute for Genetic Diseases, Paris, France.

Purpose: Intellectual disability (ID) is a clinically and genetically heterogeneous disorder representing the leading cause for referral in the Genetics Department of the Necker Hospital with about 7000 patients every year. Trio whole exome sequencing (WES) is a powerful approach to gain insight into the genetic bases of ID, but it remains costly and time consuming as far as interpretation is concerned. Moreover, it can be a source of uncertain or incidental findings, and is therefore less relevant for the diagnostic setting. As a collaborative work of the *Imagine* Institute and the Clinical and Molecular Genetics Unit of the Necker Hospital, we developed a custom targeted next generation sequencing (TNGS) strategy for the molecular diagnosis of ID, including all genes whose mutations have already been reported in at least two unrelated patients. **Methods:** Genomic DNA libraries were performed using SureSelectXT Target Enrichment Reagent Kit according to the suppliers' recommendations, and were subjected to custom targeted DNA panel enrichment. 4615 regions of interest of 253 ID genes were captured with the 120-pb cRNA baits designed using the SureDesign software (Agilent). The targeted regions, encompassing 1,556 Mb of DNA, were sequenced by synthesis on an Illumina HiSeq2500, and data were analysed through the Imagine Bioinformatics core facilities. **Results:** We analysed 128 samples including 12 with known ID mutations, 14 patient-parents trios, and 74 index cases. We confirmed the presence of mutations in controls (including point mutations and indels). We identified pathogenic or probably damaging variants in 20/88 patients (23%), involving the *SATB2, STXBP1, CDKL5, MEF2C, SYN1, KATNAL2, PTPN1, SMC1A, SHANK3, NFIX, FLNA, BCOR* and *EHMT1* genes. **Conclusions:** With a diagnostic yield of 23%, TNGS is an efficient strategy for the first-step genetic screening of ID, avoiding incidental findings, and at a lower cost compared to WES. Even if targeted approaches are restricted to a few hundred genes, interpreting variants generated by TNGS is challenging. A dynamic interaction between physicians and molecular biologists is necessary to interpret TNGS results and to draw genotype-phenotype correlations. TNGS, enabling a comprehensive screening of all known genes involved in ID, offers a rapid, cost effective and relevant selection of patients available for the identification of novel genes by WES and WGS in the research setting.

2077W

Comprehensive genetic testing of 74 probands with Waardenburg syndrome using massively parallel sequencing. *j. Cheng[1], y. lu[2], h. j. yuan[1,2].* 1) medical genetics center, Southwest Hospital, Third Military Medical University, chongqing, China; 2) Dept. of Otolaryngology, Chinese PLA General Hospital, Beijing.

Waardenburg syndrome (WS) is heterogeneous and responsible for 1-3% of total congenital deafness. The combination of targeted exome capture and massive parallel sequencing (MPS) has become a promising method for detecting novel and known mutations involved in hereditary hearing loss in a rapid and cost-effective way. In this study, we tested 74 probands in China with representative Waardenburg syndrome, using the in-house platform including 438 genes associated with deafness. They were characterized by the association of congenital sensorineural hearing loss and pigmentary disturbances of the iris, hair, and skin, along with or without other features such as dystopia canthorum, musculoskeletal abnormalities of the limbs, Hirschsprung disease, or neurological defects. The results indicated that six known WS genes (*PAX3, MITF, SOX10, EDNRB, EDN3* and *SNAI2*) were responsible for about 57% (42/74) of WS cases. We identified 11 distinct *PAX3* mutations in 12 patients, of which 2 were novel; 8 distinct *MITF* mutations in 13 patients, of which 6 were novel; 13 distinct *SOX10* mutations in 13 patients, of which 11 were novel; 4 distinct novel *EDNRB* mutations in 3 patients. None of mutation was found in *EDN3* and *SNAI2*. In conclusion, 24 novel WS gene mutations were identified, which need to be confirmed; and 18 cases (about 24%) can be diagnosed surely. Further whole exome sequencing will be applied to identify the novel WS genes with the rest 43% undiagnosed WS cases.

2078T

Comprehensive molecular analysis of the *ABCA4* gene in Stargardt disease patients of the German population. *H. L. Schulz, F. Grassmann, R. Walker, H. Roth, M. Weber, H. Stöhr, B. H. F. Weber.* Institute of Human Genetics, University of Regensburg, Regensburg, Germany.

Autosomal recessive Stargardt disease (STGD1 [MIM 248200]) is the most prevalent juvenile-onset maculopathy and is caused by mutations in the ABCA4 gene. To date, more than 800 ABCA4-associated mutations have been reported to cause STGD1 and other retinal diseases. Several patients carry monoallelic ABCA4 variants in the coding or conserved intronic sequences some of which may belong to a subgroup of clinically distinct AMD/late-onset Stargardt disease patients. In addition, recent studies indicate the rare presence of deep intronic pathologic ABCA4 variants. To identify the ABCA4 mutation spectrum in German patients diagnosed with Stargardt disease, we analyzed the ABCA4 gene in 245 individuals by either Affymetrix-based custom-designed resequencing array (RetChip) analysis or NGS. A total of 142 different mutations found in this cohort were categorized as likely pathogenic or pathogenic based on allele frequency (MAF≤0. 5%), bioinformatical and functional analysis, or as variants of unknown clinical significance. This includes 39 novel mutations. In 45% of the patients, two causative mutations were identified, 24% of patients harbor only one ABCA4 mutation and no ABCA4 sequence alterations were identified in the remaining 31% of patients. To analyze the impact of deep intronic ABCA4 variants in our cohort, we sequenced selected regions of intron 30 and 36. Two different intronic variants that are predicted to strengthen cryptic donor splice sites were found in two patients with a second rare variant in the coding region of ABCA4. Finally, multivariate logistic regression of the patient cohort and ethnically-matched subjects was used to study the association of 18 polymorphic ABCA4 variants (MAF>0. 5%) with disease. This identified three statistically significant (Pval<0. 05/18) risk-increasing variants and two protective variants. From these, a genetic risk score (GRS) was computed as the sum of risk increasing and protective ABCA4 alleles weighted by the relative effect size of each SNP from the logistic regression model, revealed that patients have a significantly higher mean GRS (0. 606, 95% CI: 0. 497-0. 714) compared to controls (0. 0, 95% CI: -0. 054-0. 054). Patients that carry only one causative ABCA4 mutation have a statistically significant higher GRS than those with no or two ABCA4 mutations. This suggests that in patients with a single ABCA4 mutation, risk-increasing variants or associated changes may contribute to the disease.

2079F

First report of molecular genetics diagnosis of Niemann Pick disease in North-West of Iran. *E. Abedini[1], A. Khorrami[2], M. saeid nejhad[1], M. Ziadi[2], M. mohaddes[2].* 1) Azad university of Tabriz, Tabriz, Iran; 2) Tabriz University of Medical Science.

Niemann Pick disease (NPD) types A and B are rare autosomal recessive disorders occur by sphingomyelin phosphodiesterase1 (SMPD1) gene mutations that result in the deficiency or shortage of lysosomal acid sphingomyelinase (ASM) activity. The disease is characterized by the accumulation of sphingomyelin in the endolysosomal compartments and patients often present hepatosplenomegaly, respiratory complications, atherogenic lipid profiles and sometimes progressive neurodegeneration. In this study we report the molecular characterization of 7 unrelated Iranian Azeri Turk patients with type A and B NPD by sequencing of SMPD1 exones. For collecting more information in the case of frequency of the polymorphic hexanucleotide sequence. P. 37LA (3-8), reside within the signal peptide of ASM, we also design a control group including 50 healthy participants with the same ethnic background. L137P missense mutation, was observed in the 3 of 7 patients in homozygote status. 4 of 7 patients were homozygous for rs1050239 (c. 1522G>A; p. G508R). In the case of P. 37LA (3-8) the allele contained four hexanucleotide repeat units is the most frequent in both patients and healthy participents followed by 5, 6and 7repeats. whereas the number of repetition in ENSEMBLE and UCSC databases, both are five while in NCBI database seven repeats is considered as common allele. L137P is the only pathogenic mutation was found in this investigaton. The existence of allele with four repetations with high frequency for P. 37LA (3-8) in our population can be an example of founder effect.

2080W

***SPATA5* mutations identified by whole exome sequencing cause a novel autosomal recessive syndrome characterized by microcephaly, intellectual disability, seizures, and hearing loss.** *K. G. Monaghan[1], A. J. Tanaka[2], M. T. Cho[1], F. Millan[1], J. Juusola[1], K. Retterer[1], C. Joshi[3], D. Niyazov[4], A. Garnica[5], E. Gratz[6], M. Deardorff[7], A. Wilkins[7], X. Ortiz-Gonzalez[8], K. Mathews[3], K. Panzer[9], E. Brilstra[10], K. L. I. van Gassen[10], C. M. L. Volker-Touw[10], E. van Binsbergen[10], N. Sobreira[11], A. Hamosh[11], D. McKnight[1], W. K. Chung[2, 12].* 1) GeneDx, Gaithersburg, MD, USA; 2) Department of Pediatrics, Columbia University Medical Center, New York, NY, USA; 3) Departments of Pediatrics and Neurology, University of Iowa Children's Hospital, Iowa City, IA, USA; 4) Department of Pediatrics, Division of Medical Genetics, Ochsner Health System, New Orleans, LA, USA; 5) Arkansas Children's Hospital, Little Rock, AR, USA; 6) Gratz & Shafrir, M. D. , Child Neurology, Baltimore, MD, USA; 7) Department of Pediatrics, Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 8) Departments of Pediatrics and Neurology, Children's Hospital of Philadelphia, University of Pennsylvania School of Medicine, Philadelphia, USA; 9) Department of Pediatrics, University of Iowa Children's Hospital, Iowa City, IA, USA; 10) Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands; 11) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD, USA; 12) Department of Medicine, Columbia University Medical Center, New York, NY, USA.

We describe a novel syndrome characterized by severe neurological impairment including intellectual disability (ID), epilepsy, microcephaly, abnormal muscle tone, and sensorineural hearing loss due to autosomal recessive mutations in the Spermatogenesis-Associated Protein 5 (*SPATA5*) gene. Clinical whole exome sequencing (WES) was performed on 3,359 patients with developmental delay, ID, or seizures. Ten affected children from 7 unrelated families were found to be homozygous or compound heterozygous for *SPATA5* variants, which were suspected to cause their phenotype. Through GeneMatcher we identified 4 additional probands from 3 families. All families were sequenced by WES either as a proband-parent trio or family analysis using affected or unaffected siblings for segregation studies, which supported autosomal recessive inheritance. *SPATA5* is a ubiquitously expressed member of the ATPase associated with diverse activities (AAA)-protein family located at 4q28.1. Involved in mitochondrial morphogenesis during early spermatogenesis, it may also play a role in post-translational modification during cell differentiation in neuronal development. Pathogenic variants in *SPATA5* may affect brain development and function, resulting in the neurological phenotype described here. All 15 *SPATA5* variants we detected were novel or rare (allele frequency <1/10,000), occurring in highly evolutionarily conserved regions. Variants included missense (9), frameshift, (3) in-frame deletion (1), nonsense (1) and canonical splice (1) mutations. The patients ranged in ages from 2-19 years and all have moderate-severe intellectual disability, with the majority of the children classified as non-verbal and non-ambulatory. Commonly reported features were developmental delay (14/14), microcephaly (12/14), abnormal EEG (14/14), seizures (13/14), abnormal muscle tone (14/14), sensorineural hearing loss (14/14), and visual impairment (13/14). Other features included brain MRI abnormalities, short stature, failure to thrive, and immunodeficiency. Two patients had muscle biopsies that revealed a possible mitochondrial disorder. Dysmorphic features were not noted. The findings presented in this case series may facilitate the identification of additional affected patients to further the knowledge regarding the phenotype associated with pathogenic variants in *SPATA5*.

2081T

The Emerging Need for Genetic Testing in Clinical Psychiatry. *C. G. Bouwkamp[1,2], J. A. Kievit[2], L. van Zutven[2], R. van Minkelen[2], I. Sterrenburg[3], V. Bonifati[2], S. A. Kushner[1].* 1) Department of Psychiatry, Erasmus MC, Rotterdam, The Netherlands; 2) Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands; 3) Delta Psychiatric Center, Poortugaal, The Netherlands.

Background: A family history of psychiatric illness is currently the single best predictor of disease risk, confirming the strong genetic basis for psychiatric disorders. However, patients with psychiatric disorders are seldom referred clinically for genetic testing. With the recent technological developments in genotyping rapidly adding to our knowledge of psychiatric genetics, we draw on the literature on psychiatric genetics and our experience of clinical genetic testing with the purpose to recommend a revision of the current standard of care in clinical psychiatry. **Methods & Findings:** Through our diagnostic service, we ascertained 31 probands with chronic severe mental illness and comorbid mild-to-moderate intellectual disability. Clinical examination included care dysmorphology screening, standard blood plasma and urine metabolic screening, Fragile X testing, and Illumina CytoSNP-12 300k DNA microarray genotyping. In 11 (35%) of 31 consecutively evaluated cases, we were able to provide an etiological molecular genetic diagnosis (n=4) or identify a potentially causative DNA copy number variant (n=7). In addition to these singleton cases, several families with a Mendelian pattern of psychiatric disease inheritance were ascertained, including one with a balanced translocation segregating with affective psychosis and recurrent spontaneous abortions. Whole genome sequencing and Sanger sequencing confirmation enabled us to identify the exact breakpoints of the translocation [t(6;15)(q26;q21)], which has intragenic breakpoints within *BCL2L10* and *PNLDC1* in all affected family members available for testing. **Conclusion:**On the basis of the existing literature and our experience with patients and families with chronic severe mental illness, we conclude that clinical genetic testing and counseling is warranted in patients with complex psychiatric syndromes, in particular those with clinically-relevant intellectual disability and/or congenital abnormalities.

2082F

Does this patient need to be tested for Lynch Syndrome? Assessing the reliability of family history for ascertainment. *H. C. Cox[1], E. Rosenthal[2], Y. Qian[1], B. Coffee[1], R. Wenstrup[2], B. B. Roa[1], K. R. Bowles[1].* 1) Myriad Genetics, Inc. , Salt Lake City, UT; 2) Myriad Genetic Laboratories, Inc. , Salt Lake City, UT.

Background: Lynch Syndrome (LS) is a dominantly inherited cancer syndrome caused by pathogenic germline variants in mismatch repair genes *MLH1*, *MSH2*, *MSH6*, and *PMS2*, as well as *EPCAM*. Mutation carriers are candidates for more aggressive clinical management such as increased surveillance and prophylactic surgery. Age and family history are key components in targeted genetic testing of individuals as outlined by Amsterdam criteria, Bethesda guidelines, and predictive algorithms. However, family history may be insufficient to identify individuals carrying pathogenic variants due to incomplete penetrance, phenotypic heterogeneity and limited generation sizes in contemporary families. **Methods:** We used pedigree simulation and statistical modelling to estimate the probability that a 40 year-old proband carrying a pathogenic germline variant in *MLH1*, *MSH2* or *EPCAM* would meet family history requirements for genetic testing. Simulated pedigrees spanned 3 generations with sibship sizes of 2 or 4. Sporadic cancer risks, birth age, and mortality were estimated from US population data. LS cancer risks were estimated from published values available at www. myriadpro. com. All analyses were performed using R version 3. 1. 3. **Results:** Preliminary results failed to identify >55% of appropriate patients using Amsterdam II criteria in sibships composed of 2 individuals. Relaxing this family history criteria as follows: 1) colorectal or endometrial cancer under age 50, 2) any LS cancer in an individual with a 1st or 2nd degree relative with a LS associated cancer at any age, 3) 2 primary LS cancers at any age, and 4) any unaffected individual who has a 1st degree relative that meets one or more of the criteria above, still failed to identify >12% of patients with pathogenic variants. Both family history analysis models failed to detect >25% of unaffected patients. **Discussion:** Genetic testing based on family history will identify a subset of Lynch Syndrome patients prior to cancer onset. However simulations demonstrate that the strength of family history declines significantly for families harboring a pathogenic variant with incomplete penetrance, particularly if the proband is unaffected and generation sizes are small. Whilst universal population testing is not yet a viable option, clinicians should consider broad pan-cancer testing rather than more limited cancer-specific testing for patients who are being assessed for inherited cancer risk.

2083W

Clinical and radiographic indications for genetic testing of spondylocostal dysostosis. *M. Lefebvre[1, 2], J. Saint Onge[2], B. Doray[3], C. Francannet[4], L. Pinson[5], A. Bazin[6], S. Berg[7], T. Attié-Bitach[8], C. Baumann[9], M. Fradin[10], G. Pierquin[11], S. Julia[12], C. Quelin[10], A. Dieux-Coselier[13], C. Vincent-Delorme[13], L. Lambert[14], N. Bachmann[15], D. Lacombe[16], B. Isidor[17], J. Roume[18], P. Blanchet[5], S. Odent[10], D. Kervan[19], N. Laporrier[20], C. Abel[21], V. Cormier-Daire[22], S. El Chehadeh[23], J. B. Riviere[2], L. Faivre[1, 2], J. Thevenon[1, 2].* 1) Medical genetics, CHU Dijon, Dijon, France; 2) Equipe EA 4271 GAD Génétique des Anomalies du Développement, FHU Translad, Université de Bourgogne, Dijon, France; 3) Medical genetics, CHU La Réunion, Saint Denis, France; 4) Medical genetics, CHU Clermont-Ferrand Clermont ferrand, France; 5) Medical genetics, CHU Montpellier, Montpellier, France; 6) Departement de Génétique et de biologie spécialisée, Laboratoire Cerba, St-Ouen-L'Aumône, France; 7) Medical genetics, CHD Felix Guyon, La Réunion, France; 8) Institut Imagine, 75015 Paris, France Département de Génétique, Hôpital Necker-Enfants Malades, Assistance Publique/Hôpitaux de Paris, Paris, France; 9) Medical genetics, Département de Génétique,Hopital Robert-Debré, Paris, France; 10) Medical genetics, CHU Rennes, Rennes, France; 11) Medical genetics, CHU Sart Tilman, Liege, Belgium; 12) Medical genetics, CHU Toulouse, Toulouse, France; 13) Medical genetics, CHU Lille, Lille, France; 14) Medical genetics, CHU Nancy, Nancy, France; 15) Zentrum für Humangenetik, Bioscientia, Ingelheim , Germany; 16) Medical genetics, CHU Bordeaux, Bordeaux, France; 17) Medical genetics, CHU Nantes, Nantes, France; 18) Medical genetics, CHR Poissy, Poissy, France; 19) Medical genetics, CHR Orléans, Orléans, France; 20) Medical genetics, CHU Caen, Caen, France; 21) Medical genetics, CHU Lyon, Lyon, France; 22) Department of Genetics, Paris Descartes University-Sorbonne Paris Cité, Necker Enfants Malades Hospital, Imagine Institute , Paris, France; 23) Medical genetics, CHU Strasbourg, Strasbourg, France.

Background: Spondylocostal dysostosis (SCDO) is radiographically defined as multiple segmentation defects of the vertebrae (SDV) affecting at least ten contiguous levels, with grossly symmetric abnormalities of the ribs. Clinical features of SCDO are a short trunk in proportion to height, a short neck and a non-progressive scoliosis. To date, 5 genes are known to cause SCDO : DLL3, MESP2, LFNG, HES7 and TBX6. Overall, the diagnostic yield associated with the sequencing of these genes approaches 25%. Objective: To identify relevant clinical and radiological findings that should orientate towards targeted gene sequencing with a relevant diagnostic yield. Methods: A national recruitment of patients with SDV was realized. For each patient, clinical data and X-ray were asked to the referring clinician. Genomic DNA was sequenced for the coding sequence of DLL3, MESP2, LFNG, HES7 and TBX6 using NexteraXT protocol on a MiSeq instrument (Illumina). Results: Overall, 53 patients were recruited in a two-years period. Clinical information was missing for nine cases. The global diagnostic yield was of 11%. Overall, 28/44 patients presented with visceral malformations and SDV. In this subgroup, no patient was found mutated for the tested genes. Among the 16 patients with isolated segmentation defect, eight had less than 10 contiguous SDV levels and were negative for gene testing. Among the eight cases with at least 10 contiguous SDV levels and no visceral malformation, six were positive for mutation screening, namely DLL3 mutations in two patients, LFNG mutations in 2 patients, MESP2 mutations in one patient and TBX6 mutation in 1 patient. In this specific subgroup, the diagnostic yield was of 75% (6/8). A radiological description of the positive cases was strongly consistent with classical description in the DLL3 and MESP2 cases. Discussion: In this cohort, the overall diagnostic yield was of 11%. A radiological selection for patients with at least 10 SDV levels and no visceral malformation would have lead to a 75% (6/8) diagnostic yield. Although DLL3 and MESP2 mutation may be anticipated based on published images, the radiological findings remained unpredictable for TBX6 and LFNG patients with one family reported in the literature for each. To conclude, our results suggest that targeted sequencing may be preferred in typical cases or in patients with extended SDV without visceral malformation.

2084T

Exome sequencing in intellectual disability and epilepsy. *J. Halvardson[1], J. Zhao[1], CS. Zander[1], E. Månsson[2], H. Edert Sävmarker[3], G. Brandberg[4], AC. Thuresson[1], L. Feuk[1].* 1) Department of Immunology, Genetics and Pathology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 2) Örebro University hospital, Örebro, Sweden; 3) Gävle hospital, Gävle, Sweden; 4) Falun hospital, Falun, Sweden.

Intellectual disability (ID) has a prevalence of 1 – 3% and is defined by an IQ less than 70 with an onset before the age of 18. It has been estimated that at least 20 – 30% of ID patients are co-diagnosed with epilepsy. This gives strong over-representation of epilepsy in these patients compared to the general population prevalence of 0. 05 to 0. 1 %. In recent years exome sequencing has proved to be an effective method to identify causative disease genes in a great number of different syndromes, including syndromes with ID and epilepsy. Here we present the results of exome sequencing in 38 trio families where the index patient is diagnosed with both ID and epilepsy. The families have previously no known causative mutation or history of disease and have previously been screened for copy number variation using SNP arrays. We report 27 validated *de-novo* SNVs or INDELs in 20 of the 38 trio families (~53%). Among the validated *de-novo* mutations identified 22 were nonsense or missense mutations, while 5 were silent mutations. Nine of the *de-novo* mutations were found in genes where mutations previously have been reported to cause disease with ID and epilepsy. Several of the remaining mutations were found in genes highly expressed in the brain and predicted to be highly disruptive (CADD score > 20), opening for the possibility of identification of new genes associations to ID and epilepsy. In one patient we found a nonsense mutation in the *SETD5* gene, mutations in this gene are known to cause ID but not epilepsy. Interestingly in the same patient we also found a *de-novo* mutation in the *ERC2* gene, mouse knockouts of this gene have a disturbed inhibitory synaptic strength. Taken together this raises the possibility of the *SETD5* and *ERC2* mutations acting together to create both epilepsy and ID in the carrier. We also identify mutations in several genes not previously linked to ID or epilepsy, but that represent excellent candidates as causative genes based on their function in the brain or important metabolic pathways. For some of these genes mouse models show phenotypes with impairment of neurological function, adding evidence for mutations in these genes to be causative in neurological disease. Taken together the results of this study point out several new potentially causative genes in ID and epilepsy that are of great interest for further studies to understand their involvement indisease etiology.

2085F

UPD1 in a Newborn with Multiple Congenital Anomalies. *A. R. Seman[1], L. Rodan[2], A. C. Woerner[2].* 1) Claritas Genomics, Cambridge, MA; 2) Division of Genetics, Boston Children's Hospital, Boston, MA.

There have been reports in the literature of uniparental isodisomy of chromosome 1 (UPD), but none has revealed a unifying phenotype. In most cases, the UPD1 led to the unmasking of an autosomal recessive condition. A newborn girl with multiple congenital anomalies was referred for the Claritas Genomics custom SNP chromosomal microarray analysis (CMA), which identified uniparental isodisomy of the entire chromosome 1. The patient originally came to medical attention prenatally for concerns regarding small kidneys, dilated ureters, and oligohydramnios beginning at 26 weeks gestation. Additional ultrasounds and fetal MRI identified bilateral dysplastic kidneys and cardiomegaly. The patient was born at 38 weeks via spontaneous vaginal delivery with Apgars of 2, 5, and 5. Postnatal imaging identified aortic coarctation, right ventricular dysfunction, tricuspid regurgitation and confirmed renal dysplasia/hypoplasia. Physical exam was notable for dysmorphic features including deep set eyes, downslanting palpebral fissures, and 3,4 syndactyly of fingers. She later was noted to have bilateral clavicle fractures and bilateral hip dislocation. Mosaic chromosome analysis to assess for potential aneuploidy rescue was normal. None of the known autosomal recessive disorders related to chromosome 1 sufficiently explained the patient's entire phenotype, raising the possibility of more than one disorder, an atypical presentation of a known disorder, or a novel disorder. The patient's fractures raised the possibility of hypophosphatasia, associated with autosomal recessive mutations in the *ALPL* gene on chromosome 1; however, plasma ALP levels were not decreased and urine phosphoethanolamine was not increased. Exome sequencing of chromosome 1 is currently pending with the goal of identifying the responsible gene(s) for this patient's features. This case highlights the ability of SNP CMA to identify UPD that likely would not have been otherwise detected. This information helped to refine our search and will hopefully lead to an ultimate molecular diagnosis.

2086W

Genetic investigation of cystic fibrosis transmembrane regulator mutations in a cohort of consecutive patients candidate for assisted reproductive techniques. *C. Vaccarella, F. Papa, M. B. Majolini, V. Mazzucchi, A. Luciano, M. CA. Rongioletti.* Clinical Pathology Department, San Giovanni Calibita Hospital, Rome, Italy.

Introduction. The present study, investigated the frequency of mutations in the CFTR gene, in a group of consecutive patients candidate for assisted reproductive techniques with the aim of identify subjects carriers of the most severe ones. **Methods.** 22,416 alleles were screened for 56 CFTR gene mutations utilizing the CFTR INNO-LiPA**Results.** CFTR mutations were detected in 6. 2% of the screened alleles. In the large group of alleles analyzed 93. 4% were wt, 4. 4% were characterized by mild mutations, and 1. 7% by severe or severe/mild mutations. Indeed, the most common severe mutation was ΔF508/N observed in 192/22,416 (0. 86%) of all alleles analyzed, followed by the N1303K mutation with the frequency of 36/22,416 (0. 16%). Whereas regarding mild mutations, the most frequent was the 5T polymorphism present in 916/22,416 (4. 1%). **Conclusions.** Our results together with previous studies, reinforce the importance of an accurate determination of mutations in the CFTR gene, including the 5T polymorphism, in order to inform the couple of their carrier risk and the possibility on having affected child. Moreover, our findings highlight the potential of genetic screening as a tool to identify possible compound heterozygous subjects without CF-like symptoms.

2087T

Detection of Translocations in Clinical Cancer Samples using Targeted NGS Data. *S. Agrawal, R. Bettadapura, K. Sandhu, H. Adil, B. Singh, M. Sen, K. Kumari, D. Vishwanath, G. Deshpande, V. Gupta, S. Katragadda, R. Gupta, V. Veeramachaneni.* Strand Life Sciences, Bangalore, Karnataka, India.

Introduction: Chromosomal translocations are one of the primary genetic alterations behind various cancers and therefore detecting them is crucial for cancer therapeutics. Traditionally, oncologists have used single gene testing of hotspot mutations along with IHC (Immunohistochemistry) for copy number variants (CNVs) and FISH (Fluorescence In Situ Hybridization) for translocations to determine the course of therapy for patients. However, recently targeted sequencing of multiple genes is emerging as an attractive alternative in the clinical setting. Here, we demonstrate how a combination of a custom panel design and development of bioinformatics tools for structural variant (SV) detection has enabled the identification of known and novel translocations in somatic cancer samples along with point mutations, small indels and CNVs. **Methods:** A custom panel was designed to detect mutations including SNVs, small Indels and CNVs in regions of 152 genes known to be important for solid cancer therapeutics. Additionally, it was designed to detect gene fusion events that would occur as a consequence of translocations in the ALK, RET and ROS1 genes. Targeted sequencing was performed on >30 clinical samples, including lung, thyroid, breast and colon cancer samples, at an average depth of coverage of >300x. A new algorithm was developed to allow for split alignment of reads that don't align with a high score (>95) to the reference genome. Split alignment splits an input read into two segments, mapping each segment to a different location on the reference. Based on the split reads, a structural variant caller was developed to call out large deletion, insertion, inversion and translocation events. This method is now integrated into StrandNGS v2. 5, which was used to perform all analysis. **Results:** Using a threshold of 10 unique supporting reads, ~20% of the samples were found to have a translocation event involving at least 1 of the 3 genes mentioned above. These included known translocation events such as CCDC6-RET fusion in thyroid cancer as well as novel events involving the ALK gene in lung cancer. We have further confirmed a few of these translocations by RT-PCR. Thus, we have demonstrated that by using a targeted sequencing panel and a structural variant detection method based on split read alignment, we can successfully detect known and novel translocations in clinical cancer samples, thereby increasing the therapeutic yield of a single NGS-based test.

2088F

Report of a patient with cystic fibrosis and homozygosity for the Q1100P *CFTR* variant. *K. E. Singh[1,2], K. Lai[1], P. Do[1], E. Nussbaum[1,2].* 1) Miller Children's Hospital, Long Beach, CA; 2) University of California, Irvine, Irvine, CA.

Variants of uncertain significance (VUS) continue to evoke frustration among clinicians and families, even in well-characterized genes such as *CFTR.* In the absence of published clinical or functional information testing laboratories must rely largely on internal data in calling sequence variants as potentially disease causing, which can result in different laboratories making different decisions on sequence changes. Until functional studies become more easily attainable, as clinicians we desire clinical data on other patients with the same variants to aid in our own interpretation. We argue for more public reporting from clinicians and laboratories on patients with these suspected pathogenic variants. We present a Guatemalan infant of consanguineous parents with elevated immunoreactive trypsinogen (IRT) by California newborn screen, failure to thrive, pancreatic insufficiency, abnormal sweat chloride, and a homozygous VUS in *CFTR.* At five weeks he was admitted to the hospital for failure to thrive and diaphoresis with feeds. IRT on newborn screen was elevated but *CFTR* common mutation panel was normal. His sweat chloride was elevated at 90 mmol/L and fecal elastase was low at 85 and 45 µg/g on two separate occasions. *K. pneumonia* grew on respiratory culture. His weight gain improved with pancreatic enzyme supplementation but at five months he was only at the 5th centile. He was given a clinical diagnosis of cystic fibrosis (CF) (AHC [MIM 219700]). *CFTR* sequencing and deletion/duplication analysis identified a homozygous VUS, p. Q1100P (c. 3299A>C). Q1100P has been published in at least four individuals with a clinical diagnosis of CF; however, their second mutation was unknown. The testing laboratory for our patient does a high volume of *CFTR* testing and has seen this mutation in one individual with a clinical diagnosis of CF and a second mutation, but phase was unknown. Q1100P has not been reported in either dbSNP or 1,000 genomes. The amino acid position is highly conserved among vertebrates and *in silico* analyses were conflicting. Functional data is not available. Given the clear clinical diagnosis in our patient, the expectation for homozygosity given known consanguinity, and the available information on this variant we believe Q1100P is the molecular cause for our patient's CF diagnosis. Additional clinical information on other patients with this VUS as well as functional data can help clarify whether this VUS should be reclassified as disease causing.

2089W

Development and Implementation of a Bioinformatics Framework to Address Pseudogenes in Clinical Whole-Exome Test. *K. Cao[1], Z. Yu[1], KM. Gibson[1], BH. Funke[2], AB. Santani[1,3], M. Sarmady[1].* 1) Pathology, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Laboratory for Molecular Medicine at PCPGM, Boston MA; 3) Dept. of Pathology & Laboratory Medicine, University of Pennsylvania, Philadelphia PA.

Emerging applications of next-generation sequencing (NGS) in clinical settings have enabled genetic diagnostics labs to interrogate thousands of genes simultaneously (e. g. whole-exome). To ensure the quality of all aspects of the NGS test including sequencing, data analysis and interpretation, the College of American Pathologists (CAP) checklist requires clinical laboratories to perform stringent validation testing. A compounding factor that could significantly impact data quality is pseudogenes and highly repetitive regions. Alignment of sequence data in these regions poses significant challenge is inherently challenging, and result in false positive and false negative calls. We have developed and implemented a framework to identify and annotate "pseudo" exons for clinical whole exome-sequencing (WES) testing. For each coding exon, we assign a score between 0 and 1, named PseudoScore (PScore). PScore is calculated based on ENCODE mappability track of UCSC Genome Browser (GRch37). 100mer window size track was chosen to match the read length used for WES. We queried the mappability score for each base pair in the coding exon of all genes and generated the exon level PS by averaging the mappability scores in the region. The PScore correlates to mappability of the exon. As such, a PScore of 1 indicates the genomic sequence of the coding exon is uniquely mapped to the genome while a PScore of 0. 5 means there is another copy elsewhere in the genome and likewise 0. 33 suggests the presence of another two copies and so on. We found that 3,373 RefSeq genes (17. 5%) have at least one coding exon with various levels of sequence homology (PScore < 1) in the genome. Specifically, 1,078 genes have one or more sequence copies of whole coding exon(s) elsewhere (PScore $\leq$ 0. 5). PScores for all the coding exons were pre-calculated, and used as an annotation resource to specifically highlight variants located in an exon with PScore $\leq$ 0. 5. When a pathogenic variant is flagged, the laboratory confirms the origin of the variant by Sanger sequencing. In summary, we proposed a framework to identify pseudo-exons by assigning a PScore to each exon to quantify the level of homology of the exon. This framework can be used by clinical labs offering NGS-based diagnostics tests to fulfill CAP validation requirements and it further enhances reliability of their tests. This may help to flag potentially false variants caused by presence of pseudogenes or repetitive regions in the genome.

2090T

Comparing gene panel and augmented exome tests using a gold-standard dataset. *G. Chandratillake, G. Bartha, A. Patwardhan, D. Church, S. Garcia, R. Chen, J. West.* Personalis, Inc. , Menlo Park, CA.

Assessing the sensitivity and specificity of next-generation sequencing (NGS) based tests poses a challenge as it requires reference samples with known variants in all genomic regions that are included in the assay. Such gold-standard datasets have only recently been available for a subset of the genome. The Genetic Testing Reference Materials Coordination Program (GeT-RM) is a CDC initiative that aims to improve the availability of appropriate and characterized reference materials for genetic testing. These materials can be used by the genetic testing community for quality control, proficiency testing, test development and validation, as well as for research. In an effort to characterize the variants in reference samples, several clinical testing labs have run clinical gene panel tests on these samples and contributed their variant data to the GeT-RM. Together with data from the National Institutes of Science and Technology Genomes in a Bottle Consortium (GIAB), this data constitutes high-quality variant datasets that labs developing new NGS tests can use to assess analytical validity. We compared the gene panel test data submitted to the GeT-RM by clinical labs for reference sample NA12878, to the gold-standard data produced for this sample by GIAB. We also compared the results to those achieved with an augmented exome assay developed by our lab: the ACE Clinical Exome Test. Most of the panels are 100% covered by ACE capture regions and nearly all of the remaining panels are also covered. Taken together, we found that the sensitivity and specificity of the ACE Clinical Exome Test was comparable to most gene panel tests for those genes specifically targeted in the panels. In many instances, the number of false positives and false negatives obtained was lower with the ACE Clinical Exome Test than obtained with gene panel tests. Increased standardization and transparency in reporting accuracy data for NGS tests will facilitate test selection by healthcare providers and therefore ultimately benefit patients.

2091F

Microcephaly panel testing reveals an exon-level CASK deletion in a patient with a previously inconclusive whole exome sequencing result. *A. Shanmugham[1], G. Morley[2], G. Hoganson[2], D. Pineda[1], K. Wierenga[1], D. Tan[1], G. Richard[1], T. Brandt[1].* 1) GeneDx, Gaithersburg, MD; 2) Rockford Memorial Hospital, Rockford, IL.

The CASK gene encodes the calcium/calmodulin-dependent serine protein kinase, a member of the membrane-associated guanylate kinase (MAGUK) scaffolding protein family. Mutations in the CASK gene are inherited in an unusual X-linked manner, as heterozygous females develop mental retardation and microcephaly with pontine and cerebellar hypoplasia (MICPCH), while males may have isolated intellectual disability or FG syndrome (FGS, Opitz-Kaveggia syndrome). Here we present a 28 year old female with severe microcephaly, intellectual disability, seizures and cerebellar hypoplasia noted on brain MRI. Whole exome sequencing and mitochondrial genome testing was performed at GeneDx. Results revealed a heterozygous nonsense mutation in the CEP152 gene, which is associated with autosomal recessive Seckel syndrome and autosomal recessive primary microcephaly type 9 (MCPH9) without significant cognitive impairment, short stature or other dysmorphic features. It was noted that exon 22 of CEP152 lacked sufficient coverage in the proband by exome analysis. Subsequent exon-level copy number analysis of CEP152 was negative, thus excluding the presence of an intragenic deletion or duplication on the other allele. As the whole exome sequencing results were not diagnostic, a multigene panel, including sequencing and deletion/duplication analysis of the CEP152 gene and 27 additional genes associated with primary and secondary microcephaly was performed. This panel identified the nonsense mutation in CEP152; however, it also identified a partial deletion of 609 kb including exon 10 of the CASK gene, which is predicted to result in haploinsufficiency. Although this specific deletion has not been reported previously, other exon-level intragenic deletions and duplications as well as protein-truncating sequence variants are a common cause of CASK-related disorders. Therefore, the presence of this intragenic deletion is consistent with a diagnosis of MICPCH in the proband. This result highlights the utility of a multigene panel that includes both exon-targeted deletion/duplication and sequencing analysis as well as the limitations of whole exome sequencing in identifying small copy-number changes.

2092W

DFNB4 linkage analysis and *PDS* mutation detection in Iranian patients with deafness. *N. Yazdanpanahi[1], M. A. Tabatabaiefar[2], E. Farrokhi[3], M. Hashemzadeh Chaleshtori[3].* 1) Department of Biochemistry, Falavarjan Branch, Islamic Azad University, Iran, Isfahan, Iran; 2) Department of Medical Genetics, School of Medicine, Isfahan University of Medical Sciences, Isfahan, Iran; 3) Cellular and Molecular Research Center, School of Medicine, Shahrekord University of Medical Sciences, Shahrekord, Iran.

Introduction: Mutations of *PDS* (SLC26A4) are associated with recessive non-syndromic deafness (DFNB4) and Pendred syndrome (PS). To date only a little is known about *PDS* mutations in Iran. In the present study the prevalence and types of *PDS* (*SLC26A4)* mutations and the relevant phenotypes in a series of Iranian patients with deafness were investigated **Methods**: Sixty patients and 61 unrelated families were included in the current study. In the 61 families, linkage was conducted for at least 4 informative short tandem repeats (STRs) of the DFNB4 locus. Selected individuals from the linked families and all of the 60 deaf individuals were subjected to sequencing of *SLC26A4*. Temporal bone CT-scan, thyroid ultrasonography and hormone assays were used to identify phenotypes of patients. **Results:** Seven out of 61 (11. 5%) families were shown linkage**.** Twelve mutations including: c. 2106delG, c. 65-66insT, c. 881-882delAC, c. 863-864insT, c. 1226G>A, c. 1238A>G, c. 1334T>G, c. 1790T>C, c. 1489G>A, c. 919-2A>G (IVS7-2A>G), c. 1412delT and c. 1197delT, were clarified in 9. 1% (12 families or patients) of total investigated alleles. Six out of 12 (50%) families with mutations had pendred syndrome (PS). **Conclusion**: The results probably suggest the high prevalence and specificity of *SLC26A4* mutations among Iranian deaf patients. It may lead to clarification of the population-specific mutation profile which is of importance in molecular diagnostics of deafness.

2093T

Clinical laboratories collaborate to resolve variant interpretation differences in ClinVar. *S. M. Harrison[1,2], J. S. Dolinsky[3], L. Vincent[4], A. Knight Johnson[5], E. C. Chao[3], D. M. Azzariti[1], S. Bale[4], H. L. Rehm[1,2].* 1) Laboratory for Molecular Medicine, Partners HealthCare Personalized Medicine, Cambridge, MA; 2) Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 3) Ambry Genetics, Aliso Viejo, CA; 4) GeneDx, Gaithersburg, MD; 5) Human Genetics, University of Chicago, Chicago, IL.

Recent efforts by the ClinGen Project (http://www. clinicalgenome. org/) to support widespread data sharing using the ClinVar database have allowed clinical laboratories to share variant interpretations that previously had been unpublished or unavailable to the larger community. In March, ACMG published guidelines for variant interpretation that provide a framework to classify variants; however, given the complexity of variant interpretation, application of guidelines still require subjective interpretation. In addition, many variants in the literature or reported in ClinVar, were assessed prior to the availability of these new guidelines. Four clinical laboratories have been working together to resolve these differences. As of June 1st 2015, Ambry Genetics, GeneDx, the Laboratory for Molecular Medicine (LMM), and University of Chicago's Genetic Services Laboratory, have shared 9822, 11706, 12092, and 7127 variant interpretations in ClinVar, respectively, resulting in 4879 unique variants interpreted by ≥2 of the four participating labs. Of these variants, 572 (12%) have one- or two-step differences between the three major categories: pathogenic/likely pathogenic, variant of uncertain significance, and likely benign/benign. To aid in understanding the basis for differences, non-identical interpretations were categorized as follows: Differences in Classification Algorithms (e. g. frequency thresholds), Difference in External Evidence (e. g. different data sources used), Differences in Internal Evidence (e. g. proband phenotype or segregation data), and Differences in Subjective Interpretation of Evidence. Identifying reasons for these differences enables the development of more specific guidance for variant interpretation and increases consistency amongst classifications. Preliminary data suggest the reasons for differences are variable and include all of the above. As the four participating clinical labs work through the 572 differences, trends are emerging to facilitate resolution of differences in bulk as opposed to evaluating variant-by-variant. Thus, data sharing through ClinVar offers a unique opportunity to identify classification differences between laboratories and to work together to resolve differences and strengthen the interpretation of variants that are used in patient care.

2094F

Beyond the ACMG 56: Parental choices and initial results from a comprehensive whole genome sequencing-based search for predictive genomic variants in children. *M. S. Meyn[1,2,3,4,5], N. Monfared[2], C. R. Marshall[6,7], D. Merico[1,6], D. J. Stavropoulos[7,8], R. Z. Hayeems[2,9,10], M. Szego[6,10,11], R. Jobling[4], M. Girdea[1,12], G. D. Bader[3,13], M. Brudno[1,12], R. D. Cohn[1,2,3,4,5], S. W. Scherer[1,2,3,6,14], R. Zlotnik Shaul[5,10,15], C. Shuman[3,4], P. N. Ray[1,2,3,6,7], S. C. Bowdin[2,4,5].* 1) Program in Genetics and Genome Biology, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 3) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 4) Division of Clinical and Metabolic Genetics, Dept of Paediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 5) Department of Paediatrics, University of Toronto, Toronto, ON, Canada; 6) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 7) Department of Paediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 8) Department of Laboratory Medicine and Pathology, University of Toronto, Toronto, ON, Canada; 9) Program in Child Health Evaluative Sciences, Hospital for Sick Children, Toronto, ON, Canada; 10) Joint Centre for Bioethics, University of Toronto, Toronto, ON, Canada; 11) Centre for Clinical Ethics, St. Joseph's Health Centre, Toronto, ON, Canada; 12) Department of Computer Science, University of Toronto, Toronto, ON, Canada; 13) The Donnelly Centre, University of Toronto, Toronto, ON, Canada; 14) McLaughlin Centre, University of Toronto, Toronto, ON, Canada; 15) Department of Bioethics, The Hospital for Sick Children, Toronto, ON, Canada.

Paediatric genomic medicine presumes that knowledge of a child's genomic variants can be used to anticipate, diagnose and manage disease. A challenging aspect of this clinical paradigm is the identification of presymptomatic health risks in children through active searches for predictive (secondary) medically actionable genomic variants (MAVs). We are investigating this issue through the SickKids Genome Clinic, a multidisciplinary research project that carries out whole genome sequencing on 150+ children/year who are undergoing clinical genetic evaluations. With parents' permission, we search their child's genomes for predictive MAVs in 2800+ disease genes listed in the NIH Clinical Genomic Database. Of 373 families approached to date, 208 (56%) agreed to participate in the project and therefore receive their child's predictive childhood-onset MAVs. 131/208 (63%) participants also chose to learn their child's secondary adult-onset MAVs. Among these parents, 104/131 (79%) decided to learn their own status for the same risk variants. Consent involved multiple encounters with parents taking an average of ~9 days to reach a decision. Parents declining secondary variants were most concerned about psychological burdens and/or insurance discrimination. Bioinformatics analysis of the first 80 patient genomes yielded 2382 candidate variants in 1117 genes that were then manually assessed (~30 variants/genome.) While ~15% were novel, ~70% of candidate variants were listed in HGMD and/or ClinVar. However, most variants listed in HGMD as disease-causing failed manual assessment. E. g. , >90% of variants for dominant diseases listed in HGMD as disease causing were rejected due to inadequate evidence of pathogenicity. After manual assessment, 24/80 (30%) of children had at least one reportable predictive MAV. 7/80 (~8%) of children had a reportable predictive MAV in one of the 56 ACMG genes. Expanding our search 50 fold to include 2800+ disease genes yielded an additional 20 reportable predictive MAVs. 5/27 MAVs predicted adult-onset disease. Return of predictive MAVs and assessment of their penetrance is currently underway. We conclude that parental opinions vary widely regarding return of predictive MAVs; comprehensive genomic searching can yield predictive MAVs in ~30% of children; most predictive MAVs are associated with childhood-onset disorders; and the number of reportable predictive MAVs is constrained by disease prevalence and imperfect variant interpretation.

2095W

Clinical diagnostic pipelines at the Exome Laboratory in Baylor Miraca Genetics Laboratories. *Y. Ding[1], J. Chandarana[2], L. K. Dolores-Freiberg[3], W. Liu[1], J. Ma[3], S. Matakis[3], P. G. Otenyo[3], C. Qu[2], N. Saada[3], R. C. Glenn[1], N. Veeraraghavan[1], T. Chiang[1], A. C. Hawes[1], H. Doddapaneni[1], C. J. Buhay[1], J. Hu[1], M. Wang[1], V. Korchina[1], H. Cui[2], J. Scull[2], A. L. Beaudet[2], R. A. Gibbs[1], C. Eng[2], D. M. Muzny[1], Y. Yang[2].* 1) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 3) Baylor Miraca Genetics Laboratories, Houston, TX 77021, USA.

The Exome Laboratory at Baylor Miraca Genetics Laboratories (BMGL) has been providing clinical Whole Exome Sequencing (WES) services since October 2011. To date, approximately 6500 germline WES cases and 230 cancer/normal pair WES (CES) cases have been sequenced on the Illumina HiSeq2500 or HiSeq2000 with average of >11. 4 Gb data per WES, 17. 5 Gb per CES respectively and more than 97% of the targeted exome regions sequenced at a depth of 20X (average coverage >140X per WES and >212X per CES). In order to generate WES data with enhanced gene coverage and to improve our overall turnaround time (TAT), multiple process improvement measures have been implemented, including WES Version 3 capture reagent, which polishes the coverage of over 3600 known disease genes, a quick WES with overnight hybridization (QWES) for enabling five working days of TAT in the wet lab as well as automation using Biomek Beckman FXP (Laboratory Automation Workstation) dual arm robots in library construction for high throughput sample preparations. . In addition to regular proband only and trio exome testing, the lab also offers prenatal trio exome and critical trio exome with an overall TAT of 2-3 weeks for final clinical report. Gene panel tests include custom gene panel test BluePrint, for which referral physicians can order any genes of interest, as well as a panel of 168 autosomal recessive genes for universal carrier testing. The latter is a new test launched in our laboratory recently with quick TAT (14 days from sample intake to final clinical report) and high throughput by utilizing QWES and automation. In summary, the high capacity and short TAT for exome and panel tests has enabled the laboratory to provide exceptional service to patients and clinicians.

2096T

Interpreting Exome Data Using Selection-Based Prioritization. *J. T. Shieh[1,2], H. Gong[2], X. Ge[2].* 1) Institute for Human Genetics, University of California, San Francisco, CA; 2) Division of Medical Genetics, Pediatrics and Benioff Children's Hospital, University of California, San Francisco, CA.

Purpose: Exome sequencing is widely available, but significant challenges lie in data interpretation. To increase the yield from exome sequencing efforts, unbiased prioritization can be used for variants in sparsely annotated regions of the exome. **Methods**: Here we developed a strategy to evaluate any coding region in the genome for disease priority based on signature of purifying selection regardless of prior annotation. Thereby any variant from exome testing could be flagged if in critical coding regions. We performed exome-wide analyses using thousands of control adult exomes and determined coding regions that lack amino acid-altering variation. We tested for missense-depleted regions (MDRs) and corresponding protein characteristics to identify regions for potentially key conditions that limit reproductive fitness. We also apply these methods to inconclusive exome data from undiagnosed disease patients from our institution. **Summary of Results**: Genes with missense depletion demonstrated modest overall enrichment for disease genes (Orphanet/OMIM annotated). A greater difference in disease prioritization by missense depletion was identified in genes for x-linked conditions, suggesting optimized utility in that mode of analysis. Exome-wide scanning revealed MDRs in genes (n=239), including ones that align with conserved protein domains. By applying missense depletion analysis to undiagnosed pediatric disease exomes from patients with abnormal brain MRIs, we identify new candidates for disease. Further application of these methods will identify genes for severe conditions, including neonatal and childhood diseases that should be prioritized in clinical interpretation.

2097F

Successful linkage analysis in classical phenylketonuria families followed by direct sequencing and mutation detection. *S. Tabei, M. Silawi.* Shiraz Medical School, Shiraz university of Medical Sciences, Shiraz, Fars, Iran.

Introduction: Phenylketonuria (PKU) is the most common disorder of inborn errors of metabolism. Prevalence of PKU is about 1:10000 live births; however, due to high rate of consanguinity in Middle East and North of Africa prevalence of PKU is more than other areas. It is estimated between 1:4000 in Turkey and 1:3672 in Iran. **METHODS***:* Ten irrelated classical PKU families from Iran-Fars province were enrolled. Linkage analysis was performed through application of highly linked genetic markers to the *PAH* gene (VNTR, PAHSTR, and *Xmn*I marker) with new designed primers for polymerase chain reaction (PCR). Reliability of approach was assessed by Sanger sequencing, mutation detection and capillary electrophoresis (CE). **RESULTS***:* Through application of linkage analysis nine out of ten families were genotyped successfully. Heterozygosity of chromosome 12 was not detected in any of the enrolled PKU patients. Specificity of new designed primers for linkage analysis was confirmed by Sanger sequencing. Obtained results from linkage analysis were confirmed by direct sequencing and detecting causative mutations in half of the genotyped families. All the results were the same as the linkage analysis results. Labeled primers were capable and linkage analysis by CE was successful. **CONCLUSION***:* Linkage analysis is a powerful and reliable approach for detecting carriers in PKU families which are not screened for causative mutations before. We suggest studying the feasibility of the approach in preliminary diagnosis of PKU and confirming autozygosity of chromosome 12, prenatal diagnosis and preimplantation genetic testing. Also we recommend using labeled primers for constructing faithful local PKU associated haplotypes databases to provide fast, cheap and reliable detection of causative mutations in new cases of hyperphenylalaninemia.

2098W

Genetic variation in the *SMAD3* gene is associated with knee osteoarthritis in north Indian population. *A. ch. Sharma[1], R. N. Srivastava[1], A. Mishra[1], S. R. Srivastava[1], K. Baghel[1], S. Raj[2].* 1) orthopaedic surgery, king george medical university, Lucknow, Uttar Pradesh, India, India; 2) MS Ramaiah Medical College - Bangalore.

Introduction: Osteoarthritis (OA) is the most common degenerative arthritis, a type of arthritis that is caused by breakdown of articular cartilage with eventual loss of the cartilage of the joints. Smad3 is a key intracellular messenger in the transforming growth factor b signaling pathway. Previous study suggested Smad3 gene mutation is a possible predisposing factor for human OA and found gene mutation in OA, providing insight into the function of SMAD3 mediated TGF-b signals in the development of OA and also suggested that Smad3 gene mutation may be a risk factor for genetic susceptibilities to OA. In this case control study, we investigated the possible correlation between the SNPs SmaI (C/T; rs6494629), FokI (A/C; rs2289263) in Smad3 gene and susceptibility to knee OA. **Material & Methods**: This study was conducted in the department of Orthopaedic Surgery, King George's Medical University (KGMU), Lucknow. In this study cases consisted of men and women ≥40 years that fulfilled American College of Rheumatology (ACR) clinical and radiographic criteria for knee OA. Venous blood samples were obtained from all cases as well as controls for genetic analysis. Polymerase chain reactions were performed for SNP analysis using specific primer. **Results:** A total of 200 cases that confirmed radiographic knee OA and equal number of age and sex matched healthy controls were enrolled. There was no significant difference in demographic characteristics between the cases and controls. SNP (rs6494629 and rs2289263) mapping to intron 1 of SMAD3 was observed to be associated with knee OA. Within the SNP rs6494629 of Smad3 gene, genotype TT (p=0. 013) and CT (p=0. 009) was found to be significantly associated with knee OA as compared with the CC genotype; within SNP rs2289263, genotype CC (p=0. 04) and CA (p=0. 02) was found to be significantly associated with knee OA as compared with the AA genotype . **Conclusion:** Our data indicate that genetic variation in the SMAD3 gene may be involved in the risk of knee OA in North Indian populations, confirming the results from previous studies on the potential importance of this gene in the pathogenesis of OA.

2099T

Miniaturized next-generation sequencing (NGS) library preparation and quantification - ultra low sample input and ultra low cost libraries. *S. Vaezeslami[1], J. Jenkins[1], C. Frazer[2], B. Paeper[2], B. Yu[3], S. Quake[3], R. Morey[4], S. Mora-Castilla[4], L. Laurent[4], B. Peters[5].* 1) TTP Labtech Ltd, Melbourn Science Park, Melbourn, Rosyton, Hertfordshire, SG86EE, UK; 2) University of Washington, School of Medicine, Department of Genome Sciences, 3720 15th Ave NE, Seattle WA 98195, USA; 3) Stanford University, Lokey Stem Cell Building, Stanford, CA 94305; 4) University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA; 5) Complete Genomics, A BGI company, 2071 Stierlin Court, Mountain View, CA 94043, USA.

Miniaturized next-generation sequencing (NGS) library preparation and quantification - ultra low sample input and ultra low cost libraries Miniaturization of reaction volumes in different molecular biology applications such as next-generation sequencing (NGS), Gateway® cloning (Life Technologies), and quantitative PCR, has become of great interest in recent years. In the case of sequencing, this is due to a significant reduction in the cost per base pair allowing for high throughput applications, such as single cell nucleic acid sequencing, haplotyping and multiplexing. There are also a number of preparation steps in NGS that would benefit from a reduction in reaction volumes, such as reverse transcription of RNA. Almost all library preparation protocols recommend volumes that are within the volume range of manual pipettes, or that of larger volume liquid handlers. This is to ensure high accuracy and precision of these procedures. Our aim was to reduce the reaction volumes to sub-microlitre values without any changes in final results, thereby saving costs on reagents and sample input, and increasing the throughput. Pipetting of these reduced volumes was performed using an automated low volume liquid handler, mosquito HTS (TTP Labtech, Ltd). Specifically, we will discuss several examples of miniaturized library preparations, such as Nextera and Nextera XT (Illumina, Inc.), reverse transcription using SMARTer Ultra Low RNA Kit (Clontech Laboratories, Inc), and DNA quantification using Kapa SYBR® Fast qPCR kit (Kapa Biosystems, Inc). Here, for the first time, we demonstrated that by using an ultra-low volume liquid handling, we were able to reduce the input of RNA or DNA down to 20 pg, while the miniaturized processes produced high quality libraries and library quantifications comparable to larger volume libraries or reactions. This also provided dramatic cost saving benefits in sample preparation and quantification steps. Using a highly accurate and precise automated liquid handler allows for the miniaturization of a number of processes involved in cloning or next-generation sequencing without compromising the quality of the results. mosquito HTS provides contamination-free, positive displacement, accurate pipetting, which is highly reproducible without any operator error.

2100T

Fetal Jugular Lymph Sacs - What is the Significance? *D. Chitayat[1,2], K. Chong[1,2], M. Bedford[4], A. Toi[3], M. M. Nezarati[1,2,4].* 1) Prenatal Diagnosis and Medical Genetics, Mount Sinai Hosp, Toronto, ON, Canada; 2) Department of Obstetrics & Gynecology, Mount Sinai Hospital, Toronto, ON, Canada; 3) Department of Medical Imaging, Mount Sinai Hospital, Toronto, ON, Canada; 4) Genetics Program, North York General Hospital, Toronto, ON, Canada.

Fetal jugular lymph sacs are accumulations of lymphatic fluid in the anterolateral region of the fetal neck. They may be isolated or occur in association with other structural anomalies and in some cases are presenting with an increased nuchal translucency or cystic hygroma. . Their significance and the guidelines for the prenatal investigation when identified in the course of routine ultrasonography in pregnancy are still controversial. Our protocol for the investigation of these cases includes microarray analysis and DNA analysis for the Noonan syndrome panel. We report our experience with ten cases of prenatally diagnosed isolated fetal jugular neck cysts identified in the last 5 years (Table 1).

Case	GA at detection of Jugular Sacs(weeks)	NT(mm)	Cystic hygroma	DNA analysis for the Noonan Syndrome Panel [GeneDX]	QF-PCR/ Microarray	Outcome
1	13. 7	1. 7	No	N/A	Normal	Normal newborn
2	18. 9	3. 5	No	PTPN11c. 215C>G	Normal	Continued Pregnancy
3	19. 1	7. 4	Yes	PTPN11c. 1381G>A	Normal	TA
4	16. 5	8. 5	Yes	PTPN11c. 184T>G	Normal	Continued Pregnancy
5	15. 8	3. 2	No	N/A	Normal	Normal newborn
6	16. 0	17. 5	Yes	N/A	Normal	TA
7	13. 1	na	No	N/A	Normal	Normal newborn
8	14. 9	16	Yes	SOS1 c. 806T>C	Normal	Continued Pregnancy
9	14	10	Yes	RAF1 c. 776C>T	Normal	TA
10	12. 5	8. 0	Yes	N/A	T21	TA

TA – Therapeutic Abortion; na – not available; N/A – not applicableIn 10 cases identified by us in the last 5 years, 50% had a pathogenic mutation in a Noonan syndrome Panel, one had trisomy 21 and none had abnormality on microarray analysis (in the case with T21 we stopped the analysis following the QF-PCR results). However, all cases with abnormal results had NT > 3. 5 mm or cystic hygroma. This study suggests that the identification of fetal jugular lymph sacs with an increased NT/cystic hygroma is associated with an increased incidence of Noonan syndrome and its related disorders and should instigate mutation analysis of the Noonan syndrome panel.

2101F

Prenatally diagnosed aqueductal stenosis (triventricular dilation) and postnatal/autopsy findings. Report on 100 cases. *R. Jobling[1,2], N. Amir[2], A. Toi[3], P. Shannon[4], K. Chong[1,2], D. Chitayat[1,2].* 1) Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Mount Sinai Hospital, Department of Obstetrics and Gynecology, The Prenatal Diagnosis and Medical Genetics Program; 3) Department of Diagnostic imaging; The University of Toronto, Toronto, Ontario, Canada; 4) Department of Laboratory Medicine and Pathobiology; The University of Toronto, Toronto, Ontario, Canada.

Hydrocephalus is etiologically a heterogenous condition. Up to 66% of the cases of childhood hydrocephalus are the result of stenosis of the aqueduct of Sylvius (Aqueductal stenosis)(AS). Most AS cases present at an early gestation with fetal ultrasound/MRI finding of dilation of the lateral and 3rd cerebral ventricles and the differential diagnosis includes infection, hemorrhage, chromosome abnormalities and a variety of single gene disorders including Dandy-Walker malformation and *L1CAM syndrome*. In many cases the underlying cause cannot be delineated making genetic counseling difficult. We report our experience with 100 cases presented prenatally with triventricular dilation on fetal ultrasound/ MRI. Of the 100 pregnancies 64 were terminated and 36 were continued. There were 12 cases which ended in still birth. In 49 of these cases pathological examination confirmed the diagnosis and in the rest the aqueduct was not found occluded or there was no pathological examination. Of the confirmed cases of AS 15 were isolated and 34 were associated with other congenital anomalies. In 5 cases intra-cerebral hemorrhage was identified and none of these were the result of maternal alloimmune antiplatelet antibodies. In one case fetal infection was identified as a cause. L1CAM gene mutation was found in four cases and in all of these cases the mother was a carrier. Chromosome abnormalities were present in 10 cases (3 cases of trisomy 21, one case of trisomy 18, one case of 45, X, one case of mosaic trisomy 22, one case of an unbalanced translocation of chromosomes 6 and 7, a duplication at 2q11. 2, a duplication of 15p13. 33 and a 110 kb deletion at 2p16. 3). In most cases an underlying etiology could not be identified. This large cohort of patients with AS provides insight into the maternal characteristics, presentation, associated anomalies, etiology and outcome of this rare and anomaly.

2102T

A study of X and Y chromosomal variations in a large fertile childbearing-age population. *P. Kezmarsky[1], E. Kirkizlar[1], Z. Demko[1], S. J. Gross[1], P. Lawson[2], C. Samango-Sprouse[3].* 1) Natera Inc. , San Carlos, CA; 2) Focus Foundation, Davidsonville, MD; 3) George Washington University School of Medicine and Health Sciences, Washington, D. C.

X and Y chromosomal variations are among the most common human whole-chromosomal copy-number variations, yet 75% of these disorders go undiagnosed during the affected individuals' lifetime. Various pre- and post-natal studies have generated incidence and prevalence estimates of these commonly occurring X and Y chromosomal disorders, but the population-based incidence and prevalence in the fertile adult population remains unclear. This study provides the first prevalence estimate for these disorders in the population of fertile individuals of child-bearing age. This retrospective analysis of prospectively collected data leveraged a routine noninvasive prenatal screening test that incorporates maternal and paternal genotypes to estimate the population-based prevalence of X and Y chromosome variations in the fertile adult population. From 141,918 fertile women and 29,581 fertile men, 119 X and Y chromosomal abnormalities (overall prevalence: 1 in 1,441) were identified (See Table 1). Among the females, fertility in 45,X/46,XX mosaic women was higher than previously reported, and demonstrated that pregnancy in non-mosaic 45,X women is possible. This is also the first documentation of incidence and fertility in 48,XXXX women, and provides important information for families with this prenatal diagnosis and individuals with tetra-X. For both cases of 48,XXXX observed in this study cohort, the duplication was observed on the q-arm only; the p-arm was disomic. Among males, reduced reproductive capabilities in those with 47,XYY has been described for the first time and supports greater variability in reproductive capacities in these disorders than previously appreciated. Together, this study expands the knowledge of X and Y chromosomal variations, and is critical to families who receive a prenatal diagnosis of one of these disorders as well as to individuals who may confront fertility issues as adults.

Table 1. Chromosomal abnormalities in the fertile adult population.

Females	n	Observed Incidence, 1 in:	Males	n	Observed Incidence, 1 in:
Total Aneuploid	103	1,378	Total Aneuploid	16	1,846
45,X	43	3,300	47,XXY	2	14,766
45,X	3	47,306	47,XXY	1	29,532
45,X/46,XX Mosaic	40	3,548	46,XY/47,XXY Mosaic	1	29,532
47,XXX	30	4,731	47,XYY	10	2,953
47,XXX	18	7,884	47,XYY	9	3,281
46,XX/47,XXX Mosaic	12	11,827	46,XY/47,XYY Mosaic	1	29,532
48,XXXX	2	70,959	Partial Y Deletion	4	7,383
46,XX Uniparental Disomy	3	47,306			
46,XY (sex reversion)	2	70,959			
Mosaicism of Unknown Type	23	6,170			

2103F

Maternal copy number variants are a significant reason for false positive noninvasive prenatal test results. *D. I. Chudova[1], K. J. Curnow[1], S. Bhatt[1], A. J. Sehnert[1], D. W. Bianchi[2].* 1) Illumina, San Francisco, CA; 2) Mother Infant Research Institute, Tufts Medical Center, Boston, MA.

Background: Noninvasive prenatal testing (NIPT) has been rapidly integrated into care for pregnant women at high risk for fetal aneuploidy. For low-risk women, a lower prevalence of aneuploidy results in lower positive predictive values (PPVs). Because PPVs are largely driven by specificity, we focused on reducing false positive NIPT results, including those that are due to the presence of maternal copy number variants (CNVs). Methods: Two primary modifications were made to the informatics algorithm in the clinical laboratory: additional sub-chromosomal coverage normalization steps and filtering. Normalization reduces variability in the sequencing data, whereas filtering over the chromosome ensures that outliers from sub-chromosomal maternal CNVs are excluded from the overall chromosome z-score. Here, we reanalyzed 11 false positive samples (5 trisomy 21, 2 trisomy 18, 3 trisomy 13, and 1 double trisomy 21/18) from the low-risk CARE study cohort (Bianchi *et al* NEJM 2014) using the updated analytic algorithm. Results: Reanalysis resulted in 6 cases that changed from aneuploidy detected to no aneuploidy detected. Of these, 3 were attributed to the improved normalization, and 3 were attributed to sub-chromosome filtering of outlier data from maternal CNVs. One maternal CNV was a duplication of ~3 Mb on chromosome 18p and two were duplications on chromosome 13q (one was ~5 Mb and the other ~8 Mb). These CNVs were identified in two independent blood samples from each individual. For the 5 cases that remained positive for aneuploidy, the sequencing results were consistent with a complete fetal chromosome gain on the previously identified affected chromosome. The corresponding newborns were all born at term, with no detectable physical abnormalities. Chromosome analyses were not performed on the infants or placentas. In the CARE study cohort, reanalysis with the refined algorithm increased the PPVs for trisomies 21 and 18 to 62. 5% (5/8) and 66. 7% (2/3), respectively. This compares with 45. 5% (5/11) and 40% (2/5) using the previous algorithm, and 4. 2% (3/72) and 8. 3% (1/12) using standard biochemical screening. Discussion: Maternal copy number variants are an increasingly recognized cause of false positive NIPT results. Continued informatics improvements in the analytic algorithm are of significant clinical value, particularly as the clinical population undergoing NIPT expands to include a greater proportion of lower-risk women.

2104T

Comprehensive Analysis of Clinical Performance of Harmony™ Prenatal Test. *R. Stokowski, A. Sparks, T. Musci, A. Batey, K. Song, A. Oliphant, E. Wang.* Ariosa Diagnostics, San Jose, CA.

Introduction: The core technology of the Harmony Prenatal Test is based on a targeted assay, Digital Analysis of Selected Regions (DANSR™) and a proprietary analysis algorithm, Fetal fraction Optimized Risk of Trisomy Evaluation (FORTE™). DANSR assays are targeted against chromosomes 21, 18, and 13 for fetal trisomy and other genomic regions to measure fetal fraction. FORTE analyzes DANSR products along with clinical information to provide a probabilistic assessment of fetal trisomy. Clinical performance of Harmony has been validated in multiple studies demonstrating high sensitivity and low false positive rates for Trisomy 21, 18 and 13 using various DANSR quantitation methods. (Sparks 2012, Ashoor 2013, Gil 2014, Nicolaides 2012, Norton 2012;2015, Verweij 2013, Juneau 2014) This study provides further clinical data using microarray for DANSR quantitation and a summary of overall performance of the Harmony test. **Methods:** Blood samples from 794 singleton, twin, natural and IVF pregnancies were collected for prenatal cfDNA testing in a prospective, multicenter, blinded study. Subjects either had invasive testing (amnio or CVS) for evaluation of fetal chromosome conditions or had karyotype at birth if aneuploidy was suspected at birth following newborn exam (twin samples only). Results of the Harmony test and karyotypes or birth outcomes were compared. Results from this study were combined with previously published studies on the performance of the Harmony test. **Results:** In the new study cohort, the mean maternal age was 35 years (range 18-48) and mean gestational age was 17 weeks (range 10-36). Using the Harmony test, T21 was identified in 106 out of 107 cases (99. 07%; 95% CI 95. 96-100. 00), T18 was identified in 28 out of 29 cases (96. 97%; 95% CI 86. 36-100. 00), and T13 was identified in 12 out of 12 cases. (100%; 95% CI 81. 60-100. 00). The specificity was 100% (95% CI 99. 58-100. 00) for trisomy 21, 18 and 13. Combining this dataset with previously published Harmony studies, the combined performance is:

Test	N	TP	FP	TN	FN	Sens	Spec	Sens 95% CI	Spec 95% CI
T21	23158	418	10	22727	3	99. 29	99. 96	98. 23-100	99. 93-100
T18	22402	147	5	22246	4	97. 35	99. 98	94. 23-100	99. 95-100
T13	14246	30	3	14211	2	93. 75	99. 98	82. 77-100	99. 95-100

Conclusion: Across multiple quantitation methods, the Harmony Prenatal Test with the core technologies of DANSR and FORTE demonstrates high sensitivity and extremely low false positive rates for the common autosomal trisomies in the general pregnancy population.

2105F

Population screening of 328,886 individuals reveals ethnic disparities in guidelines and cumulatively greater risk for severe recessive disease than for Down syndrome or neural tube defects. *I. S. Haque, G. A. Lazarin, M. Raia, H. Bellerose, D. Muzzey, K. D'Auria, H. P. Kang, E. A. Evans, J. Goldberg.* Counsyl, South San Francisco, CA.

Current US medical guidelines recommend universal ultrasound and serum screening in pregnancy for Down syndrome (T21, MIM 190685) and neural tube defects (NTDs, MIM 182940). Genetics guidelines recommend offering screening for cystic fibrosis (MIM 219700) and spinal muscular atrophy (MIM 253300) to all prospective parents and additional screenson an ethnicity-specific basis. The population risk for expanded panels of rare recessive diseases is an open question.

METHODS

We analyzed results of 328,886 tests for up to 110 Mendelian genes in a population reporting no family history, known carrier status, or infertility. Sample ethnicity was self-reported. 90% of tests used targeted genotyping of up to 417 known pathogenic sites; the remainder sequenced all exons and partial introns of the selected genes. We measured *SMN1* copy number by qPCR and *FMR1* 5'UTR CGG repeat count by PCR and capillary electrophoresis.

We interpreted novel variants using literature curation with ACMG guidelines and considered pathogenic those curated as known or likely deleterious [1]. We assigned diseases profound, severe, moderate, or mild severity by review of their symptoms [2].

From carrier frequencies, we computed the probability that a random birth would be affected by a disease (absolute risk). Fragile X syndrome absolute risk was weighted by allele size [3].

RESULTS

In all ethnicities the cumulative absolute risk of 89 severe/profound recessive diseases is larger than that of NTDs (1/1000) or of T21 in young mothers (1/1200, maternal age=20), conditions for which universal screening is recommended [4,5]. In African-Americans and the Ashkenazi Jewish, the cumulative risk of these 89 diseases is higher than that of T21 in the high-risk population (1/280, maternal age=35).

We also find ethnic disparities in care: current ACOG and ACMG guidelines have 1000x and 50x ratios respectively of absolute risk detected in Africans vs East Asians. Including 89 severe/profound diseases shrinks the maximum ratio to 4. 3x, indicating that low positive rates in current guidelines do not reflect the true distribution of disease risk.

We find no statistically significant difference between absolute risk computed from observed carrier-carrier couples and risk computed from carrier frequencies.

[1] Richards et al. Genet Med 17(5) 2015
[2] Lazarin et al. PLoS One 2014
[3] Yrigollen et al. Genet Med 14(8) 2012
[4] Cragan et al. MMWR CDC Surveil Summ 44(4) 1995
[5] Snijders et al. Ultrasound Obstet Gyn 13(3) 1999.

2106T

Next Generation Sequencing Assay Accurately Determines Carrier Status for Spinal Muscular Atrophy. *X. Wang, H. Y. Wan, C. B. Oyolu, K. R. Haas, D. Davison, K. Iori, C. S. Chu, I. S. Haque, E. A. Evans, H. P. Kang, D. Muzzey.* Counsyl, South San Francisco, CA.

INTRODUCTION: Spinal Muscular Atrophy (SMA) is a severe neuromuscular disease that is the second most common fatal autosomal recessive disorder. Nearly 95% of SMA-affected patients have *SMN1* genes that are rendered largely dysfunctional due to a splicing mutation that causes exon 7 to be preferentially omitted from the SMN protein. By convention, this single-base variant distinguishes *SMN1* from *SMN2*. Resolving SMA carrier status involves counting the number of *SMN1* genes and is traditionally determined through fluorescence-based readouts, such as allele-specific quantitative PCR or MLPA. However, these methods are relatively expensive, low throughput, and incapable of detecting novel variants. Here we report a high-throughput next-generation sequencing (NGS) method that naturally fits in the context of an NGS-based expanded carrier screen and accurately determines the *SMN1* and *SMN2* copy numbers in a sample.

METHODS: We designed hybrid-capture probes common to *SMN1* and *SMN2*. NGS reads were assigned to each gene according to the single base that distinguishes them in exon 7. Relative read depths were calculated by normalizing to the depth in regions elsewhere in the genome, where median copy number is expected to be two. Copy-number calls and associated confidence scores were calculated from a distance test that found the most likely integer copy number for each gene. Real-time quantitative PCR was used to confirm the NGS results.

RESULTS: On a set of more than 87,000 clinical samples including positive and negative controls, our NGS-based method correctly determines copy number of *SMN1* and *SMN2*, thereby resolving SMA carrier status with >95% detection rate. Since each of the tiled NGS probes individually contributes enough reads to call the copy number, the test avoids false positives that compromise the qPCR approach when a SNP is within a primer binding site. Finally, the NGS assay has higher accuracy across a broader dynamic range than qPCR, enabling confident copy-number calls in excess of two copies for *SMN1* and as many as six copies for *SMN2*.

CONCLUSIONS: We have developed an NGS-based assay for the most common cause of SMA that offers higher throughput, lower cost, and higher dynamic range than the current state-of-the-art standalone qPCR assays.

2107F

Establishing criteria for the return of results from genome sequencing for the purpose of carrier screening. *K. A. B. Goddard[1], P. Himes[2], L. Amendola[3], J. Berg[4], M. O. Dorschner[5], M. Gilmore[2], C. Harding[6], T. L. Kauffman[1], E. Morris[2], D. Nickerson[7], J. A. Reiss[2], C. S. Richards[6], A. F. Rope[2], D. K. Simpson[2], B. Wilfond[8], G. P. Jarvik[3].* 1) Ctr Hlth Res, Kaiser Permanente Northwest, Portland, OR; 2) Northwest Permanente, Kaiser Permanente Northwest, Portland, OR; 3) Department of Medicine, Division of Medical Genetics, University of Washington, Seattle, WA; 4) University of North Carolina Chapel Hill, Chapel Hill, NC; 5) Department of Pathology, University of Washington, Seattle, WA; 6) Oregon Health and Science University, Portland, OR; 7) Department of Genome Sciences, University of Washington, Seattle, WA; 8) Seattle Children's Research Institute, Trueman Katz Center for Pediatric Bioethics, Seattle, WA.

As part of the NHGRI Clinical Sequencing Exploratory Research (CSER) consortium, the NEXTGen study focuses on the use of genome sequencing in couples seeking preconception genetic screening. Although reporting carrier results from clinical sequencing is feasible and potentially relevant to individuals regardless of indication, these couples may have a greater near term interest in learning about and acting upon findings from carrier screening. We categorized the gene/condition pairs to facilitate patient autonomy in assuring that only requested information is returned. Secondary findings are also returned as part of this study. We established a Return of Results Committee (RORC) to develop criteria to select the genes/conditions to report in NEXTGen. The RORC includes expertise in clinical and molecular genetics, genetic counseling, bioethics, and epidemiology. We considered about 800 gene/condition pairs from commercial panels, publications, and contributions from RORC members. Gene/condition pairs were considered for 1) inclusion, and 2) categorization into one of five categories based on standardized criteria developed by the RORC. The RORC refined the criteria over time based on their experience, and a consensus was reached on each gene/condition pair. So far, 582 gene/condition pairs were determined appropriate for inclusion. Those that were excluded were primarily due to low evidence (e. g. , <2 people or <2 families). Based on participant feedback that they want to hear the worst case scenario, conditions with variable penetrance or expressivity were placed in the most serious category. The default category of serious encompassed the largest number (n=425) of conditions. To be categorized as lifespan limiting (n=186), more than 50% of known cases had to be deceased before the age of 10. Mild conditions (n=96) are typically not life threatening and there is only mild to moderate disruption of normal functions. Late onset (n=14) was defined as symptoms appearing for most people after the age of 20. Conditions with unpredictable outcomes (n=31) have a wide range of phenotypic expressivity that cannot be predicted based on genotype. Categorization is ongoing, but nearly complete. Placing conditions in non-overlapping categories required clear and precise category definitions, while accounting for wide variability in phenotypes. We are studying the utility of these categories for carrier screening.

2108T

Etiological Evaluation of Adverse Reproductive Outcome: A Large Study from South India. *A. Jyothy, A. Srilekha, K. Prasoona, T. Sunitha, A. Venkateshwari.* Institute of Genetics and Hospital for Genetic Diseases,Osmania University,Begumpet, Hyderabad, Telangana State, India.

Introduction: Adverse reproductive outcome (ARO) includes a wide variety of complications such as abnormal pregnancy, pregnancy loss, birth defects etc. The cause of reproductive abnormality could be the result of genetic, anatomic, endocrine, thrombotic, immunologic factors or physiological events that occur in mother, father and child. As there is a huge burden of these abnormalities in developing countries, we aimed to evaluate the causative factors in couples of South Indian origin with ARO. **Subjects and Methods:** The study included a total of 3927 couples with abnormal reproductive histories, referred to the Institute of Genetics and Hospital for Genetic Diseases and Modern Government Maternity Hospital, Hyderabad (March, 2010 to March, 2015). The pregnant women were organized into two groups: Group 1 included women carrying fetal congenital anomalies in present pregnancies (n=446) and Group 2 included women carrying normal fetuses with previous bad obstetric history and medical complications (n=3481). 3D/4D ultrasound scans and TORCH profiles were evaluated in all cases. Chromosome analysis in couples was done on peripheral blood lymphocytes using the standard protocols. Spectral karyotyping was performed to confirm the abnormal karyotypes. Prenatal diagnosis by amniocentesis has been carried out for evaluation of trisomies using FISH. **Results:** The study revealed 37% (n=165) of pregnancies carrying fetuses with central nervous system disorders as major congenital anomalies followed by 20% (n=88) renal anomalies. The seropositivity of Toxoplasmosis and Rubella were significantly higher in group 2 (p=0. 026) and group 1 (p<0. 01) respectively. Cytogenetic analysis in 1498 individuals revealed 5. 40% (n=81) chromosome abnormalities including 8. 64% (n=7) novel translocations. **Conclusion:** Adverse reproductive outcome continues to be a serious concern and cause of morbidity and mortality especially in developing countries and advanced screening methods have facilitated in early detection, appropriate management and counselling of these cases to reduce the burden.

2109F

Whole-genome amplified DNA from neonatal dried blood spot samples produces high-quality exome-sequence data. *M. V. Hollegaard[1], J. B. Maller[2,3], J. Grauholm[1], M. G. Pedersen[4,5], B. Neale[2,3], L. Olsen[6,5], C. Stevens[2], T. Hansen[6,5], P. B. Mortensen[4,5], T. M. Werge[6,7,5], D. M. Hougaard[1], M. J. Daly[2,3].* 1) Departnent Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark; 2) Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 4) National Centre for Register-based Research, Aarhus University, Aarhus, Denmark; 5) Lundbeck Initiative for Integrative Psychiatric Research (iPSYCH); 6) Mental Health Center Sankt Hans, Mental Health Services of Copenhagen, Roskilde, Denmark; 7) Institute of Clinical Sciences, University of Copenhagen, Copenhagen, Denmark.

Population based biobanks provide unique opportunities in the quest to understand the genetics of common and rare diseases. These repositories frequently contain millions of samples collected consistently over many years, and enable cost-effective and large scale studies while also minimizing bias. The Danish Neonatal Screening Biobank (DNSB) is the largest biobank in Denmark (by number of individuals), containing almost 2. 3 million neonatal dried blood spot (neoDBS) samples. This represents nearly the entire population of Denmark born since 1982. In addition Denmark have a unique system of very detailed health and social registries that are mutually connectable as well as linkable to existing biobanks. This allows researchers to study the etiology of numerous diseases in extraordinary detail. NeoDBS samples are precious and often very little material is available for research. To reach the recommended input amounts for NGS and array genotyping, the neoDBS DNA extract can be whole-genome amplified (wgaDNA). While this can be a robust process, it is critical to confirm technical consistency and accuracy for a particularly experimental design. Through exome sequencing and array genotyping of more than 1,000 samples we thoroughly evaluate the reliability and stability of neoDBS wgaDNA. First we addressed the reliability of wgaDNA by comparing genetic data from genomic DNA to wgaDNA from neoDBS. Second we use trios to quantify the de-novo mutation rate from wgaDNA. Third, we use a subset of samples with known 22q11 deletetions to evaluate if we can call large structural variations. Finally, we look at the robustness and usability of using wgaDNA from neoDBS by evaluating the exome-sequencing results of 600 case/control samples for Autism Spectrum Disorder. We demonstrate that neonatal DBS wgaDNA can be to produce high-quality genotype and sequence data. There are no signs of increased mutation rate, and the accuracy of genotype calls are comparable to standard genotyping approaches. That said few regions of the genome is not presented in the wgaDNA, suggesting a slightly sequence related bias of the amplification. Overall, this is overcome by the access to millions of samples, and the fact that the results produced are of high-quality.

2110T

An assessment of parental perspectives, understanding, and responses to fetal whole exome sequencing. *N. Vora[1], E. Hardisty[1], S. Robinson[1], C. Rini[2], D. Skinner[3]*. 1) Obstetrics and Gynecology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Gillings School of Global Public Health; UNC Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) FPG Child Development Institute, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Statement of purpose: A fetal whole exome sequencing (WES) pilot study is ongoing at the University of North Carolina-Chapel Hill on non-viable fetuses. All have congenital anomalies suspected to have a genetic etiology not identified by standard prenatal genetic testing. One of the aims of this study is to assess parental perspectives, understandings, and responses to diagnostic WES information through semi-structured interviews. Methods used: Semi-structured interviews (approximately 45 minutes each) were conducted with both parents in in-person interviews to identify patient expectations, understandings, and perceptions after the genetic counseling and consent process. Mothers completed a pre-sequencing questionnaire and genetics literacy assessment following genetic counseling. Summary of results: Six parent-fetus triads were enrolled. Demographics were as follows: mean maternal age 30. 1 (range 25-38); 5/6 white, non-Hispanic; 6/6 had some college education; mean income $73. 7K (range $37. 5-135K); 4/6 had prior genetic testing; 3/6 planned future children; 6/6 legally married. The mean scores for a pre-sequencing genetics literacy assessment administered after genetic counseling was 92% (88-100%). The average perceived likelihood of the WES result to explain the loss was 6. 25/10; the average perceived likelihood of the WES result to provide information about parental health unrelated to the loss was 5. 2/10. From data obtained from the semi-structured interviews, 5/6 families verbalized that they hoped to obtain an answer to explain their fetal loss; 4/6 verbalized a desire to provide information that would help future families and research. Conclusions: In this pilot study, after being counseled about the ~30% chance of WES providing an answer, parents' perception of the likelihood that WES would provide an answer was higher. Most families verbalized their hope that WES would provide an explanation for their loss. This study suggests that families who have lost a baby with congenital abnormalities have high hopes and expectations about WES. More research needs to be performed to determine the diagnostic capability of WES in a prenatal setting and how family's expectations complicate pre-test counseling. .

2111F

All *FMR1* premutations are not equal: impact of frequency and repeat distribution on risk for fragile X syndrome. *G. A. Lazarin, K. Wong, I. S. Haque.* Counsyl, South San Francisco, CA.

Background:Instability in the *FMR1* gene can result in CGG repeat expansion, which in turn causes fragile X syndrome (FXS, OMIM 300624) when the repeat count exceeds 200. Premutations (55-200 repeats) have a greater risk of next-generation expansion to full mutation with increasing repeat size.

Previous studies have reported FXS prevalence of 1/4000-6000. There is evidence ofethnic predilection, but this is not well defined. Since the actualFXS risk in a population depends on both premutation frequency and allele sizes, we analyzed this data derived from our carrier screening.

Methods:Samples were considered only if no family history, known carrier status, or infertility history were reported. We also excluded those originating from ART clinics to avoid bias from premutation-inducedpremature ovarian insufficiency.

We report data from 134,840 individuals passing the above filters who were screened for FXS, including at least 2,789 in each of 10 self-reported ethnicities. *FMR1* 5'UTR CGG repeat count was assessed by PCR and capillary electrophoresis. Premutations were sub-divided according to previously described risk groups (Nolin 2003).

Results:Mean frequency of any premutationwas 1/383, with substantial ethnic variation:min 1/2857 in E Asians, max 1/182in Middle Easterners. The largest ethnic subgroup, NEuropeans, exhibited a frequency of 1/315.

Premutation repeat size distribution also differed by ethnicity. In all, 46. 1% of premutation alleles were in the lowest risk group (50-55 repeats), but by ethnicity, this ranged from 22. 6%-75. 0%.

We calculated the risk for fragile X syndrome in each ethnic population by weighting the allele frequencies and their associated full mutation expansion risks (Yrigollen 2012). This provides a more accurate indication of risk than reliance on simple carrier frequency. For example, though the premutation frequency was higher in N Europeans (1/315) than S Europeans (1/412), the latter's repeat distribution favored alleles with high transmission risk: 18. 2% of premutations were 90-200 repeats, compared with 4. 1%. The result of this phenomenon is actuallya higher risk for FXS in S Europeans, 1/2222, compared to 1/3846.

The highest risk identified was among Middle Easterners (1/1370) and the lowest was among SE Asians (1/100000).

Conclusion:Premutation frequency *and* repeat length must be considered in calculating risk for FXS. Our large screening database enables tabulation of accurate, ethnicity-specific, risk estimations.

2112T

High-efficiency biomarker extraction from dried blood samples using Adaptive Focused Acoustics (AFA™). *U. Geigenmuller, A. Purdy, W. Obendorf, G. Smejkal, A. Palmer, G. Durin, H. Khoja, J. Laugharn.* Covaris Inc, Woburn, MA.

Collecting and drying small amounts of blood on a solid support, such as filter paper, has several advantages over other forms of blood collection: It can be performed outside of a medical setting, requires only small amounts of blood, allows specimen transport at ambient temperature and as non-hazardous material, and facilitates storage for extended periods of time. Dried blood spots (DBS) are routinely used for newborn screening programs and are also finding use in adult screening programs such as HIV drug resistance testing. However, DBS also have a number of drawbacks. Specifically, variations in blood collection and blood spot drying techniques can easily render specimens unusable; insufficient drying or subsequent exposure to humidity leads to biomarker degradation; inefficient biomarker retrieval limits downstream analyses; and variation of spot size with spotted volume and haematocrit complicates quantitative analyses of blood biomarkers. Here, we present two innovations that together overcome many of the drawbacks of the current DBS technology: (1) a novel sampling device for collecting capillary blood (truCOLLECT) that simplifies the collection procedure, eliminates the need for a drying period before shipment, provides integrated sample barcoding, and increases biomarker yield; and (2) use of Adaptive Focused Acoustics (AFA™) for efficient release of biomarker from dried blood samples collected on truCOLLECT devices as well as from conventional DBS. The AFA™-based extraction process is highly automated and allows processing of up to 96 samples in 4 hours or less in a 96-well plate format, with less than 30 min hands-on-time. We determined yield of DNA, RNA, and protein extracted from truCOLLECT samples or DBS from six different donors using AFA™ or traditional chemical extraction methods. Biomarker quality was assessed by gel electrophoresis and qPCR for nucleic acids and by gel electrophoresis and ELISA for proteins.

2113F

Prostacyclin Synthase (PTGIS) interacts with nonsteroidal anti-inflammatory drugs (NSAIDs) to Protect from Miscarriage. *M. J. Bray[1], K. E. Hartmann[2,3,4], E. S. Torstenson[2], S. H. Jones[5], D. R. Velez Edwards[1,2,3,5].* 1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 2) Vanderbilt Epidemiology Center, Vanderbilt University, Nashville, TN; 3) Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN; 4) Department of Medicine, Vanderbilt. University, Nashville, Tn; 5) Institute of Medicine and Public Health, Vanderbilt University, Nashville, TN.

Nonsteroidal anti-inflammatory drugs (NSAIDs) are one of the most commonly used medications reported in pregnancy. NSAIDs directly impact prostaglandin pathways, and have been proposed as potential risk factors for spontaneous abortions (SAB, gestation <20 weeks). Prior studies have observed inconsistent associations between NSAID use and SAB risk, including racial differences. Given that response to medications may vary across individuals due to genetic differences, we conducted a genetic study evaluating NSAID by gene interactions for risk of SAB in a cohort of pregnant women. Women were enrolled in the *Right from the Start* (2004-2010) cohort. Periconceptional prescription and over-the-counter (OTC) NSAIDs reported up to the sixth week of pregnancy were obtained from study interviews. We evaluated ancestry using 100 ancestry informative markers and 201 haplotype tags; candidate SNPs were selected from NSAID metabolism pathways using 600 European American women. Interaction analyses between NSAID use and SNPs were conducted in logistic regression, adjusted for BMI, SAB history, and ancestry. NSAID exposure was reported by 63% of SAB cases and 62% controls. The strongest interactions were observed in multiple SNPs within the prostacyclin synthase (*PTGIS*) gene, the majority of which were in strong linkage disequilibrium (LD). The strongest interaction was at *PTGIS* rs5602 (OR 0. 34, 95% CI 0. 19 to 0. 60, p=2. 45x10-4) in fully adjusted models, and was statistically significant after a Bonferroni correction for multiple testing. NSAID stratified rs5602 single SNP additive association analyses that modeled A as the minor allele showed that NSAID users with the AA genotype were most protected from SAB (OR=0. 78, 95% CI 0. 56 to 1. 10), while non-NSAID users were at increased risk of SAB (OR=2. 11, 95% CI 1. 35 to 3. 29). Nominally significant associations were also observed at *PTGDR* and *CYP2A6* (p<5x10-3). Our findings suggest that risk for SAB due to NSAIDs in early pregnancy may be modified by *PTGIS*. However, further validation of our associations is necessary.

2114T

Association of G1733A polymorphism (rs6152) in androgen receptor gene with Idiopatic Recurrent Pregnancy Loss in Mexican women. *A. Porras[1, 2], A. LAZCANO[3], T. DA SILVA[1], J. JUÁREZ[1,2], J. García[2].* 1) Human Genetics, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico, JALISCO, Mexico; 2) 2Division de Genetica, Centro de Investigacion Biomedica de Occidente, Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, Mexico; 3) Servicio de Ginecoobstetricia, Hospital General 180, Instituto Mexicano del Seguro Social, Tlajomulco de Zuñiga, Jalisco, Mexico.

AbstractBackground. Idiopatic Recurrent Pregnancy Loss (IRPL) is defined as two or more repeated pregnancy losses before the 20th week. Its etiology is multifactorial and some genetic polymorphisms have been associated to risk of IRPL. The G1733A polymorphism (rs6152, *Stu*I polymorphism) in the androgen receptor (AR) gene can influencein the regulation of the transcription of androgen responsive genes. **Objective.** To determine whether the G1733A polymorphism is associated with IRPL. **Methods.** We analyzed the G1733A polymorphism in the *AR* gene by polymerase chain reaction (PCR)-restriction fragment length polymorphism (PCR-RFLP) in 156 IRPL women, (≥ 2 miscarriages) without identifiable risk factors and 152 unrelated fertile women, controls (≥ 2 live births). The ages of individuals in both groups ranged from 18 to 42 years. **Results.** The genotypic frequencies of the G1733A polymorphism in controls were in accordance with Hardy-Weinberg equilibrium (HWE) (p=0. 496413). The observed frequencies of GG, GA and AA genotypes were 70. 51%, 26. 92% and 2. 56%, respectively, for the patients and 88. 82%, 10. 53% and 0. 66%, respectively, for the controls (OR=3. 321; C. I=[1. 803-6. 116]; chi2=15. 85; p=0. 00007). Allele frequencies of the G1733A polymorphism among patients and controls were 0. 840 and 0. 941, respectively, for the wild type allele (G) and 0. 160 and 0. 059, respectively, for the mutant type allele (A) (OR=3. 032; C. I=[1. 725-5. 331]; chi2=16. 01; p=0. 00006). **Conclusion.** This study suggests that G1733A polymorphism in the *AR* gene is strongly associated with IRPL in Mexican women (OR=3. 032). .

2115F

Pattern of cytogenetic aberrations in Tunisian couples with recurrent miscarriage. *R. Frikha, N. Bouayed Abdelmoula, T. Rebai.* Histology Laboratory, Univ Medicine, Sfax, Tunisia.

INTRODUCTION: Recurrent miscarriage (RM), as at least three or more consecutives miscarriages less than 24 weeks of gestations, is an heterogeneous condition with affect approximately 1% of couples who are trying to conceive. Chromosomal rearrangements are well recognized cause of miscarriage in couples with RM with an overall frequency varying from 2% to 8% of couples with RM. This study was conducted to assess the prevalence of chromosomes abnormalities and clinical characteristics of couples with history of RM from Tunisia. Further, to delineate strategies for genetic evaluation and clinical management in subsequent pregnancy. METHODS: We investigated the karyotype of 98 couples with more than 3 miscarriages, referred to our genetic counseling. Chromosomal analysis was performed using RGH banding in peripheral blood. In all the cases the detailed reproductive case histories were taken. RESULTS: Of a total of 98 couples, 4 men and 1woman had abnormal karyotype with an overall incidence of (5. 1%). The commonest abnormalities appear to be reciprocal translocation. Chromosome abnormalities seem to be more prevalent in younger couples. In addition; personal reproductive background have an impact on the risk of been carrier of structural abnormalities. CONCLUSION: These data strengthens the current guidelines for the cytogenetic evaluation of couples with RM after the third miscarriage. However, depending on the presence of other factors such as personal reproductive background, karyotype should be proposed after 2 miscarriages. For the couples with chromosomes abnormalities, a non invasive prenatal testing should be discussed.

2116T

Interleukin 1 Beta Gene Polymorphisms in Women with Early Onset Preeclampsia. *A. Venkateshwari[1], S. Sowmya[1], R. Aruna[2], N. Pratibha[3], A. Jyothy[1].* 1) Cell Biology, Institute of Genetics and Hospital for Genetic Dis, Hydrabad, Telangana, India; 2) Government Maternity Hospital, Petlaburz, Hyderabad,Telangana, India; 3) Department of Genetics, Osmania University, Hyderabad, Telangana, India.

Purpose of the study: Preeclampsia is a pregnancy-specific disorder characterized by hypertension and systemic endothelial dysfunction. Interleukin (IL)-1b is a possible mediator of maternal endothelial dysfunction in preeclampsia. Earlier studies reported increased levels of serum IL-1b as well as its natural inhibitor IL-1 receptor antagonist (IL-1Ra) in women with preeclampsia. In view of the above, the purpose of the present study is to evaluate the relationship between IL-1 beta polymorphisms +3954 C/T (rs 1143634), -31 T/C(rs1143627), receptor IL1Ra VNTR polymorphism and preeclampsia (PE), and analyze haplotype frequencies of the three polymorphic loci. **Methods:**A total of 140 patients and 135 women with normal pregnancy, from Government Maternity Hospital, Petlaburz, Hyderabad, India, were considered for the present study. PCR- RFLP was carried out for IL-1B +3954 and -31 genotyping and PCR for IL-1Ra, followed by agarose gel electrophoresis. Genotype and allele frequencies between case-control groups were compared by chi-square (χ^2) tests. Haplotype analyses was carried out with the help of Haploview software . **Results:**Significant statistical differences were found between PE and control groups regarding genotype and allele frequencies of the two polymorphisms of IL-1b and IL1Ra (For IL-1b -31C/T: χ^2= 21. 17, p=0. 000004; For IL-1b+3954: χ^2= 28. 24, p=0. 0001and IL 1Ra : χ^2= 13. 57 , p=0. 0002). Further, haplotype analysis of these SNPs revealed a significant association of STC, LTC and LTT haplotypes with patients compared to controls. **Conclusion:**The present study suggests that IL-1B cluster polymorphism could be a major genetic regulator in the pathogenesis of preelampsia.

2117F

Hypertension as a risk factor for uterine fibroids: a Mendelian randomization analysis. *K. S. Tsosie[1], D. R. Velez Edwards[1,3], T. L. Edwards[1,2].* 1) Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN; 2) Division of Epidemiology, Department of Medicine, Vanderbilt University, Nashville, TN; 3) Department of Obstetrics and Gynecology, Vanderbilt University, Nashville, TN.

Uterine fibroids [OMIM: 150699] are common benign tumors that affect up to 77% of women by menopause and account for $34. 4 billion in annual U. S. healthcare costs. Fibroids are highly heritable with greater risk for women of African ancestry. Clinical and biological studies have reported a relationship between arterial hypertension and uterine fibroids but, to date, there has not been a published analysis associating causal genetic factors for hypertension with fibroids. Aggregate genetic risk scores (GRS) were calculated using 35 markers significantly associated (p>5x10-7) with hypertension (HTN) and high blood pressure (BP) from well-powered, independently replicated studies listed in the NHGRI GWAS Catalog. A Mendelian randomization analysis was conducted using imputed genome-wide data from 1,358 unrelated African American women with hypertension and image-confirmed fibroid status determined from electronic medical records. Logistic regression was used to evaluate whether HTN and BP are genetically associated with uterine fibroids. Women with greater numbers of hypertension and BP-increasing alleles had a higher uterine fibroids risk (OR=1. 03, 95% CI=1. 00-1. 06, p=0. 08). However, this association was not seen when only hypertension-associated alleles were included in the model. Hence, Mendelian randomization suggests a potential association between hypertension via blood pressure-increasing alleles and uterine fibroids in our defined study population of African American women.

2118T

Role of Kallmann Syndrome Gene *NELF* in the JAK-STAT Pathway. *E. Ko, L. Chorich, M. Sullivan, S. Quaynor, R. Cameron, L. Layman.* Department of Neuroscience &Regenerative medicine, Georgia Regents University, Augusta, GA.

Normosmic hypogonadotropic hypogonadism (nHH) and Kallmann syndrome (KS) represent two related disorders of human puberty, which share in common a deficiency of gonadotropin releasing hormone (GnRH). Proper migration of GnRH neurons from the olfactory placode to the hypothalamus is an important prerequisite for normal hypothalamic-pituitary–gonadal (H-P-G) axis function, which plays a critical role in pubertal development. Human mutations in nasal embryonic LHRH factor (*NELF*), which is also known as NMDA receptor synaptonuclear signaling and neuronal migration factor (*NSMF*), have been identified in nHH/KS patients. In addition, *Nelf* knockdown has been reported to disrupt GnRH neuron migration *in vivo* and *in vitro*. The mechanism by which NELF regulates GnRH neuron migration is unclear. We hypothesize that NELF could be involved in GnRH neuron migration either directly or indirectly through regulation of other genes or proteins. NELF has been reported predominantly localized in the nucleus and contains two atypical zinc finger domains. Jacob protein, the rat ortholog, is 98 % identical to NELF, and following NMDA receptor activation, it translocates to the nucleus and interacts with CREB, suggesting a role in transcription. In the present study, cDNA arrays were performed on RNA from migratory GnRH neuronal (NLT) cells following *Nelf* knockdown (KD) using Nelf shRNA lentiviral particles. Ingenuity pathway analysis (IPA) indicated the JAK-STAT pathway as one of the most relevant signaling pathways, since *Stat1, Stat2, Stat5a, Jak2*, and *Irf 9* were down regulated in the *Nelf* KD sample compared to a scrambled control. To validate these genes, RT-qPCR and immunoblot were performed. By RT-qPCR, all five genes were down-regulated by 48-82% following *Nelf* KD. Protein levels of STAT1 and JAK2 were decreased in immunoblots by 87% and 55 %, respectively. Our findings suggest that the nHH/KS gene product of NELF is involved in the JAK/STAT signaling pathway, which could be a potential mechanism whereby NELF regulates GnRH neuronal migration and function.

2119F

Immotile Short Tail Sperms (ISTS) defect: Genetic evaluation in-depth of alterations in essential candidate genes. *A. Mohseni Meybodi[1], M. Sabbaghian[2], H. Hosseinifar[2], H. Hosseini[2], T. Modarresi[2], H. Gourabi[1], M. A Sadighi Gilani[2].* 1) Department of Genetics at Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran; 2) Department of Andrology at Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran.

ISTS defect in which sperm tail is short and fibrous sheath and axoneme are disorganized, is one of the syndromes that cause male infertility. Despite the studies have been made in this regard, its exact etiology in human is unclear yet. Four candidate genes causing ISTS are*SPEF2*, *RABL2B,* and A-kinas anchoring proteins genes (*AKAP3*and*AKAP4*). Proteins which coded by *SPEF2*and *RABL2B* are essential for correct sperm tail assembly and development, besides, AKAP3andAKAP4 are most abundant structural proteins in the fibrous sheath. In the present study, the variations of candidate genes were investigated in 35 infertile men with ISTS and 40 fertile men. In this purpose, DNA was extracted from peripheral blood of the patients (with more than 80% short tail sperms in at least two spermograms) and control group. Primers were designed for each target segment of candidate genes, and then PCR sequencing was carried out. Sequence analysis of*SPEF2*did not identify any mutation in exon three and 28. However, one polymorphism (363A>C) was identified in exon 3 in four of patients and three of controls (P>0. 05). Analysis of genetic data revealed that no mutations or single-nucleotide polymorphisms in exon four of *RABL2B*was identified, but an intronic variant [(C) nucleotide deletion (rs: 144944885)] was found in heterozygote form in five patients (P<0. 05). No alteration was identified in controls. Moreover, the 1499T>C polymorphism in*AKAP3*was just seen in five patients whereas none of the individuals in the control group had this alternation (P<0. 05). No genetic alteration but one was found in*AKAP4*gene which was a >1350 bp deletion in exon 5. Our results revealed that some specific gene alterations in*AKAP3*,*AKAP4*and*RABL2B*can take part an important role in sperm tail malformation, so it can be assumed as the etiology of ISTSalthough genetic alterations in*SPEF2*gene are not involved in this defect.

2120T

Genetic basis of obstructive azoospermia: Whole Exome Sequencing provides new insights into Congenital Absence of Vas Deferens. *O. Patat[1], A. Pagin[2], A. Siegfried[3], N. Chassaing[1], L. Monteil[1], V. Gaston[1], L. Bujan[4], M. Courtade-Saidi[3], F. Marcelli[5], G. Lalau[2], JM. Rigot[5], R. Mieusset[4], E. Bieth[1].* 1) Service de Génétique Médicale, Hôpital Purpan, TOULOUSE 31059, France; 2) UF Mucoviscidose, Service de Toxicologie et Génopathies, Centre de Biologie Pathologie Génétique, CHRU LILLE, Boul. du Prof. Jules Leclercq, 59037- LILLE, France; 3) Service d'Anatomopathologie, Institut Universitaire du Cancer (IUC), 1 avenue Irène Joliot Curie, 31059 TOULOUSE, France; 4) Service d'Andrologie, Hôpital Paule De Viguier, avenue de Grande-Bretagne, 31059 TOULOUSE, France; 5) Service d'Urologie et Andrologie, CHRU Lille, boulevard Professeur Leclercq, 59037 LILLE, France.

Introduction. Among the causes of male infertility, congenital bilateral absence of vas deferens (CBAVD) is a common cause of congenital obstructive azoospermia. Mutations in the gene *CFTR* are identified in nearly 80% of cases. For the remaining 20% the origin of the CBAVD is unknown and other genetics causes are strongly suspected. Thus, other genetic involvement is suggested, in a subset of these infertile men, by an increased frequency of unilateral renal agenesis. However, with the exception of *CFTR*, no other gene responsible for CBAVD has been identified to date. Due to the evident genetic consequence of infertility and as a result of psychological and sociocultural aspects, it is hard to obtain the multiplex pedigrees or complete trios which are required for family-based approaches of causal gene identification. To overcome this difficulty we compared the exomes of 23 unrelated CBAVD cases without any *CFTR* mutations detected after extensive gene screening. All the included cases were selected on the basis of a strict homogenous phenotype of CBAVD and were divided into 2 groups according to their renal status (agenesis or not). **Methods.** A total of 84 medical files of infertile men with CBAVD and a negative *CFTR* mutation screening result were reviewed by clinical experts in andrology. 23 unrelated patients who fulfilled the stringent phenotypic criteria underwent individual whole exome sequencing (WES). **Results.** After filtering, several hemizygous or heterozygous loss-of-function mutations were identified in at least two patients of the cohort. Some of these mutations affected compelling candidate genes, given their expression in the male genital tract, such as *MAGEC1, OTOG, CCDC168, ADGRG2* and *TP53TG5*. Moreover, immunohistochemical evidence was obtained for certain genes showing altered expression in the genital tract of mutated patients compared to normal subjects. **Conclusion.** The WES results strongly support genetic heterogeneity in CBAVD and offer new insights into the pathophysiology of obstructive azoospermia. Further investigations are in progress on larger cohorts of infertile men.

2121F

Molecular analysis of Y-chromosome micro-deletions and its relationship with male infertility in Zanjan, Yazd and East Azerbaijan province of Iran. *M. r Ranjouri[1], A. Biglari[1], P. Aob[1], M. Shekari Khaniani[2], M. h Sheikhha[3].* 1) Genetics and Molecular Medicine Department, Zanjan University of Medical Science, Zanjan, Iran; 2) Medical Genetic Department, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran; 3) Medical Genetic Department, Faculty of Medicine,Yazd University of Medical Sciences,Yazd,Iran.

Background:After Klinefelter's syndrome, microdeletion of Yq (YCMD) is the most common genetic cause of male infertility; Three regions in the long arm of the Y chromosome, known as AZFa, AZFb, and AZFc, are involved in the most frequent patterns of Y chromosome microdeletions. These regions contain a high density of genes that are thought to be responsible for impaired spermatogenesis. 15% of azoospermic or 5-10% of oligozoospermic males have Yq deletions. **Objective**: Our study aimed to determine the prevalence and type of Yq microdeletions and investigate and detect the Azoospermia Factor (AZF) region deletion, rearrangement and deleted azoospermia (DAZ) gene copy number variants in 3 province of Iran. **Materials and Methods:** A sample of 250 Iranian infertile males (118 with azoospermia and 132 with severe oligozoospermia) was screened for microdeletions and 250 fertile men were also studied as control group. DNA extraction and molecular analysis were done on blood samples. Multiplex-PCR (M-PCR) method was done with 14 STS primers were selected to determine Y-chromosome microdeletions to identify the presence of microdeletion in AZFa, AZFb or AZFc region and also sY587/DraI PCR-RFLP assay were used for determine DAZ copy number variants. **Results:** The Multiplex PCR method identified 21 (8. 4%) infertile males with Y chromosome microdeletions, while none was found in the controls. Amongst the AZF subregions, three cases had deletions in AZFa (1. 2%), seven in AZFb (2. 8%) and eleven in AZFc (4. 4%). the ratio of Y chromosome microdeletion in azoospermic men was higher than this ratio in oligospermic men [11% (13/118) among azoospermic men and 6% (8/132) among severe oligospermics] and also results revealed a significant prevalence of AZFc subtypes deletion and reduced DAZ gene dosage in Iranian azoospermic cases affecting Y chromosome deletion. **Conclusion:** Based on the present study, the frequency of microdeletion in Iranian population is 8. 4% and This study is the first one to investigate AZFc subtypes deletion and DAZ genes dosage in Iranian infertile men and also we concluded that DAZ gene deletion is a risk factor for spermatogenic failure.

2122T

Evaluation of the expression of *AR* and *SOX3* genes in sertoli cells of azoospermic patient. *M. Sabbaghian[1], A. Mohseni Meybodi[2], F. Sangsefidi[1], T. Modaressi[1], M. A. Sadighi Gilani[1].* 1) Department of Andrology at Reproductive Biomedicine Research Center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran; 2) Department of Genetics at Reproductive Biomedicine Research center, Royan Institute for Reproductive Biomedicine, ACECR, Tehran, Iran.

Sertoli cells located in seminiferous tubular basement membrane and surrounds different types of germ cells. Androgen Receptor (AR) is one of the most important transcription factors in sertoli cells, which binds to DNA and influences expression of genes involved in spermatogenesis. The final distinction and separation of spermatids from seminiferous epithelium is dependent and maximally sensitive to AR depletion within the testis. Moreover, during spermatogenesis, expression of *SOX3*, a transcription factor, expressed in sertoli cells, is necessary for differentiation of spermatogonia type A. The purpose of this study was to separate and purificate sertoli cells from human testis and to investigate the expression of *AR* and *SOX3* genes in sertoli cells of azoospermic patients with different spermatogenesis status. Biopsies were obtained from 10 men who referred to the Royan institution and underwent testicular sperm extraction (TESE). Tissue samples were transferred to lectin coated petri dishes after enzymatic dissociation and isolation. After few passages, all the cells were harvested and the cell type was confirmed by immunocytochemistry. *AR* and *SOX3* gene expression level was determined by real time RT-PCR. Isolation, purification and culture of the human sertoli cells were performed successfully. It was shown that *AR* and *SOX3* genes are expressed in these cells and there is significant difference in the expression of these genes in sertoli cells derived from tissues with successful sperm extraction (TESE+) compared to samples without sperm (TESE-). The results showed that the expression of *AR* gene in human sertoli cells could play an important role in spermatogenesis and its expression level could be related to the outcome of sperm extraction. Also, the expression of *SOX3* gene in (TESE-) is more than (TESE+), that indicate negative effect over expression of *SOX3* gene in spermatogenesis.

2123F

Low levels of PAPP-A as a sole parameter predicting high risk of aneuploidy in the fetus in pregnant women younger than 35 and NT <3mm. *A. Singer[1], V. Nadler[2], V. Katalan[2], M. Shohat[2], R. Berger[2], C. Vinkler[3].* 1) Dept Clinical Gen Unit, Barzilai Med Ctr, Tel Aviv, Israel; 2) Maccabi Health Services Rechovot Israel; 3) Institute of Medical Genetics,Wolfson Medical Center, Holon. Israel.

First trimester Down syndrome screening test is widely used in Israel and is part of the national screening program during pregnancy. The tests include measurement of nuchal translucency (NT), plasma free-B-hCG and plasma level of pregnancy- associated protein-A (PAPP-A). Low level of PAPP-A has been previously shown as predictor of adverse outcome of pregnancy and increased risk for aneuploidy. The aim of this work is to assess the incidence of pregnancies with aneuploidies in a group of women younger than 35y with NT measuring below 3 mm and low levels of PAPP-A as a sole parameter predicting high risk of aneuploidy in the fetus. PAPP-A levels of 0. 2 MoM or less are infrequently encountered in clinical practice, being seen in no greater than 0. 5% of the population. This level has been considered as a cut off value considering the need for both further evaluation and close monitoring of the pregnancy. Data on consecutive patients attending a first trimester screening program were collected. Of those with PAPP-A levels ≤ 0. 2 MoM only women under 35y of age with NT< 3mm were considered. First trimester screening was performed on 100,483 women with a live singleton pregnancy between 2006 and 2013. The PAPP-A level was 0. 2 MoM or less, in 360 pregnancies (0. 36%). Of these we were able to obtain information regarding karyotype for 93 women so far (25. 8%). The incidence of karyotype abnormality in this group was 11 cases (11. 8%), including trisomy 21(7/11), trisomy 18 (2/11), 47,XXY (1/11) and one case of triploidy. We are currently analyzing further information and more accurate results will be presented in our final report. Non-invasive prenatal testing (NIPT) is currently recommended in some countries to women at a high risk for aneuploidy (>1:200). Based on initially partial result, it is suggested that this high risk group should include women with low PAPP-A levels (<0. 2MOM) as a single abnormal parameter.

2124T

Noninvasive prenatal diagnosis of Huntington disease: detection of the paternally inherited expanded CAG repeat in maternal plasma. *M. J. van Belzen[1], J. M. E. van den Oever[1], E. K. Bijlsma[1], I. Feenstra[2], N. Muntjewerff[3], I. B. Mathijssen[4], E. Bakker[1], E. M. J. Boon[1].* 1) Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Human Genetics, Radboud University Medical Center, Nijmegen, The Netherlands; 3) Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands; 4) Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands.

Huntington disease is an autosomal dominant progressive neurodegenerative disorder, characterized by irrepressible motor symptoms, cognitive impairment, and psychiatric disturbances. It is caused by a CAG repeat expansion in exon 1 of the HTT gene. There is currently no cure for the disease. Prenatal testing is available for couples who do not wish to transmit the disease to their offspring. With a shift towards noninvasive testing, we have explored and validated the use of noninvasive prenatal diagnosis (NIPD) for HD. Fifteen couples have been included, assessing a total of n = 20 pregnancies. Fetal paternally inherited CAG repeat length was determined in total cell-free DNA from maternal plasma using a direct approach by PCR and subsequent fragment analysis. All fetal HD (n = 7) and intermediate (n = 3) CAG repeats could be detected in maternal plasma. Detection of repeats in the normal range (n = 10) was successful in n = 5 cases where the paternal repeat size could be distinguished from maternal repeat patterns after fragment analysis. In all other cases (n = 5), the paternal peaks coincided with the maternal peak pattern. All NIPD results were concordant with results from routine diagnostics on fetal genomic DNA from chorionic villi. In conclusion, we demonstrated that all fetuses at risk for HD could be identified noninvasively in maternal plasma. Additionally, we have confirmed results from previously described case reports that NIPD for HD can be performed using a direct approach by PCR. For future diagnostics, parental CAG profiles can be used to predict the success rate for NIPD prior to testing.

2125F

Analytical Validation of a Novel Next-Generation Sequencing Based Pre-implantation Genetic Screening Technology. *M. Umbarger[1], J. Gole[1], A. Gore[1], E. Boyden[1], M. Slevin[1], B. Milette[2], M. Henein[2], P. Saunders[1], D. Maganzini[1], G. Celia[2], G. Porreca[1].* 1) Good Start Genetics, Cambridge, MA; 2) Dominion Fertility, Arlington, VA.

Current data indicate that pre-implantation genetic screening (PGS) increases pregnancy and decreases miscarriage rates. However, high cost has limited its adoption. This may be overcome by utilizing more efficient and streamlined workflows employing next-generation DNA sequencing (NGS). We have developed a novel method for PGS that couples NGS with a technology termed FAST-SeqS which utilizes a PCR reaction to amplify repetitive elements from all chromosomes while simultaneously attaching sequencing adapters and sample specific molecular barcodes. To evaluate the analytical accuracy of our new NGS-based PGS approach we used cells derived from both previously karyotyped primary fibroblasts and trophectoderm (TE) biopsies of blastocyst stage embryos. In the cell-line studies, five-cell samples from two euploid (46,XY; 46,XX) and six aneuploid (47,XY,+9; 47,XY+13; 47,XX,+18; 47,XY,+21; 47,XXY; 47,XYY) lines were run through the developed assay and the resulting molecular karyotype calls were compared to the reported karyotypes. 168 aneuploid and 55 euploid samples were correctly called, yielding 100% sensitivity and specificity. In the embryo-based studies, paired TE biopsies were obtained from previously cryopreserved blastocysts. One biopsy was tested using FAST-SeqS and the other was tested using the Illumina 24sure array comparative genomic hybridization (aCGH) approach, one of the most commonly utilized PGS technologies. 24 aneuploid and 30 euploid samples were correctly called, and one euploid sample was incorrectly called, yielding sensitivity and specificity estimates of 100% (95% CI 86. 2 - 100%) and 96. 8% (95% CI 83. 9 - 99. 8%), respectively. One false positive and two false negative chromosome calls were made; the latter two occurred in a sample correctly classified as aneuploid, and the associated raw copy number calls were close to our calling threshold, suggesting that the embryo may have been mosaic aneuploid. Together, our cell-line and TE-biopsy-based data indicate that our NGS-based PGS approach yields molecular karyotype calls that are both highly accurate and consistent with traditional aCGH calls.

2126T

Characterization of the mtDNA bottleneck in human preimplantation embryos and zebrafish germline and non-germline cells. *A. Otten[1], S. Sallevelt[1], J. Dreesen[1], C. de Die[1], A. Paulussen[1], T. Theunissen[1], M. Winandy[2], E. Lambrichs[1], I. Eijkenboom[1], M. Gerards[1], J. Vanoevelen[1], K. Tarbashevich[3], E. Raz[3], D. Samuels[4], B. Van den Bosch[1], M. Muller[2], H. Smeets[1].* 1) Clinical Genetics, GROW, MUMC , Maastricht, Netherlands; 2) GIGA, Université de Liège, Liège, Belgium; 3) Institute of Cell Biology, Münster University, Germany; 4) Vanderbilt University School of Medicine, Nashville, USA.

In most animals, the mitochondrial DNA (mtDNA) inherits through a segregational bottleneck, during which a limited number of mtDNA molecules is passed from oocytes to the offspring. However, size, timing and possible differences between germline and non-germline cells of this bottleneck are still debated. To get more insight in this, we collected data of individual human PGD embryos from carriers of the m. 3243A>G or m. 8993T>G mtDNA mutation, as well as of zebrafish embryos. In the human oocytes/embryos, the m. 3243A>G mutation load showed in most cases a Gaussian distribution for individual carriers, whereas the m. 8993T>G was fundamentally different. In one m. 3243A>G carrier two populations with regard to the heteroplasmy level were observed. From this data we estimated the bottleneck sizes, based on the Kimura distribution, which ranged from 48-93 (m. 3243A>G) and 11 (m. 8993T>G, one carrier). This is somewhat lower than the minimum mtDNA counts in embryogenesis in mice (about 200) and selection mechanisms cannot be excluded. Oocytes and PGCs were collected from zebrafish embryosduring development. We measured, on average, 21. 0x106 mtDNA molecules (range: 5. 1x106 – 42. 2x106) in zebrafish oocytes, which is higher than in mammals, probably due to the absence of implantation during teleost development. Oocytes from different females had similar mean mtDNA copy number values with high intra-individual variation (CVs: 28-49%). During development, the mtDNA copy number per cell decreased until the segmentation period, suggesting the onset of mtDNA replication. The mtDNA copy number in FACS-sorted primordial germ cells (PGCs) from different embryonic stages was, at lowest, on average, ~170 (range: 65-287) mtDNA molecules in PGCs, similar to the mice germline bottleneck. For FACS-sorted non-PGCs, the lowest value was ~50 (range: 18-88) mtDNA molecules. Our data suggest that the lowest mtDNA copy number is reached after a fixed developmental stage, implying that considerable variation will exist among PGC and non-PGCs of the same female fish. If similar in humans, this could provide an explanation for theoccurrence of (*de novo*) mutations that reach high mutation loads within a single generation. Early differentiation of PGCs during the developing embryo most likely causes the observed differences in bottleneck size, which would suggest different segregation of familial mtDNA mutations in germline and non-germline cells.

2127F

Preimplantation genetic diagnosis for male infertility with chromosomal rearrangement. *C. C. Huang, S. U. Chen, Y. S. Yang.* National Taiwan University Hospital, Taipei, Taiwan.

Infertile men with chromosomal rearrangement (CR) are highly vulnerable to produce unbalanced gametes which result in recurrent miscarriage, affected offspring or infertility. Preimplantation genetic diagnosis (PGD) with blastomere biopsy and fluorescent in-situ hybridization (FISH) had been used to select normal/balanced embryos for transfer. However, FISH is inherent with some technical difficulties such as cell fixation and signal overlapping. Here weapplied a strategy of PGD using blastocyst biopsy and array comparative genomic hybridization (aCGH) for male reproductive problems with CR. A total of eight men diagnosed as CR were included. Blastocyst biopsy was performed and biopsied blastocysts were cryopreserved individually. Testing was performed with aCGH and the euploid embryos were transferred in the following thawing cycles. The overall diagnostic efficiency is 88. 7% (47/53) and the euploidy rate was 34% (16/47). Six cycles of embryo transfer (ET) were carried out, resulting inthree live births with a pregnancy rate of 50% per transfer cycle. Two couples had no euploid embryos for transfer. Blastocyst biopsy retrieved more genetic material and may provide more reliable results. Meanwhile, aCGH offered not only detection of chromosomal translocation but also more comprehensive 24 chromosomal analyses than traditional FISH. Our study demonstrates an effective PGD strategy of blastocyst biopsy, aCGH and thawed ET for male reproductive problems with CR.

2128T

Prenatal Caffey Disease: A Tale of Two Siblings. *M. S. Saleh[1,2], P. S. Shannon[3], A. T. Toi[4], R. S. Silver[2], R. T. Teitelbaum[2], S. K. Keating[3], O. N. Nevo[5], D. C. Chitayat[1,2].* 1) The Hospital for Sickkids, Department of Pediatrics; 2) Division of Clinical and Metabolic Genetics; Mount Sinai Hospital, Department of Obstetrics and Gynecology; 3) The Prenatal Diagnosis and Medical Genetics Program; 4) Department of Laboratory Medicine and Pathobiology; 5) Department of Diagnostic imaging; Sunnybrook Health Science Center.

Caffey disease or cortical hyperostosis is a condition characterized by massive subperiosteal new bone formation usually involving the diaphyses of the long bones. Two forms of Caffey disease have been identified 1) The classical mild, transient infantile form typically associated with fever, joint swelling and pain with onset around age two months and 2) the Lethal prenatal form which typically presents before 35 weeks' gestation, characterized by cortical hyperostosis, short, thick and angulated long bones associated with polyhydramnios and pulmonary hypoplasia. We report two cases of prenatal Caffey disease born to healthy and non-consanguineous parents of Indian-Sikh descent. **Case 1**: The mother was 26 years old G2P1L1 and the father was 28 years old. The couple was healthy and had a son who was well. The pregnancy was initially uncomplicated. Detailed fetal ultrasound at 20 weeks gestation was normal and there was no exposures to teratogens. At 30 weeks gestation an ultrasound revealed polyhydramnios, all long bones were below the 5th centile in length, angulated and thick and there was scalp edema. Echocardiography was normal. No hydrops fetalis was detected. The couple decided to terminate the pregnancy and the autopsy showed cortical hyperostosis of all long bones. The fetal autopsy showed a female fetus with X-rays and histopathology consistent with cortical hyperostosis. **Case 2**: The couple embarked on their 3rd pregnancy and this was initially uncomplicated. A fetal ultrasound at 29 weeks gestation showed polyhydramnios, all long bones were thick, bowed and measured below the 5th centile, the chest was narrow, there was frontal bossing with forehead edema, echogenic gut, absent stomach bubble, large IVC and a large and bulky placenta. The couple was counseled regarding the recurrence of cortical hyperostosis and decided to terminate the pregnancy. The fetal autopsy showed a male fetus with X-rays and histopathology consistent with cortical hyperostosis. DNA analysis of the*COL1A1* gene and a whole exome sequencing (WES) showed no detectable abnormalities. The occurrence of the same condition in sibs of both sexes born to unaffected parents is consistent with an autosomal recessive mode of inheritance. However, the negative WES raises the possibility that this prenatal and lethal condition is the result of an autoimmune condition caused by maternal antibodies directed towards a component of the fetal skeleton.

2129F

Prenatal presentation of an inherited chromosome 6p22. 3 microdeletion, encompassing the JARID2 gene. *M. Biderman Waberski[1], H. Al-Kouatly[2], MH. Fries[2].* 1) National Human Genome Research Institute, National Institute of Health, Bethesda, MD; 2) Maternal Fetal Medicine & Genetics, MedStar Washington Hospital Center, Washington, DC.

Microdeletions in the short arm of chromosome 6 are relatively rare and patients demonstrate varying degrees of intellectual disability and facial dysmorphism. The gene *JARID2* resides on chromosome 6p22. 3 and has been proposed as the most likely gene contributing to this phenotype. We describe an affected mother, daughter and fetal presentation of a 393kb microdeletion in chromosome 6p22. 3 encompassing *JARID2*. This is the first report of inheritance of a microdeletion in this region, displaying variable expressivity. The mother is a 28 year old woman G4P1021. Her personal medical history includes bilateral club feet, meningitis as an infant, blepharophimosis, and a madelung deformity as well as a learning disability. Her 6 year old daughter also had bilateral club feet, an ASD and VSD requiring surgical repair, a madelung deformity, ptosis of her left eyelid, and significant developmental delay including motor and speech delays. Genetic evaluation for this daughter revealed the microdeletion in chromosome 6p22. 3 (15,340,206-15,740,872) via SNP microarray analysis. This mutation was assumed to be de novo in the daughter until fetal anomalies were detected in a subsequent pregnancy, including a large omphalocele, a mildly enlarged nuchal translucency at 2. 5 mm and a short nasal bone. Amniocentesis revealed the presence of the same microdeletion, which was confirmed to be inherited from the mother. Additional prenatal imaging showed IUGR, bilateral clubbing of the feet, VSD, small lens cataract, scoliosis, and kyphosis. A fetal MRI showed that the omphalocele contained the entire liver, stomach and small bowel, measuring up to 10. 4cm x12. 8cm. He was delivered via scheduled c-section at 38 weeks, with a normal APGAR score and a birth weight of 2557g (5th %ile). In addition to prenatal findings he was hypotonic, however no major dysmorphic features were appreciated. This is the first description of a child with a major abdominal wall defect and a microdeletion in chromosome 6p22. 3. The microdeletion overlaps *JARID2.* Haploinsuffiency of this gene has been associated with features present in this family including impaired cognitive function and variable dysmorphic facial features. This case further reveals the phenotypic variability of this deletion syndrome and expands the phenotype to include major organ defects, including abdominal wall defects and heart defects.

2130T

Prenatal whole genome SNP array: frequency and relevance of incidental diagnoses in parental samples. *L. Govaerts[1], D. Van Opstal[1], K. Diderich[1], M. Joosten[1], S. Riedijk[1], A. Prinsen[2], F. de Vries[1], R. Galjaard[1], M. Srebniak[1].* 1) Dept Clinical Gen, Erasmus Med Ctr, Rotterdam, Netherlands; 2) Dept Obstetrics and Gynaecology, Erasmus Med Crt, Rotterdam, Netherlands.

Background: We routinely perform SNP array analysis as a first-tier test for all prenatal indications. In cases with ultrasound abnormalities, fetal and parental samples are simultaneously tested, facilitating the interpretation of fetal array results, in order to ensure short reporting times and avoid parental anxiety. However, by doing so there is a chance to detect a chromosomal aberration in parental DNA samples. These so-called incidental diagnoses are abnormalities found by chance, unintentionally, in parents of probands (Srebniak et al. , Eur J Hum Genet 2014). The aim of this study is to investigate the frequency and nature of these incidental diagnoses in order to judge whether our current protocol needs to be changed. **Methods:** In 2010-2014, 1,682 pregnant women were referred for prenatal SNP array testing because of ultrasound anomalies. Only CNVs present in the fetus were investigated in the parental samples. For quality assessment, the whole genome plots (LogR ratio and BAF) were viewed and eventually the presence of large chromosomal imbalances was noticed. **Results:** An incidental diagnosis was found in 0. 36% (6/1682) cases:In three pregnant women, mosaic Turner syndrome was found. Their phenotype was apparently normal withoutfertility problems. They were referred to our Turner expertise clinic, for evaluation of health issues such as increased risk of cardiovascular symptoms, endocrinological pathology and premature ovarian failure. In one pregnant woman the array demonstrated a mosaic interstitial deletion in 2q37. 1. She had a history of recurrent miscarriages and this deletion could explain the miscarriages. In two partners an aberration associated with an increased risk for a hematological malignancy was found. One of these men was diagnosed with constitutional mosaic trisomy 8. He had neither dysmorphic features nor structural anomalies. The other partner had a mosaic partial trisomy of the long arm of chromosome 1, without any dysmorphic features. Both were referred for extensive hematological evaluation. No malignancy was found, but both will stay under hematological surveillance. **Conclusion**: Because the frequency of incidental findings in parental samples is low and all incidental diagnoses found in our cohort were relevant for health surveillance of the parent. In our opinion, there is no reason to change our protocol with regards to the quality check of the parental samples.

2131F

Thrombocytopenia Absent Radius (TAR) Syndrome due to compound heterozygosity for 1q21. 1 microdeletion and an hypomorpic mutation in *RBM8A* gene. *D. Barel[1], A. Orr-Urtreger[1, 2], A. Bar Shira[1], S. Bak[1], M. Busilov[1], Y. Yaron[1, 2], A. Reches[1, 2].* 1) Tel Aviv Medical Center, Tel Aviv, Israel; 2) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

IntroductionCongenital anomalies of the limbs can be diagnosed on early second trimester ultrasound. It is often challenging to reach a diagnosis in such cases. We describe a case of prenatally detected limb malformations in a fetus confirmed as TAR syndrome by molecular analysis. **Case presentation**A 26 old primipara was referred for genetic counseling due to bilateral agenesis of the radii and hypoplasia of the humeri and a small left hand detected by transvaginal sonography at 17 weeks of gestation. No joint movement was observed. The parents are of non-consanguineous Iranian and Ashkenazi Jewish origin, with no family history of genetic diseases. After a multidisciplinary consultation the couple opted to terminate pregnancy. Pathologic examination demonstrated a female fetus with bilateral abnormal development of the bones in both upper limbs including: short humeri, very short radii, curvature of the ulnae, small left hand. In addition, ear deformities and clitoromegaly were also noted. Molecular studies demonstrated compound hetrozygosity for a 1q21. 1 microdeletion including the *RBM8A* gene and a *RBM8A* hypomorphic mutation (c. -21G>A). This is a common variant located in the 5' UTR of the gene leading to reduced protein expression. Parental analysis demonstrated that the 1q21. 1 deletion was inherited from the father. The hypomorphic mutation is being tested in the mother. **Discussion**TAR syndrome is a rare autosomal recessive disorder characterized by upper limb defects including agenesis of the radius and humeral hypoplasia and thrombocytopenia, which is usually transient. Other skeletal manifestations include abnormalities of the lower limbs, ribs, and vertebrae. In some cases, additional malformations are found in the heart and genitourinary system (renal anomalies and agenesis of uterus, cervix, and upper part of the vagina). Most patients demonstrate submicroscopic deletion in chromosome 1q21. 1 including the *RBM8A* gene on one chromosome and a hypomorpic mutation in the *RBM8A* gene on the other copy. Some patients have bi-allelic intragenic mutations of *RBM8A* gene. The majority (50-75%) of the 1q21. 1 deletions are inherited from an unaffected parent. The second mutation in the *RBM8A* gene may be inherited from an unaffected parent or occur *de novo*. If the mother is found to be a carrier, the parents will opt for preimplantation genetic diagnosis (PGD) for subsequent unaffected pregnancies.

2132T

"It was the last thing on my mind": couples' experiences and needs for support following the diagnosis of a fetal abnormality in pregnancy. *S. A. Metcalfe[1, 2], B. J. McClaren[1], P. Pitt[1], C. Hickerton[1], M. Menezes[1,3], J. Halliday[2,4], J. Fisher[5], K. Petersen[6], J. Hodgson[1,2].* 1) Genetics Education and Health Research, Murdoch Childrens Research Institute, Parkville, Melbourne, Victoria, Australia; 2) Dept Paediatrics, The University of Melbourne, Royal Children's Hospital, Parkville, Melbourne, Victoria, Australia; 3) Monash Ultrasound for Women, Melbourne, Victoria, Australia; 4) Public Health Genetics, Murdoch Childrens Research Institute, Parkville, Melbourne, Victoria, Australia; 5) Monash University, Clayton, Melbourne, Victoria, Australia; 6) La Trobe University, Bundoora, Melbourne, Victoria, Australia.

Background: The identification of a fetal abnormality in pregnancy can arise following screening and/or diagnostic testing. Despite this being a clinically common situation there are little empirical data about couples' experiences, especially following the diagnosis. This study aims to explore the psychosocial impact of receiving a prenatal diagnosis of a fetal abnormality and identify professional and social supports that are used and needed by couples. **Method**: A longitudinal case study approach with women and their partners from 3 clinical sites in Victoria, Australia. Quantitative (using depression, anxiety, stress scales and intimate bonds measure) and qualitative data collected via individual interviews with women and men at 3 time-points: 6 weeks, 6-9 months, and 2 years post definitive diagnosis. Initially women only were interviewed, and later this was expanded to include their male partners. **Results**: A total of 102 individuals have been interviewed at the first time-point (75 women, 27 men), regarding their experience of receiving a prenatal diagnosis. Conditions identified included chromosomal abnormalities, complex cardiac, structural/organ malformation and neural tube defects. Of the 75 pregnancies, 59 were terminated and 16 were continued. Participants commonly experienced significant grief and overwhelming sadness and many women described intense feelings of isolation. Stress levels were higher than normative controls (p<0. 0001) but women and men showed higher partner support and less control compared with normative controls (p=0. 0001). Those choosing to have an abortion frequently described feeling negatively 'judged'. Access to abortion, levels of support and the perceived quality of support varied both within and across the sites, and couples, whether terminating or not, identified a need for greater access to information and support. Later timepoint interviews indicated that women continued to focus on the need for support when trying to become pregnant or when actually pregnant again. **Conclusions**: Following prenatal diagnosis, couples describe variable and sometimes inadequate levels of follow-up bereavement care and support. The expanding scope of prenatal testing, including non-invasive prenatal screening and chromosomal microarrays, means couples are increasingly faced with complex choices. Providing prenatal testing in the absence of a full range of supportive options may be considered unethical.

2133F

Methionine Synthase (MTR) and maternal risk factor for Down syndrome child-A case-control study. *A. Kaur, A. Kaur.* Human Genetics, Guru Nanak Dev University, Amritsar, India.

Introduction: Methionine Synthase (MTR) is essential for the conversion of homocysteine (hcy) to methionine using cobalamin as a cofactor. Presence of SNPs in *MTR* gene alters homocystein/folate levels by decreasing enzyme activities. These changes influence the DNA stability and methylation pattern which may be one of the factors that plays role in the development of a Down syndrome (DS) child. A case-control study was performed to assess the role of polymorphism (MTR 2756 A-G) in *MTR* gene as a maternal risk factor for DS child in Punjab. **Methodology:** We investigated 110 mothers having DS children with mean age of 27. 5 years and 70 controls with mean age of 28. 5 years having at least one normal child and no miscarriages. Genotyping was done using PCR-RFLP method followed by restriction digestion by *Hae III* enzyme. **Results:** A total of 110 cases and 70 controls were analyzed. Out of 110 cases, 53 (48. 2%) and 52 (47. 3%) showed homozygous wild and heterozygous mutant genotype, respectively. Only 5 (4. 5%) cases exhibited homozygous mutant genotype and among controls, 5 (7. 14%) display homozygous mutant genotype, 27 (38. 6%) revealed homozygous wild and 38 (54. 3%) heterozygous mutant genotype. Higher percentage of homozygous mutant genotype was observed among controls. The chi square value (χ^2=1. 829, p= 0. 400) and genetic models [dominant-0. 68 (0. 37-1. 24), codominant-0. 70 (0. 42-1. 17), recessive-0. 62(0. 17-2. 22)] suggested non-significant association between cases and controls. On the other hand, minor allele frequency in controls (G=34. 29%) was observed to be higher as compared to cases (A=28. 18%), however, there was evidence of no significant difference (p=0. 267). Both cases and controls were found to be in Hardy Weinberg equilibrium (χ^2=6. 032, p=0. 049). **Conclusion:** No association of the studied polymorphism with birth of DS was observed. Present data do not support a role for MTR 2756 A-G as maternal risk factor for a DS birth. Future studies are needed to investigate the potential function of this polymorphism as maternal risk factor for DS.

2134T

Prenatally initiated workup for Desbuquois dysplasia type 1: The utility of SNP-microarray in identifying a candidate gene. *K. Rock[1], J. Hoover-Fong[2], J. E. Hooper[3], K. J. Blakemore[4], A. A. Baschat[1].* 1) The Johns Hopkins Center for Fetal Therapy, Department of Gynecology and Obstetrics, The Johns Hopkins University School of Medicine, Baltimore, MD, USA; 2) Greenberg Center for Skeletal Dysplasias, McKusick-Nathans Institute of Genetic Medicine, Department of Pediatrics, Johns Hopkins University, Baltimore, MD, USA; 3) Department of Pathology, The Johns Hopkins University School of Medicine, Baltimore, MD, USA; 4) Maternal Fetal Medicine, Department of Gynecology and Obstetrics, McKusick-Nathans Institute of Genetic Medicine, The Johns Hopkins University School of Medicine, Baltimore, MD, USA.

Desbuquois dysplasia (DBQD1 [MIM 251450]) is an autosomal recessive chondrodysplasia marked by micromelia, severe joint laxity and dislocations, a characteristic radiographic 'monkey wrench' appearance at the proximal femur due to femoral neck shortening and a medial metaphyseal 'notch'. Type 1 Desbuquois dysplasia is caused by *CANT1* mutations and is distinct from type 2 which is caused by mutations in *XYLT1,* in that the former has unique hand anomalies including accessory phalangeal ossification centers, advanced carpal bone maturation, and/or axial phalangeal deviation. Severe prenatal presentations have been rarely reported. A 27 year old, G2P0010, Pakistani woman from a consanguineous relationship presented at 17 weeks gestation with multiple fetal anomalies including bilaterally short ulna, radius, tibia and fibula (<5th%ile), bilateral severe knee abduction and clubfeet, suspected bilateral hand syndactyly with unusual thumb positioning, and kyphosis. Findings were almost identical to a 16 week ultrasound in her first pregnancy. That pregnancy was terminated without postmortem examination, but SNP-microarray revealed loss of heterozygosity (LOH) of 330. 9 Mb (~10. 3%). No specific genetic diagnosis was made. The parents requested termination of this 2nd pregnancy at 18 weeks given concerns of recurrence. Autopsy revealed severe, bilateral, symmetric genu recurvatum; broad, short neck; micrognathia; cleft palate; wide spacing between the 3rd and 4th fingers bilaterally (prenatally interpreted as syndactyly of adjacent fingers) with bilateral bony fusion of the first two phalanges; clubbed proximal ulnae; and short, superiorly displaced first toes. SNP-microarray showed LOH of 177 Mb (~6. 16%). Considering the common physical features of the fetuses, examination of all genes within overlapping regions of LOH revealed *CANT1* as a candidate gene. With limited fetal DNA available, sequencing of *CANT1* revealed a c. 643G>T (Glu215Term) heterozygous mutation in both parents, thereby establishing a diagnosis of Desbuquois dysplasia type 1 in the fetuses. In this case, independent SNP-microarray results did not provide a diagnosis, but knowledge of the specific areas of LOH with consideration of the phenotype targeted one gene of interest. In clinical situations in which ample DNA is not available, utilization of SNP-microarray in consanguineous families at risk for rare autosomal recessive disorders may dramatically narrow the list of candidate genes.

2135F

Successful array CGH analysis of fetal cells derived from maternal blood: A step towards cell-based NIPT. *A. Breman[1], E. Normand[1], S. Qdaisat[1], I. Van den Veyver[1], R. Singh[2], N. Uldbjerg[4], L. Hatt[2], P. Schelde[2], S. Kolvaa[2], J. Chow[3], L. U'Ren[3], J. Stilwell[3], R. Seubert[3], E. Kaldjian[3], A. Beaudet[1].* 1) Baylor College of Medicine, Houston, TX, USA; 2) ARCEDI Biotech ApS, Aarhus, Denmark; 3) RareCyte Inc. Seattle, WA, USA; 4) Aarhus University Hospital, Aarhus, Denmark.

Detection of genomic copy number abnormalities from one or a few a single cells using array CGH offers a promising non-invasive alternative for prenatal diagnosis. Towards this goal, wehave established a collaborative method for performing fetal cell-based, noninvasive prenatal testing (NIPT) during the first trimester. The long-term goal is to develop a test that would allow the use of array CGH and next generation sequencing to detect clinically important copy number variants (CNVs) and point mutations to accomplish all of the diagnostic testing that is possible at present using amniocentesis or chorionic villus sampling. Through multiple collaborations, we have performed successful array CGH on fetal cells isolated from the maternal circulation. ARCEDI Biotech ApS has developed methods for paraformaldehyde fixation of whole blood followed by immunomagnetic bead enrichment, immunostaining on glass slides, custom scanning for positive cells, and picking using commercially available laser capture or cell picking equipment. RareCyte, Inc has focused on developing methods for preservation, density based enrichment, immunostaining, custom high-resolution scanning and analysis, and integrated single-cell picking. Baylor College of Medicine has focused on clinical use of the RareCyte picker, whole genome amplification, array CGH, and interpretation of clinically relevant copy number abnormalities. Using various combinations of these strategies, the group has successfully recovered and analyzed fetal cells from maternal blood samples using short tandem repeat (STR) analysis for confirmation. Array CGH was successfully performed on whole genome-amplified DNA from fetal cells in cases where the fetus was male. Our data suggests between 6 and 12 fetal cells, likely trophoblasts, can be recovered from 30 ml of maternal blood. The immediate goal is to implement a research version of cell-based noninvasive array CGH that would detect all aneuploidy andclinically significant copy number variations (CNVs) down to 3 Mb resolution or less. The longer term goal is to validate the test for commercial use by comparing array CGH from concurrent invasive and noninvasive samples obtained during the first trimester. Cell-based NIPT has dramatic potential advantages over cell-free testing primarily because of the ability to analyze fetal DNA free of contamination by maternal DNA. We conclude that cell-based NIPT can be developed as a routine clinical test in the near future.

2136T

Women in Western Australia; knowledge, understanding and attitudes towards the potential for population Non-Invasive Prenatal Testing (NIPT) for de novo single gene disorders. *S. M. Long.* Genetic Services of Western Australia, King Edward Memorial Hospital, Perth, Western Australia, Australia.

Background: The diagnostic options available through the public health system are chorionic villus sampling (CVS) and amniocentesis, both of which carry a small risk of pregnancy loss. Non-invasive prenatal testing using cell free fetal DNA (cffDNA) in maternal blood (NIPT) has been eagerly anticipated by both consumers and health care providers and from November 2012 has been offered in Western Australia. With the widespread uptake of NIPT and expanded panelof conditions tested for, such as Prader-Willi/ Angelman syndrome, Cru-di-chat, 22q11 deletion and 1p36 deletions the question of how far away population screening for single gene disorders is arises. A proof of concept study at King Edward Memorial Hospital has already commenced to investigate the feasibility of single gene disorder testing through cff DNA in maternal blood as have other studies. However prenatal testing in the presence of a family history of a condition and genetic counselling pre-test is an extremely different situation to population screening in a prenatal testing with no family history. **Statement of purpose**: The proposed research aims to investigate current knowledge and attitudes surrounding NIPT for single gene disorders in the general population, by both qualitative and quantitative research methods. **Methods:** The initial study consists of qualitative interviews with three cohorts of women to assess what is already known about prenatal testing, what attitudes are towards disability and termination of pregnancy and what they think of the potential for new screening options. The transcripts are analysed by NVivo and checked for thematic saturation. This qualitative study will then inform the design of a much larger online survey. **Results:** The women interviewed have a basic knowledge of prenatal screening and diagnostic tests, with most women aware of screening for Down syndrome and the diagnostic test called amniocentesis. Women have little knowledge about other conditions screened for and unless they have a child with a genetic condition they have little knowledge of genetic conditions in general. While most women are supportive of a woman's right to choose termination of pregnancy in the setting of fetal abnormality they are unsure about decision making in the context of uncertain results. The option of potentially prenatally testing for hundreds of different de novo genetic conditions at once has perceived benefits and drawbacks.

2137F

Demonstration of a Subchromosomal Noninvasive Prenatal Testing Method Using Genome-Wide Sequencing. *R. C. Tim[1], A. R. Mazloom[1], J. Tynan[1], T. Liu[1], Y. Wu[1], C. O'Neal[1], Z. Qazi[1], M. Ehrich[2], P. Oeth[1].* 1) Sequenom Laboratories, San Diego, CA; 2) Sequenom, Inc. , San Diego, CA.

Noninvasive prenatal testing (NIPT) has gained widespread acceptance for the detection of fetal whole chromosome abnormalities such as trisomies 21, 18 and 13. However, a broad range of clinically relevant subchromosomal aberrations can also be identified with high accuracy when using genome-wide sequencing of circulating cell free fetal DNA (ccffDNA), especially events ≥ 7 Mb in size. This study demonstrates a methodology for the identification of such microdeletions and microduplications (MDs) using low coverage sequencing data (0. 4 - 0. 5X genome coverage). Assay sensitivity was tested using a model system composed of fragmented and size selected genomic DNA (gDNA), representing a range of subchromosomal abnormalities, mixed into ccfDNA from non-pregnant female donors to mimic ccffDNA at various fetal fractions. Sixty-three independent mixture model samples with events ≥ 7 Mb and a median mixture ratio corresponding to 5. 8% fetal fraction were interrogated. The measured analytical sensitivity was found to be slightly higher than the in-silico predicted sensitivity, therefore validating the simulations. The expected sensitivity of microdeletions > 7 Mb is greater than 95%. It should be noted that the observable event sizes based on sequencing were smaller than the clinically reported sizes (often by 15% or more) due to poorly mappable regions within the genome. Assay specificity was assessed using ccfDNA extracted from presumably unaffected (euploid) maternal plasma samples. The 395 samples tested were not known *a priori* to possess subchromosomal events ≥ 7 Mb based on previous low coverage sequencing results. After filtering for the typical quality control criteria, no subchromosomal events > 7 Mb were detected. Consequently the data in this study suggests high sensitivity and specificity for the detection of subchromosomal events > 7 Mb can be achieved.

2138T

Contribution of NIPT in case of ultrasonc malformations: more than 2 years of French experience. *G. Viot, L. LHOMANN, C. BERNABE-DUPONT, C. LACAM, P. BOUHANNA, I. THUILLIER, F. JACQUEMARD.* prenatal diagnosis unit, Neuilly sur Seine, France.

Detection of fetal DNA in maternal blood allows non-invasive prenatal testing (NIPT) of aneuploidy with very high sensitivity and specificity. The implementation of NIPT increases in lot of countries with time with a recognized indication of all for the diagnosis of trisomy. In our unit, we asked ourselves the question whether this test could be proposed in other indication, especially after discovery of ultrasonic malformations. From January 2013 to April 2015, we carried out 3 000 tests mainly because of advanced maternal age or abnormal first trimester screening (FTS). For 117 women (3. 9%), fetal DNA was proposed while scan detected fetal malformations (28. 5%: thick nuchal translucency; 21. 4%: soft signs of trisomy 21; 12. 85%: intrauterine growth retardation; other: 37. 27%). For 2 cases, the result was positive for trisomy 21. Classically, fetal DNA testing is not a good option in case of ultrasonic anomalies. For all these women, we granted the request for specific reasons: refusal of patients to perform an invasive procedure due to the risk of induced miscarriages, risk of premature delivery or history of premature labor, decision to continue the pregnancy whatever the outcome, assisted procreation history. Genetic counseling is mandatory before testing. Test limitations were explained. An ultrasound control was always performed in our unit. Clinical evaluation of newborns and follow-up were always practiced after birth. After ultrasound discovery of anomalies and in very specific situations, NIPT can be performed after genetic counseling and explanation of its limitations. In the near future, it appears important to evaluate the help of this non-invasive approach, mainly after discovery of soft signs in favor of Down syndrome.

2139F

First Report of Prenatal Diagnosis for Severe Genodermatoses in Egypt. *K. Amr[1], K. Gaber[2], G. Kamah[3], M. Farag[2], H. Naser[1].* 1) Medical Molecular Genetics, National Research Centre, Cairo, Egypt; 2) Prenatal Diagnosis Unit , National Research Centre, Cairo, Egypt; 3) Clinical Genetics, National Research Centre, Cairo, Egypt.

Background: Genodermatoses are mostly severe inherited disorders. A great success in identifying responsible genes & characterizing mutations within such genes paved the road for DNA-based prenatal diagnosis. Examples of severe genodermatoses candidate for prenatal diagnosis include autosomal recessive congenital ichthyosis (ARCI), Xerodemapigmentosa (XPA), Sjögren–Larsson syndrome (SLS) and papillonlefeuvre syndrome (PLS)where clinical severity affects span &/or quality of life hence urging prenatal diagnosis. **Materials and metho**ds: The study included five amniotic samples (AF) from carrier mothers descending from five pedigrees with history of affected sibs with severe genodermatoses including; two mothers of previous ARCI cases, one XPA, one SLS and one PLS. DNA was extracted from AF samples by QIA gene extraction kit followed by mutational screening for XPA, TGM1, ALDH3A2 and CTSC genes. **Results**: prenatal diagnosis was successfully performed in all cases. For the family with history of XPA, the fetus was found to be heterozygous carrier for E111X mutation; For ALDH3A2 gene the fetus was affected for E331X nonsense mutation; for TGM1 gene the two fetuses were heterozygous carriers for R264W,R143Hmissense mutations. The fifth AF sample PLS showed homozygous wild type genotype. **Conclusion**: The high incidence of consanguinity & consequently AR rare disorders combined with the lack of curative therapy, points to the importance of implementing preventive programs. Prenatal diagnosis and genetic counseling represent an important step in prevention & alleviating the burden of severe genodermatoses on the family & community.

2140T

5-tiered Scheme for Classification of High-Resolution Oligo-SNP Array Results in Invasive Prenatal Diagnosis: Findings from 3,091 Clinical Cases. *A. Anguiano, J. C. Kelly, L. P. Ross, L. W. Mahon, R. Owen, F. Z. Boyar, C. M. Strom, J. C. Wang.* Cytogenetics, Quest Diagnostics Nichols Institute, San Juan Capistrano, CA.

Precise classification of copy number variants identified by high-resolution (HR) oligo-SNP array in prenatal diagnosis delivers key genetic information that is often a significant factor for families making decisions about their pregnancies. Our laboratory reports HR oligo-SNP results using a standardized 5-tiered scheme: 1) abnormal/pathogenic, 2) likely pathogenic, 3) uncertain clinical significance, 4) likely benign, or 5) normal. We assessed the findings of HR oligo-SNP array testing of chorionic villus and amniocentesis samples in clinical practice. The HR whole-genome CytoScan™ HD oligo-SNP array was implemented among a cohort of 3,091 high-risk pregnancies with a threshold set at 50 kb for chromosomal loss, 200 kb for gain, and 5 Mb for regions of allelic homozygosity (ROH). The major indications for invasive testing were abnormal ultrasound (64%), abnormal maternal serum screen (16%), and advanced maternal age (15%). Benign gains or losses, gains or losses < 1 Mb not encompassing genes, and gains < 0. 5 Mb of unknown clinical significance were not reported. Presence of ROH was reported for samples with either ≥2 ROH with a total size >10 Mb or a single ROH >10 Mb. Pathogenic findings were identified in 290 (9. 4%) cases, and variants of uncertain clinical significance (VOUS) were found in another 272 (8. 8%). ROH was found in 88 (2. 8%) cases (range, 11-515 Mb; median, 51 Mb). The most frequent pathogenic findings were trisomies (103 cases; 35. 5%), microdeletions/microduplications <10 Mb (95; 32. 8%), and large deletions/duplications and unbalanced translocations >10 Mb (53; 18. 3%); DiGeorge syndromes, STS deletion, and 17q12 and 16p13. 11 deletions and duplications were the most prevalent. Among the 272 VOUS cases, 45% were likely benign and 53% were of uncertain clinical significance (most due to partial duplications affecting genes implicated in autosomal dominant disorders). Only 4 (1. 5%) cases were classified as likely pathogenic. The overall detection rate was about 9. 5% for predicting a phenotypically abnormal pregnancy, and close to 5% for copy number variants requiring detailed clinical assessment for genotype-phenotype correlation. The results from this cohort provide further support of HR oligo-SNP array as a first-tier genetic test for prenatal diagnosis in high-risk pregnancies. The 5-tiered standardized interpretation scheme provides valuable information to inform pregnancy decisions.

2141F

Prenatal diagnosis in Huntington disease and reproductive choices. *H. Bouchghoul[1], S. Clement[2], M. Gargiulo[3,4], C. Cazeneuve[5], S. Noel[5], E. Schaerer[5], M. Dommergues[1], D. Vauthier[1], J. Nizard[1], A. Durr[2,5].* 1) Service de Gynécologie Obstétrique, Groupe Hospitalier Pitié-Salpêtrière, APHP, Sorbonne Universités, UPMC Univ Paris 06, CNRS UMR 7222, INSERM U1150-, Paris, France; 2) 2-Institut du Cerveau et de la Moelle, INSERM U1127, CNRS UMR7225, Sorbonne Universités - UPMC Université Paris VI UMR_S1127, Paris, France; 3) 3-Laboratoire de Psychologie Clinique et Psychopathologie, EA 4056, Université Paris Descartes, Sorbonne Paris Cité, Institut de Psychologie, Paris, France; 4) Institut de Myologie, Groupe Hospitalier Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, Paris, France; 5) APHP Department of Genetics, Groupe Hospitalier Pitié-Salpêtrière, Paris France.

In France, persons carrying the mutation responsible for Huntington disease (HD), an expanded CAG repeat in the HTT gene, who wish to prevent transmission of the mutation to their offspring can request a prenatal diagnosis (PND) and, eventually, terminate the pregnancy. Our goal was to determine the impact of PND on subsequent reproductive decisions, as well as to investigate how the information was disclosed to children born after a PND. Between April 1, 1996, and January 29, 2015, 62 couples requested 102 PND for HD in our center. This represents only 18% of the 346 carriers aged <45 years who requested presymptomatic testing during this period. There were 31 carriers and 31 spouses of male carriers; 74% and 87% were nulliparous, respectively (p= 0. 20). Women were contacted by phone between March and May 2015 and invited to participate in a survey; 55 responded positively (89%), 1 declined and 6 could not be reached. The median delay between the last PND and phone contact was 6. 5 years [4. 5-9. 9]. Multiple PND were sometimes requested: 1 (n=33), 2 (n=24), 3 (n=2), 4 (n=2) and 6 (n=1). Those who requested more than one PND were carriers as well spouses (p= 0. 78). Spouses and female carriers were aged 27. 1 [24. 9-31. 3] and 30. 5 years [26. 9-32. 6] (p= 0. 15). Pregnancies were terminated after an unfavorable result (52/102), except for one carried to term and one miscarriage. Repeated PND were significantly more frequent after an unfavorable initial result (21/36) than after a favorable result (8/26) p= 0. 03. During follow-up, there were 16 pregnancies after a PND, 33% (10/30) in carriers versus 24% (6/25) in spouses (p= 0. 46). Twelve out of 45 children born after a PND (27%) were informed of their status by both spouses (8/21, 38%) and female carriers (4/24, 17%, p= 0. 11). In conclusion, PND is rarely requested after presymptomatic testing for HD, and is comparable in female carriers and spouses of male carriers. The hypothesis that future reproductive choices in women are impeded after direct testing of the fetus is supported by our data; a second PND is requested more often after an unfavorable first result. Only 16 children were born without a PND during follow-up. The traumatic dimension of the PND procedure was well reported by the women in general. The fact that the parents disclosed their status to 27% of the children born after a PND, despite their youth, was unexpected and may reveal a long-lasting suffering of the parents due to their own risk.

2142T

Needs and views of Quebec pregnant women on personalized reproductive medicine. *G. Lapointe[1], R. Drouin[1,2], C. Bouffard[1].* 1) Université de Sherbrooke, Sherbrooke, Québec, Canada; 2) Université Laval, Québec, Québec, Canada.

INTRODUCTION: Whole-genome sequencing tools are entering the field of human reproduction and gradually replacing conventional diagnostic techniques. These techniques produce a large quantity of genomic data and medical information about the fetus, thereby giving rise to personalized reproductive medicine (PRM). Thus PRM engages parents to take unprecedented responsibility on unborn children's health. This new responsibility, combined with the explosion of genetic information, the advent of non-invasive prenatal tests and direct-to-consumer genetic testing, exposes PRM to several pitfalls. Our research objective was therefore to develop empirical knowledge on the quantity and type of information that pregnant women would like to have about their fetus. **METHODOLOGY**: Descriptive-interpretive qualitative design and general inductive analysis of 15 semi-structured interviews with pregnant women who did not have a known risk of genetic disease, and whom had previously been briefed on available and forthcoming prenatal tests. **RESULTS**: Preliminary results show that women refer to the lack of information provided by physicians in the context of trisomy 21 prenatal screening to express their concerns about more complicated tests. They talked about the types of information, medical or not, that they would like to receive, or not, concerning their fetus' health. They also emphasized the importance of being informed on the medical and psychosocial consequences of their prenatal genetic testing choices. The quality of this information seems highly important to them, in order to evaluate the burden of their responsibility in line with their choices and the impact of these latter on the future of the pregnancy. This new «active participation» and responsibility of parents that PRM implies seems excessive to them and some women fear that it will be harmful for themselves and their child. **CONCLUSION**: At last, the pregnant women we interviewed want to have the choice, but not without complete information about genetic, medical and social issues. In addition, participants are open to learn more about the genome of their fetus, but propose that it should be done in a regulated framework with proper medical assistance.

2143F

Frequency of Total Uniparental Isodisomy and Triploidy in a Series of Oligo-SNP array analyses of Products of Conception and Prenatal Diagnosis. *B. T. Wang[1], R. Owen[1], F. Z. Boyar[1], L. Ross[2], L. Mahon[2], J. C. Wang[1], M. El Naggar[1], J. Kelly[1], M. Hemmat[1], A. Anguiano[1].* 1) Cytogenetics Dept, Quest Diagnostics,Nichols Inst, San Juan Capistrano, CA; 2) Genetic Counseling Dept, Quest Diagnostics, San Juan Capistrano, CA.

Molar pregnancy occurs in 0. 1% of all pregnancies. A molar pregnancy occurs when tissue that normally becomes a fetus, instead becomes an abnormal growth in the uterus. Two types of molar pregnancies have been observed. The first is complete hydatidiform mole (1 in 1500 pregnancies) due to a diploid conception that is essentially a total uniparental disomy (TUPD); this type of molar pregnancy carries a significant risk for developing persistent gestational trophoblastic disease. The second type is partial hydatidiform mole (1 in 700 pregnancies), seen in most cases with triploidy. Products of conception (POC) are often evaluated with routine chromosome analysis, which has a failure rate of up to 25% due to degradation of tissue after fetal demise and cannot detect TUPD. Oligo-SNP array analysis, on the other hand, has a much lower failure rate (<10%), a shorter turnaround time and can detect cases of TUPD, as well as triploidy. In this study, we determined the frequency of TUPD and triploidy in 1,471 consecutive POC samples and 3230 amniotic fluid specimens analyzed using the CytoScan HD Oligo-SNP array (AffymetrixTM). Oligo-SNP analysis identified TUPD in 6 (0. 4%) and triploidy in 56 (5%) of the 1,471 POC cases. Among the 3,230 amniotic fluid specimens, oligo-SNP detected TUPD in 1 case (0. 03%) and triploidy in 9 (0. 3%). The number of cases with TUPD do not include cases with total uniparental heterodisomy. In conclusion, this study demonstrates the clinical usefulness of the oligo-SNP array for identifying chromosome complements associated with complete and partial molar pregnancies. Oligo-SNP provides additional value over the non-SNP based array platforms. Choice of a SNP-based array is beneficial for patient care after spontaneous abortion, including the detection of abnormal chromosome complements that could lead to persistent gestational trophoblastic disorders, thus enabling proper monitoring and clinical management. Thus, oligo-SNP array should be considered as a first-tier analysis for pregnancies ending in spontaneous abortions.

2144T

Integrated Analysis of a Longitudinal Preterm Birth Cohort Reveals Significant Inflammatory Candidate Genes. *M. K. Veerapen[1,2], E. Rampersaud[3], N. Morris[4], J. E. Potter[5], M. M. Rodriguez[6], W. Shu[7], Z. Stowe[4], O. A. Bodamer[1,2].* 1) Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA; 2) Department of Biochemistry and Molecular Biology, University of Miami, Miami, FL; 3) Department of Computational Biology, St Jude's Hospital, Memphis, TN; 4) Department of Psychiatry, University of Arkansas, Little Rock, AR; 5) Department of Obstetrics and Gynecology, University of Miami, Miami, FL; 6) Department of Pathology, University of Miami, Miami, FL; 7) Department of Neonatology, University of Miami, Miami, FL.

Introduction: Although affecting up to 13% of total deliveries, and the largest cause of perinatal morbidity and mortality, the longitudinal aspect of preterm births (PTB) during pregnancy is relatively unknown. *Objective:* We proposed to identify temporal differences in the transcriptome and DNA methylation in PTB and term births: during pregnancy, delivery and post-partum. All maternal samples were controlled for age, gravida and gestational weeks; and selected for complication free pregnancies and vaginal deliveries, to eliminate potential confounding factors. *Method*: Total RNA and DNA from peripheral blood were extracted using MagMAX Tempus Blood RNA extraction kit and QIAamp DNA Blood Midi kit, respectively. Ribosome reduced RNA was used to evaluate the transcriptome for known genes. Bisulphite-treated DNA was genotyped for methylation using Infinium Human Methylation 450K BeadChip arrays (Illumina). RNA was library prepared according to TruSeq RNA library prep protocol and run at 2 samples per lane on the Illumina HiSeq2000 generating 25-35 million reads per library. We aligned and assembled raw reads to UCSC Known Genes (hg19). Differential expression (DE) was evaluated using generalized linear regression and Wald test for significance. Pathway analysis using MetaCore was performed to evaluate biological clustering of transcripts. DNA methylation was evaluated using F-tests and linear regressions for differentially methylated positions (DMP). The transcriptome and DNA methylation results were integrated using Spearman correlation. *Results and Discussion*: The established regression model resulted in 3408 DE genes (q<0. 05). These genes were overrepresented for infection and inflammation networks, supporting current physiological causes of PTB. However, only 232 genes were highly upregulated (fold change (FC)>1. 0). Most of the 20,448 DMPs (q<0. 05) were overrepresented in open seas CpG island type. We identified 6 candidate genes (FC>2. 0) with DMPs in upstream genomic regions, which were moderately negatively correlated (Spearman's χ=0. 61). Potentially, these genes have biological significance in PTB etiology through their involvement in neutrophil activity, oxidation, innate immunity and macrophage activation. *Conclusion*: To the best of our knowledge, this is the first longitudinal and integrated analysis of a PTB cohort. The 6 genes identified in this study may be relevant for the identification of biomarkers during pregnancy in the prevention of PTB.

2145F

Association of fetal genotypes with cytokine protein expression levels and preterm birth. *R. F. Clark[1], N. C. Gaddis[1], S. W. Erickson[1], G. P. Page[1], C. M. Cotten[2], D. J. Zaccaro[1].* 1) RTI International, Research Triangle Park, NC; 2) Dept. of Pediatrics, Duke University School of Medicine, Durham, NC.

Preterm birth (PTB)-related disorders are the leading cause of neonatal morbidity and mortality. Inflammation is strongly associated with PTB, and cytokines are markers for inflammation. Several studies have suggested that fetal genotypes may play a role in determining the magnitude of the inflammatory response at the fetomaternal interface. Although associations have been found between maternal genotypes and PTB-related disorders, few studies have examined and identified fetal genotypes that are associated with the morbidities and mortality associated with PTB-related disorders, and those that have demonstrated modest association require validation/replication. We investigated which fetal cytokine levels in whole blood (and their specific regulatory genotypes) are associated with specific neonatal outcomes. We analyzed the Neonatal Research Network candidate gene and genome-wide association datasets to investigate associations between single nucleotide polymorphisms (SNPs) and cytokine levels. We also used quantitative trait loci (QTL) analysis to: 1) test if genetic markers act in cis to regulate cytokine protein levels, 2) test genetic association to identify variants in genes affecting cytokine protein levels, and 3) identify QTLs that are associated with cytokine protein levels. We found highly significant (all p<10-8) associations for cis-SNPs with the *CRP*, *IL6*, and *IL18* protein amounts. We also revealed highly significant associations for trans-SNPs of *IL6* and *NTF4* for CXCL8; of *NTF4*, *IL17C*, and *IL6R* for CCL3; and of *IL12A* and *TREM1* for IL6R blood protein concentrations. Further analysis examined whether any of the cytokines measured over 5 time points were associated with altered risk of adverse outcomes in PTB infants. Survival analysis using Cox Proportional Hazards Models with time-dependent covariates was employed to relate outcomes to cytokines in a longitudinal framework. CXCL8 and IL6 were both positively associated with death as an outcome, and CXCL8 was positively associated with intraventricular hemorrhage (IVH) (all p<0. 001). Combined, these analyses have related SNPs that alter IL6 and CXCL8 protein levels to death and IVH. More complex analyses have been undertaken to further delineate the influences of inflammatory pathway genes and cytokines to PTB-related disorders. Identifying these genetic contributions may lead to more effective prevention strategies, attenuating the individual and societal burdens of premature births.

2146T

Molecular genetic study to elucidate complex molar pregnancy: differential diagnosis of an unusual case of hydatidiform mole with fetus. *A. C. M. Malinverni[1,2], M. I. Melaragno[1], L. Pola[1], M. E. S. Colovati[1], E. Iwamura[2], L. D. Kulikowski[3], A. F. Moron[4], S. Y. Sun[5], M. M. S. Ishigai[2].* 1) Genetic Division, Department of Morphology and Genetics, Universidade Federal de São Paulo, São Paulo, São Paulo, Brazil; 2) Department of Pathology, Universidade Federal de São Paulo, São Paulo, Brazil; 3) Pathology Division,Cytogenomics Lab, LIM 03, Universidade de São Paulo, São Paulo, Brazil; 4) Department of Obstetríc, Division of Fetal Medicine, Universidade Federal de São Paulo, São Paulo, Brazil; 5) Department of Obstetric, Universidade Federal de São Paulo, São Paulo, Brazil.

Hydatidiform mole (HM) is a complication of pregnancy, characterized by several degrees of trophoblastic hydropic degeneration and proliferation of chorionic villi with potential risk of malignancy. There are two types of HM: complete hydatidiform mole (CHM) and partial hydatidiform mole (PHM). The more difficult differential diagnosis is between PHM and dysmorphic, non-molar spontaneous abortions. While PHMs are triploid, non-molar abortion can be diploid or triploid. We report on a primigest patient, with bhCG 201,039 mIU/Ml at 27 weeks gestation. Ultrasound showed several cystic areas. Fetus of 27 weeks gestation passed away within 15 minutes. Fetal blood karyotype showed 69,XXY, suggesting PHM. Histological examination showed two villous populations: trophoblastic hyperplasia and atypia, villous dysmorphism with epithelial pseudoinclusions, terminal and stem villi edema and cistern formation with stem vessels thrombosis, suggesting PHM/placental mesenchymal dysplasia (PMD). Immunohistochemistry study showed Ki67+ in trophoblastic atypia spots. P57 showed a discrepant pattern with lack of expression in cytotrophoblast and stromal cells and positive expression in others, but with diffuse expression in stromal cells with focal positivity in cytotrophoblast cells. The positive p57was less expressed in villi cells and the syncytiotrophoblasts were uniformly negative leading to suspicion of cellular mosaicism/placental mesenchymal dysplasia. Fluorescence *in situ* hybridization (FISH) using paraffin material with centromeric X, Y and 18 probes in 100 nuclei showed 51% of diploid cells (XX) and 48% triploid (XXX). The trio Genotyping by AmpFLSTR® MiniFiler™ PCR Amplification Kit with parent´s whole blood and patient´s placental tissue revealed only pattern alleles. While cytogenetic analysis and the clinical findings were suggestive of PHM, FISH corroborated with the pathological findings of placental mosaicism associated with molar behavior. Molecular genetic testing for the diagnosis of HM showed only a paternal contribution to the genome, but with a mosaic cell line, which indicated the diagnosis of twin molar pregnancy and mosaic conception or non-molar fetus and hydatidiform mole mosaic. These data showed the importance of molecular techniques in the differential diagnosis of HM, and should be considered when the morphological screening does not allow definitive diagnosis, and also to better understand the mechanisms of HM formation. CNPq and FAPESP.

2147F

Maternal genome-wide association study identifies a fasting glucose variant associated with offspring birth weight. *B. Feenstra[1], A. Cavadino[2,3], J. Tyrrell[4], G. McMahon[5], M. Nodzenski[6], M. Horikoshi[7,8], F. Geller[1], R. Myhre[9], R. C. Richmond[5,10,11], L. Paternoster[10], J. P. Bradfield[12], E. Kreiner-Møller[13], M. -R. Jarvelin[14], S. Metrustry[15], K. L. Lunetta[16,17], J. N. Painter[18], J. J. Hottenga[19], C. Allard[20], S. J. Barton[21], A. Espinosa[22,23], J. A. March[24], C. Potter[25], M. -F. Hivert[26,27], J. F. Felix[11,28], E. Hyppönen[29], W. L. Lowe, Jr.[30], D. A. Lawlor[5,10], T. M. Frayling[4], R. M. Freathy[4,10], EGG Consortium.* 1) Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark; 2) Centre for Environmental and Preventive Medicine, Queen Mary University of London, UK; 3) Institute of Child Health, University College London, UK; 4) Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK; 5) School of Social and Community Medicine, University of Bristol, Bristol, UK; 6) Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; 7) Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, UK; 8) Wellcome Trust Centre for Human Genetics, Oxford, UK; 9) Department of Genes and Environment, Norwegian Institute of Public Health, Oslo, Norway; 10) MRC Integrative Epidemiology Unit, University of Bristol, UK; 11) The Generation R Study Group, Erasmus Medical Center, Rotterdam, the Netherlands; 12) 13. Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, PA; 13) Danish Pediatric Asthma Center, Gentofte Hospital, University of Copenhagen, Denmark; 14) Institute of Health Sciences and Biocenter, University of Oulu, Oulu, Finland; 15) Department of Twin Research, King's College London, St. Thomas' Hospital, London, UK; 16) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 17) Framingham Heart Study, Framingham, MA, USA; 18) QIMR Berghofer Medical Research Institute, Herston, Qld, Australia; 19) EMGO Institute for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands; 20) Department of Mathematics and Centre de recherche du Centre Hospitalier Universitaire, Universite de Sherbrooke, QC, Canada; 21) MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK; 22) Centre for Research in Environmental Epidemiology and Hospital del Mar Medical Research Institute, Barcelona, Spain; 23) Pompeu Fabra University (UPF), Barcelona, Spain; 24) School of Women's and Infants' Health, The University of Western Australia, Perth, Australia; 25) Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK; 26) Department of Population Medicine, Harvard Medical School, and Diabetes Center, Massachussetts General Hospital, Boston, MA; 27) Department of Medicine, Universite de Sherbrooke, QC, Canada; 28) Departments of Epidemiology and Pediatrics, Erasmus Medical Center, Rotterdam, the Netherlands; 29) School of Population Health, University of South Australia, Adelaide, Australia; 30) Division of Endocrinology, Metabolism and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL.

Several common fetal genetic variants have been associated with birth weight, but little is known about how maternal genetic variation influences fetal growth through the intra-uterine environment. To identify maternal genetic variants associated with birth weight, we performed a meta-analysis of 11 genome-wide association studies (GWAS; n = 19,626 women of European descent). We selected 18 single nucleotide polymorphisms (SNPs) for replication analysis in up to 13 further studies (n = 18,319 women of European descent). One SNP reached genome-wide significance (rs10830963, *P* = 2. 0 x 10-11) in a combined analysis of discovery and replication results. Rs10830963 is intronic in *MTNR1B* and is known from previous GWAS to be associated with fasting glucose levels, type 2 diabetes and gestational diabetes. Each copy of rs10830963-G (the allele associated with higher fasting glucose) corresponded to a 31g [95%CI: 22, 41g] higher offspring birth weight. The association between maternal rs10830963 and birth weight was unaltered by adjustment for fetal genotype in 8716 maternal-fetal pairs. We also found evidence of a collective effect of known maternal fasting glucose-associated variants on offspring birth weight: more low *P*-values than expected were observed among 13 variants selected, with the glucose-raising allele corresponding to higher offspring birth weight for the 5 most significantly associated SNPs. In conclusion, our study demonstrates that maternal genetic variation at *MTNR1B* influences offspring birth weight and supports a broader role of genetic variation affecting maternal glucose levels in fetal growth. Our study also highlights that the effect sizes of associations between other maternal genetic variants and birth weight are unlikely to exceed 20g per allele, and therefore much larger sample sizes will be required to detect them.

2148T

Low folate levels and MTHFR polymorphism C 677T in case mothers of chuldren with neural tube defects and control mothers of Pakistani origin. *N. Nauman[1], S. Jalali[2], S. Shami[2], S. Rafiq[1], G. Grobe[3], A. Hilger[3], M. Draakan[3], M. M. Nöthen[3], M. Ludwig[4], H. Reutter[5].* 1) Department of Pathology, Rawalpindi medical college, Rawalpindi, Holy Family Hospital, Pakistan; 2) Department of Animal Sciences,Quaid-i-Azam University, Islamabad;Pakistan; 3) Institute of Human Genetics, Bonn, Germany) (Department of Genomics Life &Brain Center, Bonn, Germany; 4) Department of Clinical Chemistry and Clinical Pharmacology, Bonn, Germany; 5) Institute of Human Genetics, Bonn, Germany) (Department of Neonatology,Bonn, Germany.

Introduction: Neural tube defects (NTDs) are congenital malformations of the central nervous system which result from a failure of the neural tube to close during the fourth week of embryogenesis. Low folate levels are associated with increased risk of neural tube defects. Genetic studies, examining the gene coding for the folate metabolizing methylenetetrahydrofolate reductase (MTHFR) enzyme suggest, that the functional677C/T thermolabile polymorphism contributes a genetic risk to NTDs. Previousstudies suggest that not only the fetal genotype but also the maternal genotype mighthave an impact on fetal development. The deleterious effects of this variantmutationcan be overcome if folate levels are adequate. Here we carried out a case control studyof case and control mothers of Pakistani origin. **Patients and Methods:** We examined 109 case mothers of children born with NTDTs and of100 control mothers without history of NTDs in their offspring. Case mothers and control mothers were of Pakistani origin. Red blood cell folate and serum folate levels were determined using Abbot Axym ion capture protocol. The C677T MTHFR polymorphism was genotyped by polymerase chain reaction followed by restriction digestion (PCRRFLP). Genotype and folate level comparisons were carried out using chisquare analysis. **Results:** Mean RBC folate and serum folate levels were significantly lower in case when compared to control mothers (p<0. 0001). Genotype analysis of case mothers and controlmothers revealed 11(10. 09%) and2(2%), respectively, to be homozygous for the MTHFR 677T allele. 32(29. 35%) case and 26(26%) control mothers were heterozygous. Genotype comparison by chisquare analysis of case and controlmothers was significant (p=<0. 0393). **Conclusion:** We found the homozygous MTHFR 677TT genotype to be more frequent among case when compared to control mothers. Mean RBC and serum folate levels were significantly lower in case when as compared to control mothers. Combined analysis of these results with the information of folate supplementation during the periconceptional period of case and control mothers is warranted to further elucidate the role of the MTHFR genotype and folate supplementation per se in the risk of NTDs in pregnancies of women of Pakistani origin.

2149F

***MED12* mutations in uterine leiomyomas: clinical implications.** *HR. Heinonen[1], A. Pasanen[2], R. Bützow[2], M. Mehine[1], P. Vahteristo[1], J. Sjöberg[3], O. Heikinheimo[3], N. Mäkinen[1], LA. Aaltonen[1].* 1) Department of Medical and Clinical Genetics, Genome-Scale Biology Research Program, University of Helsinki, Helsinki, Finland; 2) Department of Pathology, Haartman Institute, University of Helsinki and HUSLAB, Helsinki, Finland; 3) Department of Obstetrics and Gynecology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.

Uterine leiomyomas are hormone-dependent smooth muscle neoplasms that affect nearly 70% of females during their reproductive lifespan. Approximately 25% of these women suffer from symptoms such as abnormal menstrual bleeding, pelvic pain and discomfort, pregnancy complications, and even infertility. Various drug therapies are available, but commonly serve as a temporary relief of symptoms or a preoperative measure. Consequently, leiomyomas are the primary indication for hysterectomy. Risk factors for leiomyomas include early age-at-menarche, nulliparity, high body mass index (BMI), and African origin. Furthermore, certain genetic variations increase the risk of leiomyomas. In 2011, we discovered that *MED12* is frequently mutated in uterine leiomyomas. Subsequently, we showed that up to 86% of leiomyomas harbor a *MED12* mutation and the mutations associate with a smaller tumor size and a larger number of tumors within the uterus. The analysis included also other clinical variables, such as BMI, but no statistically significant associations were observed. These results were derived, however, from a relatively small sample set, thus further studies with larger sample sizes are needed to confirm the observed trends. In this study, we have collected all feasible leiomyomas ≥ 1 cm in diameter, a total of 540 tumors, and the respective normal myometrium tissue samples from 180 patients undergoing hysterectomy at Helsinki University Central Hospital. Location, size, histopathological characteristics have been recorded and *MED12* mutation status determined for each lesion. In addition, we will collect comprehensive clinical data from patients' medical records including symptoms, known leiomyoma risk factors, and responses to different treatments. We will compare the association of *MED12* mutation status with these clinical and tumor-specific variables. This study aims to identify the clinical characteristics unique for *MED12* mutation-positive leiomyomas that should be relevant when targeted treatments become available, and could provide new knowledge on the tumorigenic mechanism underlying these extremely common tumors.

2150T

Transcriptional profiling analyses of cell free RNA in the amniotic fluid with pregnancy of intrauterine growth retardation. *D. H. Cha[1,2], S. H. Shim[2], Y. J. Shin[2], J. E. Park[2], S. H. Kim[1], Y. W. Jung[1].* 1) Dept OB/GYN, Kangnam CHA Hosp, Seoul, South Korea; 2) Genetic Laboratory of Fetility Center, Kangnam CHA Hospital, South Korea.

The intrauterine growth restriction (IUGR) is commonly defined as slow growth of fetus with low weight. Fetal growth restriction leads to the risks of fetal death or morbidity. Because the earlier IUGR leads to the greater risks to the unborn baby, it is important to the serial monitoring of fetus with IUGR. Amniotic fluid (AF) includes nutrients and growth factors for fetal development and provides an important source for the study of multiple functions in fetal development. Because the supernatant of AF has abundant cell-free nucleic acids originated from the fetus, the studies of AF cell-free fetal RNA are valuable for understanding of fetal maturation during each stage of pregnancy. In our study, we performed a comparative analysis based on Affymetrix microarray using cell-free fetal RNA isolated from amniotic fluid of fetuses with IUGR and controls to understand the prenatal development of fetuses with IUGR. We found 448 differentially expressed genes, in which, there are 204 up-regulated genes and 244 down- regulated genes with > 1. 5 fold-change in expression by IUGR (p-value < 0. 05). And we analyzed these DEGs using the both tools such as DAVID and IPA database etc. In analysis by DAVID, the results showed that many of up-regulated genes are involved in various organs while most of down-regulated genes are involved in brain in fetus with IUGR. In functional analysis by IPA, it revealed that up-regulated genes mainly belong to pathways related with protein synthesis. On the other hand, down-regulated genes belong to several pathways including protein degradation. Therefore, the results suggest the retardation of brain development in fetus with IUGR comparison with normal.

2151F

Pathway-based analysis of preeclampsia genome-wide association data suggests hematologic and immune pathways are integral to disease pathogenesis. *K. J. Gray[1], V. P. Kovacheva[2], H. Mirzakhani[3], A. Bjonnes[4,5], B. Keating[6], B. T. Bateman[3], R. Saxena[3,4,5].* 1) Maternal-Fetal Medicine, Brigham & Women's Hospital, Boston, MA. , US; 2) Department of Anesthesiology, Brigham and Women's Hospital, Boston, MA, US; 3) Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston MA, US; 4) Center for Human Genetic Research, Massachusetts General Hospital, Boston MA, US; 5) Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, US; 6) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, US.

Preeclampsia is a severe disorder of pregnancy characterized by new-onset hypertension and proteinuria. While genetic background is known to influence risk, the genetic loci contributing to risk are not well understood. To better understand preeclampsia genetics, we conducted a large-scale gene-centric association analysis of women of European ancestry with 516 preeclampsia cases and 1097 normotensive controls using a cardiovascular gene-centric 50K SNP array. None of the loci genotyped were independently associated with preeclampsia based on a p-value of $< 5x10^{-8}$for genome-wide significance. The Data-driven Expression Prioritized Integration for Complex Traits (DEPICT) tool was then used to prioritize genes, pathways, and tissue/cell types involved in preeclampsia based on association signals with $p < 10^{-3}$. DEPICT identified 40 prioritized genes (false discovery rate < 5%) present in 172 overlapping pathways (see Table for top genes). Potentially relevant biologic processes identified include: abnormal immune system physiology, abnormal immune serum protein physiology, abnormal physiological neovascularization, hemorrhage, decreased glycogen level, abnormal blood vessel physiology, increased leukocyte cell number, abnormal innate immunity, abnormal respiratory system physiology, and increased neutrophil cell number. The prioritized genes were significantly enriched in the hematologic and immune systems and specifically included monocytes, phagocytes, neutrophils, and hematopoietic cells. Taken together, these data strongly suggest that genetic loci central to immune and hematologic processes are integral to preeclampsia pathogenesis. Ongoing work aims to replicate and extend this analysis in a GWAS from 3000 preeclampsia cases and matched controls.

Table. Top genes prioritized in DEPICT analysis in women with preeclampsia.

Locus	Gene symbol	Nominal p-value
rs4509570	*IL6R*	2. 44E-08
rs599924	*SURF4*	8. 26E-08
rs7900065;rs8946	*BAG3*	5. 71E-07
rs1110470;rs3024630;rs2239349	*IL4R*	1. 49E-06
rs11159086	*NPC2*	2. 20E-06
rs4646062	*AGMAT*	2. 49E-06
rs2482419	*ABCA1*	2. 56E-06
rs11672613	*C3*	5. 93E-06
rs12726525	*GALNT2*	7. 81E-06
rs748889;rs11048615;rs901835	*ITPR2*	3. 87E-05
rs152033	*ABCC1*	5. 14E-05
rs4853002	*ANXA4*	7. 11E-05
rs1008267	*CSF2RB*	8. 68E-05
rs4646062	*CELA2B*	1. 13E-04
rs655445	*THRSP*	1. 20E-04
rs5743594	*FAM114A1*	1. 38E-04
rs103294	*LILRA3*	2. 39E-04
rs595022	*SRPR*	2. 58E-04
rs1799880	*ADA*	2. 76E-04
rs6065904	*PLTP*	4. 50E-04

2152T

Natural Conception in a Patient with 46,XX/46,XY Chimerism: A Case Report. *L. N. Parsons[1,2], B. Szlendakova[3], P. M. Lemen[3], P. vanTuinen[1,2], J. F. Peterson[1,2].* 1) Department of Pathology, Medical College of Wisconsin, Milwaukee, WI; 2) Wisconsin Diagnostic Laboratories, Medical College of Wisconsin, Milwaukee, WI; 3) Department of Obstetrics & Gynecology, Medical College of Wisconsin, Milwaukee, WI.

Disorders of gender differentiation are a heterogenous group of conditions resulting in a wide range of chromosomal, anatomical, or gonadal features; 46,XX/46,XY chimerism is a rare finding within this category of atypical sexual development. Because of the associated phenotypic characteristics in internal and external sexual organ development in these individuals, fertility and successful pregnancy without the assistance of reproductive medical intervention is exceedingly rare. Herein, we report a case of unassisted conception in a 31-year-old patient with 46,XX/46,XY chimerism. Classical cytogenetic analysis of 77 PHA-stimulated lymphocytes (peripheral blood) revealed a 46,XY karyotype in 75 cells, and a 46,XX karyotype in two cells. Upon further review of the patient's history, it was noted that the patient was born with ambiguous genitalia and had undergone multiple corrective surgeries including perineoplasty, corporal resection, and clitoral resection at 18 months of age, followed by introitoplasty and vaginoplasty at 13 years of age. Abdominal and pelvic ultrasounds in childhood revealed a cervix, unicornate uterus, bilateral fallopian tubes, left ovotestis with epididymis, and right ultrasonographically normal ovary. The patient subsequently underwent removal of the left ovotestis/epididymis, and a biopsy of the right ovary revealed histologically unremarkable ovarian tissue. To date, the patient's pregnancy is at 22 4/7 weeks gestation and of female gender; however, due to the prior surgical interventions described above the patient is not a candidate for vaginal delivery. Thus, we describe an exceedingly rare case of fertility and unassisted conception in a 46,XX/46,XY chimeric patient. It is unclear whether the presence of male cells in this pregnant woman, in proportions and distribution of which is yet not fully known, will influence fetal development.

2153F

Delineating Epigenetic Dynmaics in Spermatogonial Stem Cell during Neonatal Development at Single-Base Resolution. *TL. Lee[1], J. Liao[1], SH. Ng[1], K. Yip[2], Q. Cao[2], WY. Chan[1].* 1) School of Biomedical Sciences, The Chinese University of Hong Kong, Shatin, Hong Kong, China; 2) Department of Computer Science and Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong, China.

Spermatogonial stem cells (SSCs) are male germline stem cells that either self-renew or differentiate to produce progenitor spermatogonia to initiate spermatogenesis. Similar to other stem cells, it is believed that these processes are tightly regulated through precise epigenetic and transcriptional dynamics, and its aberrant alterations could lead to significant developmental and reproductive impact. Although the gene expression and epigenetic profile shift dramatically when SSCs are committing differentiation in adult testis, the interplay between epigenetic and transcriptional regulations in SSCs remains largely unknown in neonatal mice. Here we reveal the genome-wide dynamics of DNA methylation modifications and gene expression in undifferentiated (Kit negative) and differentiating (Kit postive)neonatal SSCs . Epigenetic modifications on DNA methylation and demethylation marks in form of 5-methylcytosine (5mC) and 5-Hydroxymethylcytosine (5hmC) were revealed by whole genome 5mC/5hmC profiling at single-base resolution, whereas the transcriptomes were profiled by RNA-Seq. First, transcriptome analysis suggests striking signaling and transcriptional differences during kit transition, which are characterized by key self-renewal and proliferation pathways. Second, we have unraveled the baseline epigenetic status of SSC in the first wave of spermatogenesis for the first time. To our surprise, in contrast to other stem cell differentiation models, there are no significant global changes in 5mC or 5hmC levels during neonatal SSCs differentiation. Instead, we identified substantial differences on 5mC/5hmC distribution in Kit+ and Kit- SSCs. Third, analysis of the relationship between 5mC/5hmC marks and gene expression suggests that both 5mC and 5hmC at promoter regions repress the expression of the corresponding genes, while 5mC and 5hmC at gene bodies positively correlate with gene expression in neonatal SSCs. Taken together, our results provide comprehensive information of the transcriptional and epigenetic landscapes in neonatal SSC development and highlight the unique 5mC/5hmC signatures associated with Kit+/- SSCs, which may contribute to gene expression regulation. These epigenetic marks and gene targets will allow identification of novel molecular mechanisms in neonatal SSCs development.

2154T

Sexual dimorphism in placental telomere length over gestational age. *SL. Wilson[1,2], Y. Liu[1,2], WP. Robinson[1,2].* 1) Department of Medical Genetics, Child and Family Research Institute, Vancouver, BC, Canada; 2) Medical Genetics Dept. University of British Columbia, Vancouver, BC, CA.

Telomeres are repetitive sequences at the end of the chromosomes that protect chromosome ends from degradation and end fusion. Reduced telomere length (TL) is associated with biological and social stresses, age, and disease state. TL has been reported to be significantly shorter in placentas from pregnancies complicated by preeclampsia (PE) and/or intrauterine growth restriction (IUGR). This has been attributed to oxidative stress and hypoxic conditions that occur in PE/IUGR. However, such effects could alternatively be due to normal variation in telomere length related to gestational age or fetal sex. To further clarify these relationships we assessed whether placental TL was associated with i) gestational age or fetal sex within a control cohort; ii) PE/IUGR after taking into account confounding variables; iii) genome wide changes in DNA methylation (DNAm); and iv) targeted changes in DNAm (such as DNAm changes in *TERT, TERC,* or stress response genes). Average TL was measured by qPCR in 94 control placentas from first trimester to term, as well as 22 early-onset PE (EOPE, <34weeks), 19 late-onset PE (LOPE,>34weeks), and 9 normotensive IUGR placentas. We observed a slight trend towards shorter TL with increasing gestational age, (r=-0. 02, p=0. 09). TL was on average longer in female compared to male placentas (t-stat=-3. 37, p=0. 01). Adjusting for sex, we found no difference in TL when comparing EOPE (p=0. 17), LOPE (p=0. 22), and IUGR (p=0. 27) to control placentas. Of our samples, 18 EOPE, 18 LOPE, 9 IUGR, and 13 controls were also run on the Illumina Infinium HumanMethylation450 array, measuring >480,000 CpG sites across the genome. A p-value distribution comparing the association between DNAm measures at all the probes and TL showed a minimal skew to significance, indicating that some CpG sites may be associated with TL. However, no individual CpG sites were found to be associated with TL once corrected for multiple comparisons using a FDR<0. 05. Nor were any specifically associated with DNAm changes at *TERT, TERC,* or other candidate genes. Our data do not confirm TL as a marker of placental stress in PE or IUGR. Longer TL in female placentas may be attributable to increased estrogen exposure during pregnancy, as estrogen has been found to activate telomerase in a number of tissues, due to the estrogen response element upstream of the *hTERT* promoter. We did not find evidence that genome wide DNAm can explain any of the variance in TL in placental tissue.

2155T

High-security DNA bank accounts to protect and share your genetic identity. *J. Den Dunnen.* Leiden University Medical Center (LUMC), Leiden, Netherlands.

With the cost of genome sequencing decreasing every day, DNA information has the potential of affecting the lives of everyone. In general, an individual has little knowledge about his own DNA information, can rarely access it and has hardly any control over its use. This may result in preventable health-related situations and significantly inhibits its use and scientific progress. What we urgently need is a “DNA bank”, a resource providing a secure personal account where, similar to a financial institution, you store your DNA sequence. For any genetic study performed the data generated must be transferred (paid) to your DNA account, including the report showing what has been tested, what was concluded, and based on what evidence. This demand will stimulate trusted public repositories to emerge where consensus opinions on variants and their predicted consequences are available. To ensure independence and a focus on the individual’s interest, the bank should be run by independent trusted mediator. Regulations could be established and oversight arranged by global governing bodies such as the WHO, UNESCO or others. Using their private and secure DNA bank account individuals govern their sequence-related business, regulate access, knowing for what purpose (informed consent) it will be used and only to the genetic data they are willing to share. Depending on the individual’s wishes, the bank may get the right to share data in summary statistics (e. g. variant and phenotype frequencies). The bank serves as an intermediate between parties that have questions or information in relation to specific variants without revealing a person’s identity. The bank may ask individuals to participate in a genetic study by completing a phenotype questionnaire and share this information anonymously. The DNA account ensures the individual is in the driver’s seat, knows what’s known and controls what’s happening.

2156F

Illumina Methylation Array Information in Public Databases May Be Used to Impute Alcohol and Tobacco Consumption Histories on Potentially Identifiable Individuals. *R. A. Philibert[1, 2], N. Terry[3], C. Erwin[4], W. Philibert[1], S. R. H. Beach[5], G. H. Brody[5].* 1) Psychiatry, University of Iowa, Iowa City, IA; 2) Behavioral Diagnostics Inc. , Iowa City, IA; 3) McKinney School of Law, Indiana University, Indianopolis, IN; 4) Department of Medical Education, Texas Tech University Health Sciences Center, Lubbock, TX; 5) Center for Family Research, University of Georgia, Athens, GA.

Illumina methylation arrays are increasingly used as tools in studies of complex medical disorders. Current policy dictates that data from these arrays, like those from genome wide genotyping arrays, be deposited in publicly available databases. Unlike the genotyping information, access to the methylation data is often not restricted. An underlying supposition in the current non-restricted access to methylation data is the belief is that protected health and personal identifying information cannot be simultaneously extracted from these arrays. This assumption appears to be incorrect. For example, in a recent publication we have shown that smoking status can be inferred with a high degree of certainty from methylation data (Receiver Operator Characteristic (ROC) area under the curve (AUC) =0. 99). Here, we demonstrate the potential likelihood for a similarly powerful imputation (AUC >0. 95) for heavy alcohol consumption using Illumina data. In addition, we show how highly informative genotype fingerprints can be extracted using a simple algorithm. Taken together with recent events demonstrating that personally identifying genetic data can be surreptitiously obtained, these findings suggest that protected health information can be obtained about potentially identifiable individuals from public databases. Furthermore, this risk will grow more pronounced as researchers identify additional manners through which methylation arrays record illness and health behavior. As a consequence, currently accepted practices for handling methylation array data do not sufficiently protect the potentially individually identifiable protected health information contained in public databases. Given the stigma as well as potential social/legal consequences of such de-identification, we suggest that more attention needs to be paid to whether access to genome wide epigenetic data should be restricted to investigators who follow generally agreed practices and accede to data use agreements prohibiting re-identification. .

2157T

Consent, confidentiality, and information-sharing in genetic medicine: ways forward for the UK and beyond. *S. Dheensa[1], A. Lucassen[1,2], A. Fenwick[1].* 1) Clinical Ethics and Law, University of Southampton, Southampton, United Kingdom; 2) Wessex Clinical Genetics Service, University Hospitals Southampton Trust, Southampton.

When patients do not notify their family about relevant genetic risks, healthcare professionals (HCPs) can, according to UK guidelines, disclose information without patient consent when the benefit of doing so outweighs the public and patient's interest in maintaining confidence. US guidelines say the risk must be serious, imminent, foreseeable and likely. These guidelines take an individual, rather than familial, view of confidentiality, and take maintaining confidence as default over disclosure. But should we question these norms? US and UK court cases (e. g. ABC v St George's NHS Trust 2015) have not determined a legal duty to warn, but this does not preclude an ethical responsibility. Several interventional studies show that direct action by HCPs (inviting relatives to counselling/education sessions, or confirming disclosure by checking family files, etc.) increases testing uptake for hereditary cancers among relatives. Genomic medicine will make such issues more important, since HCPs will need to engage family to clarify VOUSs, or test them for IFs. We thus aimed to explore views about consent, confidentiality, and information-sharing. METHOD: Interviews with patients (n=33) and focus groups with 80 genetic HCPs. RESULTS: Patients and HCPs had contrary views. While both considered genetic information to be familial, HCPs in practice fell back on an individual definition of confidentiality, and thought their duty to relatives was discharged by giving patients 'relative letters'. We will outline several reasons for this view. (1) They worried professional involvement could exacerbate poor family relationships. By contrast, patients thought family secrets made family dynamics worse, and favoured HCP intervention. (2) HCPs assumed familial communication would happen eventually without their intervention (e. g. , via word-of-mouth). Yet HCPs gave many examples of where this had not occurred and people developed preventable disease. (3) HCPs assumed they could convince patients to disclose at follow-up appointments. However, patients did not always know they could have such appointments, which were dwindling anyway, due to resource limitations. (4) Resources alone would not lead HCPs to take action: they worried about liability and would want a 'cultural shift', involving all HCPs, towards taking action. CONCLUSION: HCPs must develop a consensus about their responsibility to family members. We will outline practical ways for HCPs to enact this responsibility.

2158F

Patients' attitudes towards collecting and sharing genetic information for clinical care versus research in an integrated healthcare system from participants and nonparticipants of the Kaiser Permanente Research Program on Genes, Environment, and Health. *J. Harris[1], C. G. Tai[2], B. A. Koenig[3], C. Somkin[1].* 1) Kaiser Permanente Division of Research, Oakland, CA; 2) Department of Epidemiology & Biostatistics, UCSF, San Francisco, CA; 3) Social & Behavioral Sci/Inst Hlth & Agning, UCSF, San Francisco, CA.

Historically, many have argued that the objectives of research and clinical care are fundamentally different and to make this distinction clear for research participants, a boundary line is drawn between activities conducted for research purposes and for clinical care. However, especially in the realm of genetics, several activities often cross this boundary. The return of individual results and incidental findings to research participants has the potential to take research activities into a clinical care context. Concurrently, rapid adoption of whole genome or whole exome sequencing in clinical genetics often outpaces current efforts to determine the clinical utility of genetic variants, especially rare variants, and thus many care decisions based on sequencing results can cross into the realm of research. The purpose of this study is to explore patients' understandings about the boundary between clinical care and research in the context of an integrated healthcare delivery system such as Kaiser Permanente (KP). We conducted two focus groups, one comprised of KP members who are participants in the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and a second comprised of KP members who did not participate in the RPGEH. We developed a moderator's guide with examples from clinical genetics and addressed the following questions: How do patients define clinical care and research? What are patients' expectations about research and data sharing within an integrated healthcare system with a comprehensive electronic medical record? How do patients balance tradeoffs between benefits and concerns (i. e. medical progress versus data security and privacy) in blurring or maintaining the boundary between research and clinical care? Does previous participation in a research biobank influence these views? Recordings of focus group discussions were transcribed and qualitatively coded to identify themes and supporting quotations. Findings from this study provide important insight useful for updating research practices to address current challenges in genomic research. Understanding the perspectives of patients and research participants can help improve informed consent procedures, attend to concerns around the therapeutic misconception, and develop appropriate governance policies for data sharing.

2159T

Attitudes about data sharing in the age of personal genome sequencing: Preliminary findings from the PeopleSeq Study. *M. H. Helm[1], S. C. Sanderson[2,3], K. D. Christensen[1,4], T. M. Hambuch[5], M. P. Ball[6], M. D. Linderman[2], N. M. Pearson[7], J. S. Roberts[8], G. M. Church[4], E. E. Schadt[2], R. C. Green[1,4].* 1) Department of Medicine, Division of Genetics, Brigham and Women's Hospital, Boston, MA; 2) Icahn School of Medicine at Mount Sinai, New York, NY; 3) Health Behaviour Research Centre, University College London, London, UK; 4) Harvard Medical School, Boston, MA; 5) Illumina, Inc. , San Diego, CA; 6) PersonalGenomes. org, Boston, MA; 7) New York Genome Center, New York, NY; 8) University of Michigan School of Public Health, Ann Arbor, MI.

Background: As personal genome sequencing (PGS) becomes more common, it is important to evaluate participants' comfort and interest in sharing their genome data (GD). Several programs currently offer PGS to healthy adults for education, research, wellness, and prognostic purposes. Most programs return reports with interpreted results to participants. Additionally programs may provide participants with their GD, which some programs also make available publicly. This issue became even more pertinent as the NIH Genomic Data Sharing Policy took effect in 2015. The PeopleSeq Study surveys ostensibly healthy adults who have undergone PGS to assess their attitudes towards data sharing. Methods: PeopleSeq Study participants that have been surveyed to date are all individuals who paid for PGS through Illumina's Understand Your Genome program, which returns a clinically focused report to participants through their ordering physicians, followed by a chance to explore their results at a 1-2 day symposium. Our survey included questions on participants' comfort with GD sharing. Results: Respondents (n=87) were on average 53 years old, white (86%), with high income (89% annual household income ≥$100,000), educated (90% doctorate or professional degree), married (85%), male (71%), and had biological children older than 18 years (29%). When asked how comfortable they were with the idea of sharing their GD with others, 56% of respondents reported being 'very comfortable,' 29% 'somewhat comfortable,' 7% 'somewhat uncomfortable,' and 5% 'very uncomfortable. ' Three percent of respondents reported that they would not consider sharing their GD; 33% reported they would share it with select individuals if it remained confidential; 56% would share it publicly, but anonymously; and 30% would share it publicly even with their identity attached. Most respondents thought it would be 'difficult' (25%), 'very difficult' (30%), or 'impossible' (13%), for someone to identify them via only their GD, while others thought it would be 'not very difficult' (23%), or 'easy' (6%). Conclusion: Overall, PeopleSeq Study respondents were comfortable with GD sharing, depending on how the data would be shared. Most thought it would be challenging to be identified by GD alone, but almost 1/3 recognized that GD could be linked to identity. Future goals of this study include comparing other PGS cohorts to explore how attitudes towards GD sharing vary between different models of results disclosures.

2160F

Harmonizing Privacy Laws to Enable International Biobank Research. *M. Rothstein[1], B. Knoppers[2].* 1) Inst Bioethics, Univ Louisville, Louisville, KY; 2) Centre of Genomics and Policy, McGill University, Montreal, CA.

Large-scale genomic research increasingly relies on international collaboration in compiling and analyzing biological specimens and associated medical information. In some countries, however, laws directed at biobanking, research, or privacy present obstacles to the cross-border sharing of human specimens or health data. Pursuant to an NIH-funded grant, we identified 25 countries engaging in substantial biobank-enabled genomic research. For each country, we invited a leading authority (or authorities) on the biobanking, research, and privacy laws of the country to research and analyze the applicable laws. Using a standard template, they considered the state of biobanking in the country, the privacy framework, the governance of the biobank, and the barriers to access and use of the specimens and data for research. Then, they drafted an interdisciplinary article to be published in the *Journal of Law, Medicine & Ethics* in 2016. The lead investigators distilled the information and prepared a series of recommendations to facilitate the use of biobank resources for genomic research while maintaining essential privacy protections. This presentation will be the first public discussion of the preliminary findings and recommendations of the study.

2161T

Can an online decision-aid help parents of children with treatment-resistant epilepsy make whole exome sequencing decisions? *S. Adam[1], N. Bansback[2], P. H. Birch[1], R. Coe[1], M. B. Connolly[3], E. Toyota[3], M. J. Farrer[4], M. K. Demos[3], J. M. Friedman[1].* 1) Dept Medical Genetics, Univ British Columbia, Vancouver, BC, Canada; 2) School of Population and Public Health, Univ British Columbia, Vancouver, BC, Canada; 3) Div Neurology, Dept. Pediatrics, Univ British Columbia, Vancouver, BC, Canada; 4) Centre for Applied Neurogenetics, Djavad Mowafhagian Centre for Brain Health, Univ British Columbia, Vancouver, BC, Canada.

As clinical genome-wide sequencing becomes more widely utilized by non-genetics specialists, alternatives to conventional genetic counselling, which is often available only through clinical genetics services, are necessary to optimize decisional support and informed consent for families. We have developed an interactive online educational tool and decision aid called "DECIDE" to help families make choices about whole exome sequencing (WES) for diagnosis of serious mendelian diseases. To determine the efficacy of this tool, we are testing it with parents who are considering research WES for a child with treatment-resistant epilepsy through the Neurology Department of a tertiary children's hospital. Prior to testing, families are randomly assigned to receive either DECIDE first or counselling by an experienced genetic counsellor (GC) first, followed by the other intervention. Outcome measures assessed before, between and after the interventions include patient knowledge, empowerment, and decisional quality and certainty. Results on the first 30 families studied show that mean (SD) knowledge score at baseline was 9. 0 (5. 1) out of 20, increasing to 13. 8 (3. 1) after the first intervention ($p< 0. 0001$). There was no significant difference in knowledge improvement between the two study arms, and there was little additional gain in knowledge between the first and second interventions. Empowerment scores also increased after each intervention, but the change was not statistically significant, and there was no significant difference in empowerment between the groups. Most parents in both groups reported that they were confident about their decisions, had sufficient support and advice, understood the pros and cons of testing, and knew which risks/benefits mattered most to them. Almost all parents chose WES for their child. Issues identified as most important in making this decision were: "getting an answer", the possibility of better management, better access to resources, and helping others/advancing science, each of which was mentioned by over 2/3 of participants. Impact on insurability and future reproductive planning were each important for about 1/3 of families. Follow-up analysis is ongoing, and 70 more families are currently being enrolled in the study. As WES use by non-genetics specialists becomes widespread, use of tools such as DECIDE may improve pre-test decision-making and consent processes.

2162F

DECIDE: Development of an interactive tool to guide parents' choices regarding genomic incidental findings results. *J. M. Friedman[1], S. Adam[1], P. H. Birch[1], R. Coe[1], J. Hicklin[1,2], N. Bansback[2].* 1) Dept Medical Genetics, Univ British Columbia, Vancouver, BC, Canada; 2) School of Population and Public Health, Univ British Columbia, Vancouver, BC, Canada.

We have developed DECIDE, an online educational tool and decision aid, to support genetic counselling and decision-making in individuals considering trio diagnostic genome-wide sequencing (GWS). The content was determined through interviews and focus groups with parents who had undergone research GWS, from discussions with genetic health professionals, and from the literature, and was modified after usability testing of pilot versions. DECIDE presents multi-media information about the genetics of disease, the testing process and return of results, including variants of unknown significance, at varying user-selected levels of detail and complexity. Users' decisions are guided by facilitating consideration of consequences in the context of their own priorities and values. Users first consider whether or not to have diagnostic GWS on the basis of issues such as the effect of GWS results on availability of resources for their child and family, the possibility of management options, and implications for insurance or reproductive decisions. A second component of DECIDE informs and guides users' choices regarding receipt of incidental findings (IF). To simplify choices, IF are binned into conditions that are typically of childhood onset or typically present after childhood, and, in each case, those with and those without effective treatments. DECIDE suggests choices based on values assigned by the software to each issue, weighted according to users' responses, but the user can accept or decline these suggestions. Decisions about whether or not to have diagnostic GWS and about return of IF for each of the four bins are made separately, and each parent makes choices for her- or himself and separate choices for the affected child. Assessments of users' knowledge and understanding of the consequences of having GWS and of the quality and certainty of each decision are integrated into DECIDE. The decisions of record are made during subsequent genetic counselling and consent sessions. DECIDE offers a number of advantages, including accessibility at times and locations convenient to users, participation by multiple family members, choice of multiple levels of complexity and learning formats, and the opportunity to review information at a later date. The utility and cost-effectiveness of DECIDE with respect to improving genetic counselling and consent for diagnostic GWS is currently being tested through randomized controlled trials.

2163T

***e-consultative* informed consent for a social research in health among teenagers with Neurofibromatosis type 1.** *F. Duplain-Laferrière[1], G. Bouffard[1], D. Drouin[1,2], C. Bouffard[1].* 1) Laboratory of Transdisciplinary Research in Genetics, Medicines and Social Sciences, Div Genetics, Dept Pediatrics, Fac Med Health Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada; 2) Dept Pediatrics, Fac Med, Laval University, Quebec City, QC, Canada.

Following an ethnographic study conducted among adolescents with Neurofibromatosis type 1 (NF1), aged 15 to 19 years, we were interested to know their concerns regarding the impact of this genetic condition on their social life (*school, work, society*), their love-life and intimate relationship, their reproductive expectations, the quality of family and social interactions (peers, teachers, friends, etc.) and intergenerational dialogue. During the preparatory phase of the study, we realized that face-to-face interviews would not work with adolescents. They agreed to participate, but responded in an evasive manner. This attitude was in stark contrast with the enthusiasm they had demonstrated for the study. To better understanding the situation, we consulted a group of late adolescents affected by NF1. They identified three sources of problems due to the intimate nature of the questions. They were uncomfortable with 1) the idea of their parents consenting for them as they are underage, 2) discussing these topics with an adult face-to-face, and 3) the possibility that the researchers were likely to meet their parents in NF associations activities. Eventually, they advised us to opt for an anonymous online questionnaire. While this approach allowed them to exercise their decision-making autonomy and to freely express themselves anonymously by using a familiar mode of communication, it required more care about the ethical issues raised by online consents and the consent of a teenager. In this context, we developed an *e-consultative* informed consent. The consultative part consists to ask them to give their opinions on the ethical dimensions of the study and the relevance of the research questions and objectives. They had to answer all the questions to get access to the research questionnaire. This presentation shares our results on the effectiveness of this new form of consent. In addition to informing the teenagers, it increases the probability that it will be read and understood. Despite some improvements, the *e-consultative* informed consent form increases the chances that the participants will read and understand the issues of the research, while having their intelligence and taking their needs into consideration.

2164F

Incidental findings; figures and hurdles of informed consent in a diagnostic setting. *M. Kriek[1,2], M. J. V. Hoffer[1], R. E. van Hellemondt[3], S. G. Kant[1], E. K. Bijlsma[1], M. Koopmans[1], M. E. Y. Laurense-Bik[1], A. C. J. Gijsbers[1], C. J. van Asperen[1], G. W. E. Santen[1], A. Tibben[1], C. A. L. Ruivenkamp[1].* 1) Department of Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Department of Human Genetics, Leiden University Medical Center, Leiden, Netherlands; 3) Department of Ethics and Law, Leiden University Medical Center, Leiden, Netherlands.

The purpose of whole exome sequencing in diagnostics is to increase the number of persons with a molecularly verified diagnosis. In accordance with European guidelines, variants will initially be sought in genes with a known relationship to the patient's disease (targeted analysis). If targeted analysis fails to yield pathogenic variant(s), the next step is exome-wide trio analysis. During this step, findings may emerge that are related to a disease other than that for which DNA testing was initiated. In our clinical diagnostic setting, these incidental finding (IF) are divided into two categories:a) Findings suggestive of a condition for which prevention or treatment is available. b) Findings suggestive of a condition for which no treatment/prevention program is available at present but knowledge of the condition may be important to the patient or family members. Due to the health implications, IF under category a) are always reported to the patient/ parents. IFs in category b) are only reported if one has agreed to be informed of such findings. Should patient/parents decide not to be informed of IFs at all, exome analysis will remain restricted to relevant disease genes. These choices are discussed in a pretest counselling by a clinical geneticist and before exome-wide analysis is initiated, a signed informed consent (IC) form is mandatory. From January 2014 till June 2015, exome sequencing was requested 331 times of which 69 comprised targeted analysis and 262 included exome-wide analysis. The main reason for a targeted approach included the inability to obtain DNA from both parent (62/69: 90%), in two cases no IC form was returned and the remaining 5 cases wanted to avoid the chance of identifying IFs. Fifteen of the 262 exome-wide cases initially forgot to indicate an IF category on the IC form. In 16 cases only one parent gave consent for a minor. From the remaining 231, 50 cases chose category a) (22%) and181 chose category b) (78%). This survey shows that the vast majority of patients wants to be informed about all IFs. However, there is a small but non-negligible subset of patients (5 out of 331: 1. 5%) who doesn't want to be informed about any IF. By offering patients/ parents the opportunity to choose their most appropriate way of reporting IFs, they can play an active role in this complex but important decision process.

2165T

High participation and engagement rate in a systematic hospital-based genomic medicine research project with broad consent. *V. Mooser[1], C. Currat[1], L. Chapatte[1], C. Roth[1], M. Bochud[2].* 1) Biomedicine Service, Lab Department, CHUV University Hospital, Lausanne, Switzerland; 2) Institute for Social and Preventive Medicine, Lausanne, Switzerland.

BACKGROUND AND STATEMENT OF PURPOSE :Little is known about hospitalized patients willingness to engage into genomic research project. This type of knowledge is critical for the planning of precision medicine initiatives. The purpose of the present study was to evaluate the participation rate of hospitalized patients to engage in a systematic hospital-based genomic medicine project based in Switzerland, and to identify factors associated with this engagement. METHODS :Within the framework of the Lausanne, Switzerland, University Hospital Institutional Biobank (BIL), inpatients and selected outpatients are systematically invited to grant researchers broad access to their biomedical data and to donate blood for future whole genome sequencing. Additionally participants are offered the options to be re-contacted in case of incidental findings and to receive an electronic newsletter. Multivariable logistic regression analysis was used to identify personal factors associated with willingness to participate in BIL and with interest in these options. Analyses were restricted to the initial 11099 invited patients for whom full dataset was available, and were stratified by age groups. RESULTS :Overall participation rate was 82. 4% (9141/11099) and was higher in the < 64-year old group (odds ratio [OR] 1. 70; 95% confidence interval [CI] 1. 53 to 1. 90). In the ≥ 64-year old group, participation was lower among women (OR 0. 77; 95% CI 0. 68 to 0. 89), among non-Swiss citizens (OR 0. 66; 95% CI 0. 55 to 0. 79) and those with emergency admissions (OR 0. 59; 95% CI 0. 51 to 0. 69). A total of 8576 (93. 8%) and 3020 (33. 0%) participants were willing to be re-contacted for incidental findings and to receive the newsletter, respectively. CONCLUSIONS :A large proportion of patients are willing to actively participate in this particular systematic hospital-based genomic research program. Hospitals adopting broad consent represent an efficient setting to recruit participants into precision medicine initiatives.

2166F

Family tree and ancestry inference: is there a need for a 'generational consent'? *S. Wallace, E. G. Gourna, V. Nikolova, N. A. Sheehan.* Health Sciences, University of Leicester, Leicester, United Kingdom.

Purpose: Genealogical research and ancestry testing, long used in scientific and demographic research, are now a popular recreational activity. Clients provide their self-reported and/or DNA to ancestry or direct-to-consumer companies and can receive information including ethnic origins, possible relatives, and health status. But little research has gone into the impact of the use of these services on biological and social families. The ASHG has noted the need to investigate these issues because of questions around the privacy and security of ancestry-related databases. We examined consent materials, which are designed to inform individuals about the implications of participation, for language on the implications for third parties. Method: We used content analysis to examine web pages, consent materials, privacy statements, terms of condition and other applicable documents provided to potential clients of ancestry and direct-to-consumer genetic testing companies to determine what consent is required, what risks associated with participation were mentioned, and whether the consent or notification of third parties was required. Results: We identified four categories of companies: 1) those providing services based only on self-reported data, such as personal or family history; 2) services based only on DNA provided by the client; 3) services using both; and 4) services using both that also have a research component. Language varied across the companies. Category 1 companies generally stated that clients should obtain consent from, or notify, third parties when using their services. Category 2 companies made no mention of third parties. Category 3/4 companies make some reference to the implications for third parties but do not specifically call for them to be informed about a client's participation. While in some cases such language was prominently placed, more often it was in text within terms and conditions or privacy statements that may not be read by potential clients. Discussion: Because both family history and DNA can provide information on more than just the individual providing them, companies should make it clear that clients should inform third parties about plans to participate and possible implications. More generally, ways that third parties can be involved in consent need to be explored. Can there be a 'generational consent' that would include more than just the individual in decisions about participating in genetic investigations?

2167T

Consenting in the time of genomic sequencing: Experiences from the BabySeq Project. *C. M. Weipert[1], M. C. Towne[2], S. Pereira[3], J. O. Robinson[3], P. B. Agrawal[2,4,5], R. B. Parad[4,5,6], I. A. Holm[2,4], A. L. McGuire[3], A. H. Beggs[2,4], R. C. Green[1,7].* 1) Department of Medicine, Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 2) Division of Genetics and Genomics, Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA; 3) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 4) Department of Pediatrics, Harvard Medical School, Boston, MA; 5) Division of Newborn Medicine, Boston Children's Hospital, Boston, MA; 6) Department of Pediatric Newborn Medicine, Brigham and Women's Hospital, Boston, MA; 7) Partners Center for Personalized Genetic Medicine, Boston, MA.

Background: Genomic newborn sequencing (gNBS) has the potential to provide comprehensive and clinically useful information on a wide range of conditions. The BabySeq Project is a randomized, controlled trial exploring the utility of gNBS for parents and healthcare providers. Two cohorts incorporating 240 infants and their parents at Brigham and Women's Hospital (BWH) and 240 from the Neonatal Intensive Care Unit at Boston Children's Hospital (BCH), are being randomized to standard newborn screening (NBS) alone or gNBS plus NBS. Genomic results related to childhood conditions are being returned to the gNBS arm. Given the breadth and complexity of potential results, and uncertainties surrounding the downstream impact of genomic information, the development of the informed consent process was complex, compounded by the logistics of obtaining consent and samples in the short, post-partum period when the infant is in the hospital. Our goal was to address these challenges by creating a multi-step, in-depth consenting process. Methods: Utilizing feedback from our IRBs at BWH and BCH, we designed a consent process that includes educational tools and a series of 'Consent Understanding Questions' for parents, intended to be minimally invasive but maximally informative. In light of potential future ethical implications, eligibility criteria include requiring consent from both biological parents (if known) and excluding multiples. Only parents who clinical staff feel are capable of making decisions in the post-partum period are approached. Results: Enrollment began in May 2015. Four out of the 9 initial families approached have agreed to participate (2/5 at BCH and 2/4 at BWH). Reasons for declining participation include feeling overwhelmed, the time commitment required, concerns about genetic discrimination and declining health status of the infant. The average time for the consent process was 65 minutes, and the average score for the 'Consent Understanding Questions' was 17. 3/18 (n=4). We will be actively enrolling more participants in the upcoming months and will plan to present our most up-to-date numbers. Conclusion: The development and use of a comprehensive informed consent process has allowed us to explore gNBS in both healthy and sick newborn populations. This study will provide important insights into the thoughts and feelings of parents regarding gNBS, while also providing essential data on the feasibility of implementing gNBS on a population-based level.

2168F

Context Matters. Attitudes, Experiences and Meanings of Consent to Stakeholders within Newborn Screening. *K. W. Miller[1], S. G. Nicholls[1,2], H. Etchegary[3], L. Tessier[2,4], C. Simmonds[3], B. K. Potter[1], J. C. Brehaut[5], D. Pullman[3], R. Z. Hayeems[6], S. Zelenietz[4], M. Lamoureux[4], J. Milburn[4], I. Turner[7], P. Chakraborty[4], B. J. Wilson[1].* 1) University of Ottawa, Ottawa, Ontario, Canada; 2) Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 3) Memorial University, St. John's, Newfoundland and Labrador, Canada; 4) Newborn Screening Ontario, Canada; 5) Ottawa Hospital Research Institute, Ottawa, Ontario, Canada; 6) Hospital for Sick Children Research Institute, Toronto, Ontario, Canada; 7) Provincial Medical Genetics Program Eastern Health, St. John's, Newfoundland and Labrador, Canada.

Background: Newborn bloodspot screening (NBS) is the program through which babies are screened for a variety of conditions shortly after birth. In Canada, NBS generally proceeds on an implied consent basis, potentially generating an ethical tension between the goals of promoting high uptake and of promoting fully informed consent. **Objectives:** Examine how current consent practices are described and experienced by different stakeholders, and how discussions of participation in NBS map across different definitions of consent practices. **Methods:**Aligned with qualitative methodology, semi-structured interviews were conducted with parents, healthcare professionals, and policy decision-makers. The study was undertaken in two Canadian provinces (Ontario and Newfoundland & Labrador). Interviews were audio recorded, transcribed, and analyzed for key themes. **Results:**Participant responses were context dependent: in addition to individual experiences, participant conceptions of key terms appeared to influence attitudes. Major themes related to varied responsibilities and requirements of parents and professionals and the level of expectation for parental involvement. A minor theme related to a perceived overemphasis on individual consent as an ethical panacea, at the expense of alternate governance structures. **Conclusions:**Findings highlight variability in experiences of the process of obtaining informed consent as well as how consent approaches are perceived by different stakeholders in NBS. The results further our understanding of attitudes toward consent and how these affect the process in practice, illustrating that the context and language of consent plays a key role in attitudes to consent-related issues for newborn screening. These results have informed the next stage of research: a self-completion survey involving a larger, more representative sample of the same stakeholder groups will differentiate attitude categories and assess whether attitudes are associated with stakeholder grouping or other important participant characteristics.

2169T

Ethical, Legal, and Social Implications of Personalized Genomic Medicine Research: Current Literature and Suggestions for the Future. *S. Callier[1], R. Abudu[1], M. Mehlman[3], M. Singer[4], D. Neuhauser[4], G. Wiesner[2].* 1) George Washington University School of Medicine and Health Sciences, Washington, DC; 2) Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee; 3) The Law-Medicine Center, Case Western Reserve University School of Law, Cleveland, Ohio; 4) Case Western Reserve University School of Medicine, Cleveland, Ohio.

Purpose: The goal of this systematic review was to assess the ethical, legal, and social implications (ELSI) literature on personalized genomic medicine (PGM) research. **Methods:** The abstracts of 2,029 articles extracted from scholarly databases and published during a 5-year period (2008-2012) were systematically reviewed. A total of 402 articles met our research criteria and were organized thematically to assess the representation of ELSI issues, stakeholders, health specialties, journals, and empirical studies. **Results:** Empirical and normative ELSI analyses were published in both scientific and ethics journals. Investigational research comprised 34 percent of the literature reviewed (136 articles) and the remaining articles (266 or 66 percent) comprised normative analyses, commentaries, and news stories. Traditional ELSI concerns dominated the discourse including discussions about disclosure of research results. In fact, there was a dramatic increase in the number of articles focused on disclosure and the implications of discussing incidental findings with research participants. Few ELSI papers focused on particular health disorders, the use of racial categories in research, low and middle income countries, or special populations, such as adolescents, elderly patients, or ethnic groups. **Conclusion:** Considering that personalized medicine will become more tailored to individuals, groups and populations in the future, further analysis is needed on how ELSI scholarship can better serve the growing global, multidisciplinary, and translational goals of PGM research.

2170F

Direct to Consumer Genetic Testing – exploring the ethical and legal dilemmas and considering regulation in Europe. *A. C. de Paor.* School of Law and Government, Dublin City University, Dublin 9, Ireland.

With rapid scientific and technological advances, the past few years has witnessed the emergence of a new genetic era and a growing understanding of the genetic make up of human beings. These advances have propelled the introduction of companies offering direct to consumer (DTC) genetic testing, which facilitates the direct provision of such tests to consumers, (for example via the internet). Although DTC genetic testing offers benefits by argaubly enhancing consumer accessibility to such technology, promoting proactive healthcare and increasing genetic awareness, it presents a myriad of challenges, from an ethical, legal and regulatory perspective. As DTC genetic testing usually eliminates the need for a medical professional in accessing genetic tests, this lack of professional guidance and counselling may result in misinterpretation and confusion regarding results. In addition, an evident concern relates to the scientific validity and quality of these tests. A further problem arising is the lack or inadequacy of regulation in this field, which may result in privacy violations as well as other unauthorised use of genetic information, and potentially discrimination. This legislative vacuum is apparent at European Union (EU) level, where there is no concrete legislation, as well as amongst EU Member States, few of which have introduced regulatory frameworks specifically addressing this area. In light of the increasing accessibility of DTC genetic testing, and the lack of a comprehensive regulatory regime in Europe (and indeed elsewhere), a need arises to highlight the emerging ethical and legal dilemmas and consider an up to date regulatory framework. Further, with inevitable advances in science, as well as increasing internet accessibility, the challenges presented are likely to become more amplified. In consideration of the ethical and legal challenges arising, this paper highlights the regulation of DTC genetic testing as a key concern of the EU, recognising its importance to both the scientific community as well as in respect of enhancing consumer confidence in such technologies.

2171T

Regulating Genome Editing Technologies: Towards a Proportional Approach. *R. Isasi, E. Kleiderman, B. M. Knoppers.* Centre of Genomics and Policy, McGill University, Montreal, PQ, Canada.

Human genome editing represents an innovative and powerful tool offering great scientific and therapeutic promise. Yet, it has rejuvenated old socio-ethical and legal (ELSI) debates and society is faced, once again, with assessing the implications of intentional human germline modification. Professional organizations and regulatory and funding agencies are calling for the scientific community to proceed with caution by establishing robust frameworks so as to assess the safety and efficacy of this technology, while also considering its socio-ethical dimensions. For some, a cautious approach represents adopting tiered policy measures, for example, by means of self-imposed moratoria limited to research and/or clinical activities directed at germline modification. For others, it means maintaining a vigilant attitude while allowing the gathering of evidence on the benefits and risks of the technology. Still, for some, a precautionary approach entails adopting outright legal restrictions that could go as far as limiting somatic genome modification. To contextualize and assess the strengths and weaknesses of current proposals, and to help inform ELSI and scientific debates, this presentation will focus on the results of an international comparative study of policy frameworks regulating somatic and germline modification in 16 jurisdictions (e. g. Australia, Belgium, Brazil, Canada, China, France, Germany, India, Israel, Japan, Mexico, Singapore, South Korea, the Netherlands, the United Kingdom and the United States). We will address convergence and divergence of positions on issues related to the permissibility of germline and somatic technologies in research and clinical applications, the effectiveness of current policy restrictions and the adequacy of existing oversight mechanisms. Finally, we will identify strategies to aid in the development of a coherent approach to consensual policy and promote proportional oversight of genomic technologies.

2172F

Ethical and Legal Issues Associated with the Use of Genomic Information in the Context of Pandemic Influenza. *M. Lewis, T. Bailey.* Johns Hopkins University, Baltimore, MD.

Advances in genomics are contributing to the growing body of evidence about genetic determinants of infectious disease in humans. President Obama's Precision Medicine Initiative likely will further drive scientific inquiry to deepen our understanding of how genetic factors can impact the human body's response to infectious agents, their treatments, and vaccines developed to prevent disease. In the scientific community, there is general consensus that a significant influenza pandemic is inevitable. In the near future, it may be possible to classify individuals as having increased or decreased susceptibility to influenza based upon the individual's genetic profile. Specific genetic variants have been shown to be associated with development of severe disease. In contrast, other variants have been shown to limit the mortality and morbidity associated with influenza infection. If an influenza pandemic (or other type of outbreak) with significant morbidity and mortality were to occur, this type of genetic information could be used to facilitate decision-making in policies related to quarantines, workforce availability, the allocation of scarce resources, and other issues. This presentation will examine the important ethical and legal issues that could be raised if genetic information is used in this way and will explore the extent to which state and federal laws, including genetic privacy laws, could potentially impact the use of genetic information in this context. The presentation will develop a framework to consider the legal and ethical issues raised by the use of genetic information in this setting. This type of framework could be used to consider in real time the ethical, legal and social challenges that could arise.

2173T

What are the clinical and socioethical recommendations of researchers and physicians about inherited chromosomally integrated human herpesvirus 6? *V. Noel[1], R. Drouin[1,2], L. Flamand[3], C. Bouffard[1].* 1) Laboratory of Transdisciplinary Research in Genetics, Medicines and Social Sciences, Div Genetics, Dep. Pediatrics, Fac. Medicine & Health Sciences, Univ Sherbrooke, Sherbrooke, QC, Canada; 2) Dep. Pediatrics, Faculty Medicine, Laval Univ. , Quebec City, QC, Canada; 3) Dep. microbiology, infectious disease and immunology, Faculty of Medicine, Laval Univ. , Quebec City, QC, Canada.

The human herpesvirus 6 (HHV-6) has the unique capacity to integrate telomeres of the chromosomes (ciHHV-6) of nucleated cells after infection. ciHHV-6 can therefore be found in gametes and be transmitted from one generation to the next and through organ and tissue donation. This worrisome situation affects between 40 and 70 million people around the world and may be associated with illnesses (encephalitis, infertility, cancers, etc.), of which some are fatal. Despite its consequences, this phenomenon has attracted little interest in scientific and public health milieus. To highlight the importance of this issue, several researchers and physicians go so far as to publish warnings in basic research articles in purely scientific journals. This uncommon phenomenon, prompted us to investigate these concerns. Thus, our objectives are to know: 1) the positions of these researchers and physicians about the current situation, and 2) the solutions they propose in order to identify their clinical and socio-ethical concerns. **METHOD**: Qualitative design used asynchronous online text-based interviews (combine questionnaire and chat box) with 24 researchers and physicians working on HHV-6. Qualitative generale inductive analysis with the use of NVivo10 software. **RESULTS**: The participants suggested four types of interventions: 1) the establishment of a screening program in the context of grafts and transplants; 2) medical information on the integrative capacity of HHV-6 and pathogenicity of ciHHV-6; 3) reevaluation of types of antiviral drugs and doses to counteract the adverse effects of pharmacological treatments and standardize the therapeutic treatment plan; 4) standardization of diagnostic tests to prevent false positives or negatives. Participants reflect on the clinical and ethical scopes of a screening program in the contexts of clinical practices and grafts or transplants. **CONCLUSION**: It was very interesting to note the convergence of opinions and the recommendations of the participants. The concerns are numerous and justified considering the lack of knowledge about the ciHHV-6's clinical and ethical consequences. These first data on the topic respond to needs expressed by researchers and physicians concerned by the ciHHV-6 in their practices. In this sense, the results of this research can guide the thinking and decision-making process to improve and harmonize practices and the wellbeing of affected individuals and populations.

2174F

Attitudes of genetics professionals and anthropologists toward race, ancestry, and genetics: results of a national survey. *J. Yu[1], S. C. Nelson[2], T. M. Harrell[1], J. K. Wagner[3], C. D. M. Royal[4], M. J. Bamshad[1,5].* 1) Department of Pediatrics, University of Washington, Seattle, WA; 2) Institute for Public Health Genetics, University of Washington, Seattle, WA; 3) Department of Anthropology, Pennsylvania State University, State College, PA; 4) Department of African and African American Studies, Duke University, Durham, NC; 5) Department of Genome Sciences, University of Washington, Seattle, WA.

Genetics professionals (GPs) influence public opinion about race and ancestry. Claims about population genetics research and GPs' attitudes toward these topics are pervasive in the public domain and contribute to ongoing debates about the role of race and genetic ancestry in research, medicine, and society. Indeed, some of the most vocal concerns about GPs' use of race and genetic ancestry have come from anthropologists. We assessed GPs' and anthropologists' perspectives on the relationships between race, ancestry, and genetics to identify concerns and inform the development of relevant policies. We conducted a 49 item online survey with GPs (n=1513) and anthropologists (n=3286) about race, ancestry, and genetics and their relevance to science, health, and society. Notably, responses to common statements about race differed between GPs and anthropologists. For example, a majority of GPs disagreed with common statements such as "races don't exist" (61%) and "race has no genetic basis" (67%) whereas most anthropologists agreed with such statements. The majority of respondents (52%) agreed that genetic ancestry testing should be subject to federal regulation and rejected its use in adoption (80%), university admissions (88%), and by law enforcement (57%). More than 70% of respondents agreed that the term "race" should be replaced by a more appropriate and precise term, but fewer GPs (51%) than anthropologists (71%) agreed that the use of the term "race" to describe human groups should be discontinued. Nearly 80% of respondents agreed that genetic ancestry—inferred from genetic markers—rather than race, is a better proxy for genetic relationships among sub-Saharan Africans, Asians, Europeans, Pacific Islanders, and Native Americans; yet fewer anthropologists (37%) than GPs (61%) agreed that continental populations – Africans, Asians, Europeans – are useful for examining genetic relationships (i. e. relatedness) among people. Qualitative analysis of 515 GPs' survey comments revealed that GPs recognize many different definitions of race that they use selectively depending on context. Overall, GPs and anthropologists appear to disagree about the biological relevance of race yet mainly agree in many other respects about the use of race and genetic ancestry. We suggest that recommendations about the use of race and genetic ancestry should move beyond debate over definitions and instead focus on specific contexts such as in scientific research, medicine, and society.

2175T

An Unreasonable Lack of Accommodation? GINA and the Legal Issues of Using Human Genetic Information in Workplace Infectious Disease Control. *T. C. Bailey, M. H. Lewis.* Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD.

The 2008 Genetic Information Nondiscrimination Act (GINA) broadly prohibits uses of genetic information in employment-related decisions concerning hiring, firing, and the terms and conditions of employment. In this, it proceeds from the flawed assumption that there are no scenarios in which genetically differentiated employment decisions are on balance preferable and justifiable in comparison to GINA-compliant genetically neutral employment policies. In this paper, we challenge this assumption in the context of workplace infectious disease control policies like mandatory employee vaccination, infectious disease screening and treatment requirements, and work assignments that under GINA must be applied in a genetically neutral fashion without consideration of individual human genetic factors that affect the acquisition, clinical course, transmission, testing, treatment, or prevention of infectious diseases. We contend that GINA unreasonably precludes employers from taking cognizance of and accommodating the genetically heightened risks (as opposed to phenotype based risks) that some employees may face compared to others from genetically neutral workplace infectious disease control policies. We defend this contention with a focused analysis of GINA's application to three hypothetical health care employment scenarios involving genetically neutral workplace infectious disease control policies applied to employees with genetic factors that adversely affect the course or control of infectious diseases:1) A genetically neutral workplace vaccine mandate applied to employees with information of personal genetic factors associated with heightened propensity to develop serious adverse reactions to the required vaccine or reduced propensity to develop protective immunological responses to the vaccine. 2) A genetically neutral work assignment of employees to care for patients with infectious diseases despite their having information of personal genetic factors associated with heightened propensity to acquire those diseases or, if infected, develop clinically severe forms of the diseases. 3) A genetically neutral annual screening of employees for particular infectious diseases with subsequent requirement of treatment applied to employees with information of personal genetic factors associated with reduced propensity to develop transmissible disease or with heightened propensity to develop serious adverse effects to the required treatment.

2176F

Crafting Tools to Engage the Public in Newborn Screening Policy. *R. Ryan[1], B. Koenig[1], H. Longstaff[2, 3], J. Harris-Wai[1, 4].* 1) Institute for Health and Aging, University of California, San Francisco, San Francisco, CA, CA; 2) Engage Associates, Vancouver, BC Canada; 3) Office of Research Ethics, Simon Fraser University, Burnaby, BC Canada; 4) Kaiser Permanente Northern California, Division of Research, Oakland, CA.

Nearly every child born in the U. S. is screened for multiple lethal or debilitating inherited disorders at birth through state-run newborn screening programs (NBS). Recently, two NBS programs were sued following high profile media coverage, resulting in the destruction of over 5 million bloodspots. In the face of growing public concern and pressure for expansion based on technological innovation, it is critical to seek authentic community guidance of NBS programs. While there are several promising models of community engagement for genomics policy development, few of these models have been systematically evaluated. The purpose of this analysis is to describe: 1) the development and implementation of a deliberative community engagement (DCE) convened to inform policy decisions in the California NBS program and 2) the creation of novel evaluation tools to examine whether and how our DCE model was effective at obtaining informed, legitimate, and representative input for the CA NBS program. We use a deliberative democracy framework to inform our four-day DCE. The first phase utilizes various communication formats to educate 30 deliberants about the NBS program from a diverse range of perspectives. Afterwards, deliberants participate in large and small group discussions to generate a set of policy recommendations on several decisions facing the NBS program. These include how stored bloodspots can be shared with researchers in a way that maintains public trust, how and when permission should be obtained from families for research use of bloodspots, and ongoing community involvement in the newborn screening program. The four key variables used to evaluate the DCE are: 1) quality of recommendations; 2) legitimacy of process and recommendations; 3) capacity for future recommendations; and 4) impact of the DCE on participant knowledge and attitudes about NBS and the DCE process. We use a mixed-methods approach to assess and evaluate each of these key constructs, including pre/post-test surveys, transcript analysis of selected group discussions, post-DCE survey, and expert review of the DCE and final recommendations by an external evaluator using a structured checklist. Few deliberative engagements conduct systematic evaluations, and selecting evaluation criteria has been noted as a distinct challenge to the field. Our evaluation tools are useful in evaluating the quality, process, and legitimacy of the DCE and its impact on deliberants and policy outcomes.

2177T

Advanced cancer patients' attitudes regarding incidental findings arising from tumor genomic profiling. *J. G. Hamilton[1], E. Shuk[1], M. Genoff[1], P. Gaissert[2], V. M. Rodríguez[1], J. L. Hay[1], K. Offit[2], M. E. Robson[2].* 1) Department of Psychiatry & Behavioral Sciences, Memorial Sloan Kettering Cancer Center, New York, NY; 2) Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY.

Background: Tumor genomic profiling is used in patients with advanced cancers to guide targeted treatments and clinical trial participation. This process involves sequencing a patient's germline DNA to aid in the interpretation of somatic variants, and can thus reveal complex incidental findings regarding a patient's inherited disease risks. As the medical genetics community debates the management of incidental findings, it is critical to understand how cancer patients view the implications of this information for themselves and their families. Method: We interviewed 40 patients with advanced breast, bladder, colorectal, or lung cancer who had tumor genomic profiling at our institution (10 patients for each cancer; 63% female, 85% white, ages 30-82). Patients were not offered their incidental findings, but were asked to discuss anticipated benefits and harms of these results. Transcribed interviews were evaluated using a thematic content analysis approach consisting of coding and interpretation of the data by team consensus. Results: Most participants were interested in learning their incidental findings due to perceived benefits including the inherent value of information, and the ability to inform future disease prevention, detection, or treatment efforts; altruistically help others; provide health benefits to family; and explain their cancer diagnosis. Yet, some expressed disinterest due to concerns about the limited clinical utility of these results given their terminal disease status and/or age, and psychological effects. Many could not identify harms of learning incidental findings, whereas others noted adverse psychological effects and concerns about the complexity or uncertainty of information, privacy and insurance implications, and family communication challenges. A majority also believed that their immediate family would be interested in learning their incidental findings. Nearly all participants believed that doctors should offer to return actionable incidental findings, but most thought that patients should have a choice of whether or not to receive this information. Conclusion: Advanced cancer patients are interested in learning incidental findings, although concerns exist about the clinical and emotional implications of this information for patients and their families. Such concerns must be addressed to ensure that patients make informed decisions about learning incidental findings, and to derive maximal health benefits from this information.

2178F

Interest in and outcomes with return of individual genetic research results for inherited susceptibility to breast cancer. *A. R Bradbury[1], L. Patrick-Miller[2], B. E. Egleston[3], A. Brandt[1], J. Brower[1], L. DiGiovanni[1], J. M. Long[1], J. Powers[1], J. Stopfer[1], K. L. Nathanson[1], K. N. Maxwell[1], S. M. Domchek[1].* 1) Medicine, Division of Hematology-Oncology, University of Pennsylvania, Philadelphia, PA; 2) Center for Clinical Cancer Genetics. The University of Chicago Medical Center, Chicago, IL; 3) Biostatistics and Bioinformatics Facility, Fox Chase Cancer Center, Temple University Health System, Philadelphia, PA.

Background: Interest in, optimal communication models and patient reported outcomes (PROs) of returning individual genetic research results (IRR) are unknown. **Methods:** Breast cancer patients who provided a biosample for research were offered the opportunity to receive IRR for 25 cancer susceptibility genes. Participants completed pre- (V1) and post-test (V2) counseling and surveys evaluating PROs. We used linear regressions with estimation by GEE where appropriate. **Results:** Of 854 IRR available, 64 participants were deceased; 81 did not have contact information. 400 participants have been contacted. 90 (23%) consented to study, 9 (2%) declined, 107 (27%) did not respond to repeated contacts, and 40 (10%) are considering participation. 142 (36%) have not had a follow-up contact. To date, 72 have completed V1 and 65 (90%) chose to receive IRR (V2). 11 (17%) participants received a positive IRR including 3 deleterious/likely deleterious (D/LD) results in high-penetrance genes (*MSH2, MSH6, TP53*) and 8 in moderate-penetrance genes *(ATM, CHEK2, MYH Het, RAD51D*). 14 (22%) participants learned of a VUS. Confirmatory testing was recommended for all D/LD IRR and VUS in high-penetrance genes. To date, 13/14 (93%) were confirmed clinically; confirmatory testing was covered by insurance in all cases. Anxiety, depression, and uncertainty decreased after receipt of IRR (Table 1). Those with a positive result had greater decreases in uncertainty (p=0. 03 compared to negatives but not VUS, p=0. 20) and increases in satisfaction (p=0. 01 compared to negatives but not VUS, p=0. 07).

Table 1	Baseline**N=55^N=43^^	After V1**N=55^N=43^^	After V2**N=43^^
Anxiety (0-21)	6. 6 (3. 7)*6. 5 (3. 4)*	5. 7 (3. 7)*5. 6 (3. 6)*	4. 6 (3. 8)*
Depression (0-21)	2. 7 (2. 8)2. 5 (2. 9)*	2. 6 (3. 0)2. 4 (3. 0)*	1. 7 (2. 4)*
State Anxiety (20-80)	35. 6 (11. 5)35. 7 (10. 6)*	33. 4 (12. 2)33. 4 (11. 9)*	31. 6 (12. 3)*
Cancer Worry (0-75)	16. 4 (14. 7)16. 4 (14. 3)	15. 9 (14. 0)16. 1 (14. 4)	14. 6 (12. 7)
Knowledge (23-112)	93. 6 (77. 4)96. 7 (87. 4)	86. 3 (6. 4)85. 9 (6. 3)	85. 1 (6. 3)
Uncertainty (0-15)	6. 1 (3. 9)6. 4 (3. 4)*	5. 8 (4. 3)5. 6 (4. 4)*	4. 5 (3. 0)*
Perceived Utility (24-120)	71. 9 (19. 0)72. 6 (20. 7)	69. 5 (13. 2)70. 8 (13. 2)	67. 1 (12. 8)
*= p<0. 05. **Mean (SD). ^completed V1. ^^ completed V1 & V2.			

Conclusions: A minority of research participants are interested in receiving IRR. Most chose to receive IRR after genetic counseling and do not experience increases in distress or uncertainty. Confirmatory testing is important and to date has been covered by insurance.

2179T

Assessment of Current Practice and Athletic Trainers' Knowledge of the NCAA Sickle Cell Trait Screening Mandate. *M. Crandall[1], J. Scott[1], S. Dixon[1], R. Macatangay[1], M. Blitzer[1], J. Boughman[2].* 1) University of Maryland, School of Medicine, Baltimore, MD; 2) University System of Maryland, Alephi, MD.

Purpose: Keeping student athletes safe is an ongoing goal among all schools and sports, yet unfortunately, there are still exercise-induced sudden deaths occurring each year. One suspectedcause is sickle cell trait (SCT). SCT is a typically asymptomatic genetic condition yet evidence shows under extreme physiological conditions (including intense exercise or dehydration) SCT can cause significant problems and potentially sudden death. Following a lawsuit, the National Collegiate Athletic Association (NCAA) is now mandating universal SCT screening for all student athletes within their conferences but not much is yet understood about the implementation of the mandated screening. This study was designed to assess the current practice of this mandate as well as the overall awareness and comfort of athletic trainers working within this capacity, to identify inconsistencies where additional resources may be useful. **Methods**: An online survey was administered to all certified athletic trainers in the mid-Atlantic region, recruiting 330 responses, an 8% response rate. **Results**: Approximately 52% of trainers working in the collegiate setting reported athletic trainers are responsible for disclosing and discussing SCT screening results with student athletes. Trainers reported less comfort discussing these results than solely disclosing a positive or negative result. Comfort level was significantly correlated with type of previous experience or education in genetics. Trainers that reported formal experiences expressed greater comfort discussing SCT results (p=0. 006). Trainers also reported a wide variety of additional healthcare professionals that they would contact for information on this topic; collegiate trainers reported seeking a general practitioner most often. Most (71%) trainers reported their institution does not provide them access to healthcare professionals trained in genetics. Those provided such access, did not report consistent procedures for consultation. **Conclusion**: This study has identified significant variability in the current practice of implementation of the NCAA SCT screening mandate that may impact achievement of its intended goal. Additional resources should become more routinely accessible, and more involvement of genetics professionals, including genetic counselors, in the follow-up education and dissemination of this genetic material should be considered.

2180F

Is there an ethical and/or legal obligation for healthcare providers to re-contact former patients in light of new genetic findings? *D. Carrieri[1], S. Dheensa[2], A. M. Lucassen[2], P. D. Turnpenny[3], A. J. Clarke[4], S. E. Kelly[1].* 1) University of Exeter, UK; 2) University of Southampton, UK; 3) Royal Devon and Exeter Hospital, UK; 4) University of Cardiff, UK.

Advances in genetics and genomics have the potential to allow more accurate diagnoses, improve knowledge of disease aetiology and risk, and inform therapeutic decisions. For health benefits of genomics to be achieved, a number of issues in clinical implementation need to be addressed, including ethical and practice issues that have yet to be examined empirically and conceptually in any depth. Key among these is the clarification of professional responsibilities and obligations, as well as patient expectations, for recontacting patients in light of new genetic findings and the reinterpretations of earlier findings. These issues are becoming more urgent with the arrival of Genomics England and related endeavors. Our current three year (2014-2017) ESRC funded project 'Mainstreaming Genetics: Re-contacting patients in a dynamic healthcare environment' http://ex. ac. uk/mgc addresses these questions. In particular it aims to: 1. Survey current clinical practices regarding re-contacting in the NHS and other European healthcare systems 2. Analyse the ethical and legal issues surrounding a potential responsibility or obligation to re-contact patients 3. Investigate patient and healthcare professional perspectives concerning re-contacting in different medical specialties using semi-structured interviews, vignettes, and questionnaires 4. Engage with stakeholders to integrate the above findings and analyses in the drafting of professional guidance or to develop a professional framework for making decisions about re-contacting patients We will present and discuss the framing and initial findings of this project.

2181T

Reporting secondary whole genome sequencing variants in pediatrics: Challenging relational best interests. *R. Z. Hayeems[1,2,3], J. Anderson[4,5,7], MJ. Szego[6,7,12], MS. Meyn[3,8,9,10,11], C. Shuman[8. 9. 10], N. Monfared[3,10], S. Bowdin[3,9,10,11], R. Zlotnik Shaul[4,7,11].* 1) Child Health Evaluative Sciences, Hospital for Sick Children, Toronto, Ontario, Canada; 2) Institute of Health Policy Management and Evaluation, The University of Toronto, Toronto, ON, Canada; 3) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, ON, Canada; 4) Department of Bioethics, The Hospital for Sick Children, Toronto, ON, Canada; 5) Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON, Canada; 6) Centre for Clinical Ethics, St. Joseph's Health Centre, Toronto, ON, Canada; 7) Joint Centre for Bioethics, University of Toronto, Toronto, ON, Canada; 8) Genetics and Genome Biology, Hospital for Sick Children, Toronto, ON, Canada; 9) Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; 10) Division of Clinical and Metabolic Genetics, Dept of Paediatrics, The Hospital for Sick Children, Toronto, ON, Canada; 11) Department of Paediatrics, University of Toronto, Toronto, ON, Canada; 12) Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada.

Whole genome sequencing (WGS) in paediatrics promises to provide a transformative set of tools for gene discovery and clinical care, but evidentiary, ethical, and implementation challenges remain. While reporting adult onset medically actionable secondary variants (SVs) in children is the subject of vigorous ethical debate and emerging best practice guidelines, there is a paucity of research examining the perceived benefits and risks of WGS from the perspective of parents of children enrolled in WGS protocols. To this end, we explored parental experiences with WGS, as well as their preferences and reasoning related to pursuing or declining SVs for themselves and their children. The Hospital for Sick Children's Genome Clinic analyzes genomic data from children with developmental delay and/or congenital anomalies for both primary and secondary variants. A choice-based model, grounded in a relational best interests rationale, is used for managing the return of SVs to parents of index children. Parents of children enrolled in the Genome Clinic were invited to participate in semi-structured interviews following completion of an informed consent discussion and pre-test counseling. Interviews were transcribed verbatim and analyzed thematically, using a mixed coding strategy whereby pre-determined codes from the interview guide and emergent codes from the data were combined. We report the perspectives of 23 parents of 18 Genome Clinic participants. The majority of parents were supportive of WGS; they valued its role in providing diagnostic 'answers' to potentially guide treatment and family planning, and in advancing science. However, many parents expressed ambivalence about the receipt of SVs and importantly, some chose to learn SVs for their child but not for themselves. In part, ambivalence stemmed from the struggle to weigh and balance perceived benefits and risks to their child and to themselves. But ambivalence also stemmed from the perception that parents have a moral obligation to learn about their child's SVs regardless of any personal stress associated with this knowledge. Our findings align with previous reports indicating that parents endorse genome-based diagnostic testing. However, our findings expose a degree of moral pressure on parents to know what WGS makes knowable and, inidentifyingparents' interests inSVs fortheirchildren but not for themselves, challenge arelationalbest interests rationale for disclosing adult-onset SVs. .

2182F

My46 usability and preferences for results among African Americans. *S. M. Jamal[1], J. Yu[1], J. Crouch[2], A. Shankar[1], M. J. Bamshad[1,3], H. K. Tabor[1,2,3].* 1) UW Department of Pediatrics, Seattle, WA; 2) Seattle Children's Hospital and Research Institute, Seattle, WA; 3) UW Department of Genome Sciences, Seattle, WA.

The majority of translational research on return of genome sequencing results has focused on European Americans affected by rare genetic conditions. In this context, web-based digital tools have been developed and used to provide education during informed consent, solicit preferences, and return results. As genome sequencing is increasingly used in racially and ethnically diverse populations, testing novel approaches to soliciting preferences and returning results will be critical to ensuring parity in genomic healthcare and research. We conducted interviews with 24 African Americans (AA) to assess the usability of My46, a web-based tool, to solicit preferences for genome sequencing results. In pre-preference setting interviews, a third of participants expressed interest in learning ancestry inference results, followed by results for cancer, diabetes and sickle cell disease/trait often motivated by family history and interest in conditions that disproportionately impact AAs. When presented with options for results classified by trait categories, the majority (58%) sought all available results. Everyone wanted carrier results; 96% wanted results about medication response, genetic ancestry, and risk of heart disease; and 79% wanted results about digestive and muscle conditions. Half thought they might or would likely change their preferences for results return in the future, especially if their health changed. My46 usability scores (ranging 1-5) were consistently high: overall mean score was 4. 47; subscale scores for system use (4. 52), information quality (4. 44), and interface quality (4. 46). Almost all participants indicated high satisfaction using My46 to choose results, felt that My46 provided information to select preferences, and would recommend My46 to others. Despite high satisfaction, 25% would have rather chosen preferences with a person than a website. By comparing participants' initial expectations for results with their My46-solicited preferences, we highlight the importance of offering specific categories of results. These results also suggest that soliciting preferences may be a dynamic process rather than a one-time event, and that a fraction of research participants will prefer face-to-face interaction despite expressed satisfaction with a web-based tool. These findings support the use of web-based approaches to solicit preferences for sequencing results from AA research participants and characterize the usability of one such tool, My46. org.

2183T

Content and delivery preferences for genome sequencing results among young breast cancer patients. *K. Kaphingst[1,2], A. Elrick[1,2], J. Ivanovich[3], B. Biesecker[4], S. Lyons[3], M. Goodman[3].* 1) Huntsman Cancer Institute, University of Utah, Salt Lake City, UT; 2) Department of Communication, University of Utah, Salt Lake City, UT; 3) Division of Public Health Sciences, Washington University School of Medicine, St. Louis, MO; 4) Social and Behavioral Research Branch, National Human Genome Research Institute, Bethesda, MD.

Introduction. Genome sequencing raises the challenge of returning secondary findings. To inform policies for return of results, we examined what genome sequencing results young breast cancer patients would want returned, from whom, and when, since various results could be relevant at different life stages. **Methods.** 1080 women diagnosed with breast cancer age 40 or younger completed an online survey. We examined level of interest, preferred source, and preferred timing for return of seven types of possible genome sequencing results: risk of preventable/treatable disease or unpreventable/untreatable disease; treatment response; unknown meaning; health of children/other relatives; and non-health (ancestry/physical traits). We built seven multivariable logistic regression models to assess whether prior genetic testing, family history of breast cancer, BRCA mutation status, genome sequencing knowledge, and importance of health information were associated with preferred timing for return of each result type. **Results.** The proportion very interested in each result type ranged from 77% (risk of preventable/treatable disease) to 16% (unknown meaning). The majority of participants (64% to 81%) preferred that each type of result be returned by a genetic counselor or clinical geneticist, except for variants affecting treatment response for which the most common preference was an oncologist (37%). The proportion that wanted a result type returned before or at breast cancer diagnosis ranged from 96% (treatment response) to 31% (ancestry/physical traits). In multivariable models, no variable predicted preferred timing of return for all result types. Three factors were significantly associated with preference for results return before/at diagnosis in at least four models: knowledge of genome sequencing limitations (odds ratios [ORs]: 0. 89-0. 93); importance of health information (ORs: 1. 1-1. 5); and prior genetic testing (ORs: 0. 49-0. 65). **Discussion.** Participants were most interested in actionable findings delivered by a genetics professional. Preferred timing varied across types of results; participants with less knowledge of genome sequencing limitations, no prior genetic testing, and greater reported importance of health information were more likely to want return of results before or at diagnosis. A one-size-fits-all policy for return will likely not meet these patients' needs or preferences, and options to return results at different times may be warranted.

2184F

Genomic Sequencing of Newborns: Exploring Psychosocial Risks to Families. *S. Pereira[1], L. Frankel[1,2], J. O. Robinson[1], C. M. Weipert[3], M. C. Towne[4], A. Gutierrez[1], K. Lee[1], I. A. Holm[4,5], D. Dukhovny[6], A. H. Beggs[4,5], R. C. Green[3,7], A. L. McGuire[1].* 1) Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX; 2) Department of Psychological, Health, and Learning Sciences, University of Houston, Houston, TX; 3) Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; 4) Division of Genetics and Genomics, Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA; 5) Department of Pediatrics, Harvard Medical School, Boston, MA; 6) Department of Pediatrics, Oregon Health & Science University, Portland, OR; 7) Partners Center for Personalized Genetic Medicine, Boston, MA.

The integration of genomic sequencing into the clinical care of newborns has been met with both enthusiasm and trepidation. While genomic sequencing has the potential to provide clinical and personal utility to physicians and families, there are also concerns about the psychosocial risks of disclosing genomic information. Empirical studies exploring the psychosocial risks of returning genomic information to individuals have found no measureable changes to participants' psychological health, though these studies have primarily focused on symptoms of depression and anxiety in adults in response to receiving their own results. There is little data on how returning genomic information to parents about their children impacts family relationships. Furthermore, despite growing interest around exploring how genomic sequencing can be integrated into newborn screening, there is almost no data examining the impact on families of returning genomic information on newborns. Thus, research to fully understand the benefits and harms of genomic sequencing of newborns is vitally needed. The BabySeq Project is a randomized, controlled trial evaluating the benefits and potential harms of providing genomic sequencing to families of newborns and their physicians. Two cohorts of newborns and their parents are being enrolled: 240 newborns from the Well Baby Nursery at the Brigham and Women's Hospital in Boston, and 240 from the Neonatal Intensive Care Unit (NICU) at Boston Children's Hospital (480 total); their health care providers are also being enrolled. Newborns are randomized to receive standard newborn screening (NBS) only, or NBS plus genomic sequencing. Parents complete longitudinal surveys at 4 time points over the course of their baby's first year of life assessing psychosocial response, family impact, and attitudes toward and perceived utility of NBS and genomic sequencing. Impact on family relationships is assessed in three specific domains: perceived child vulnerability, parent-child bonding, and self and partner blame. Drawing from psychology and genetic counseling literature, this presentation will examine the relevance and implications of these family dynamic phenomena to the use of genomic sequencing in the clinical care of newborns.

2185T

"We all get dealt our cards": Participant Responses to Return of Secondary Results from Exome Sequencing. *H. K. Tabor[1,2,3], J. Crouch[2], S. Nelson[1], K. B. Shutske[2], S. M. Jamal[1], A. Shankar[1], J. Yu[1], M. J. Bamshad[1,3].* 1) Dept. of Pediatrics, University of Washington, Seattle, WA; 2) Treuman Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, WA; 3) Dept. of Genome Sciences, University of Washington, Seattle, WA.

Exome sequencing (ES) is being used increasingly in a broad range of research and clinical settings. Consequently, much debate continues about whether and how secondary results (SR) should be offered for return, which SR should be offered, and the potential benefits and harms of SR to both pediatric and adult patients. Yet, there are little empirical data about responses to and impact of SR return, and no empirical data about the impact of SR that may have clinical utility but may not be considered "actionable. " Studying such responses is important to guide policy development on return of SR. We recruited 131 people, who had previously undergone ES as a part of research studies, to select preferences for SR return. The study population included 52 parents receiving SR for their children and 79 adults receiving their own SR. Primary diagnoses of participants included: autism; brain, limb and craniofacial malformation disorders; cystic fibrosis; thoracic aortic aneurysm/dissection; and dilated cardiomyopathy. Participants were told that no results related to their primary condition would be identified or returned. On average, 4 results were returned (range of 0-7), and included results for risk of type 2 diabetes and age-related macular degeneration, a range of carrier status results, and medication response. We interviewed 48 participants one month after they received SRs and conducted qualitative content analysis of interview transcripts. We identified several key themes in responses to SR. First, most participants were not upset by SR and did not view them as upsetting or as "a crisis. " Second, they described the impact of SR being mitigated in context of other life events, such as health or personal issues. Third, they planned to apply health information management approaches that they used for their primary condition to their SR. Fourth, they planned to use SR to plan for the future, especially among parents of children with cognitive delays. Fifth, they planned to use SR as a motivator for health-promoting behavior. Sixth, they viewed SR as explaining family history or confirming diagnoses/risks they were already aware of. Seventh, they expressed disappointment that they had not received more genetic results and a desire for additional information. In conclusion, empirical data about SR-specific responses can provide important guidance for the offer, receipt and translation of SR as a part of clinical ES and its application to precision medicine.

2186F

Public Attitude Toward Disclosure of Incidental Findings in Genomics Research. *Z. Yamagata[1], I. Ishiyama[2], K. Muto[3], J. Minari[4], G. Yoshizawa[4], K. Kato[4].* 1) Dept Health Sciences, University of Yamanashi, Chuo, Yamanashi, Japan; 2) Tokoha University, Shizuoka, Japan; 3) The University of Tokyo, Tokyo, Japan; 4) Osaka University, Osaka, Japan.

[Objectives] The return of the findings to the participants, in particular incidental findings, is a controversial issue. This study aims to determine the public attitude toward the return of incidental findings obtained from genomics research. [Methods] From the general Japanese population, 2,400 people were selected using a random sampling and a mail survey administered in February 2014. The questions that were used in this study were as follows: (1) Do want to know the incidental findings?; (2) When do you want to declare your intention whether or not you want to know the incidental findings?; (3) What kind of incidental findings do you want to know?. A multiple logistic regression analysis to determine the factors related to whether they want to know the incidental findings was also conducted. [Results] The response rate was 56. 4%. The response that indicates they want to know incidental findings was 72. 0%; the response that indicates they do not want to know was 6. 4%; the response that indicates they were undecided was 21. 6%. There were no differences according to gender. For the timing of declaration of intention whether or not they want to know incidental findings, the response that indicates they want to declare it when they participated in a study was 22. 1%; the response that indicates they want to declare it when incidental findings are known was 38. 9%; the response that indicates they want to declare it both when they participated in a study and when incidental findings are known was 37. 4%. For the type of incidental findings they want to know, the response that indicates findings related to diseases of which prevention and treatment methods are available was 62. 2%; the response that indicates findings related to diseases of which prevention and treatment methods are unavailable was 44. 0%; the response that indicates findings related to effect of drugs and the presence or absence of side effects was 70. 0%. For the results of the multiple logistic regression analysis, the odds ratio of the response that indicates they want to know the findings was as follows: genomic literacy score, 1. 38 (95% confidence interval [CI] 1. 09 - 1. 74); age, 0. 96 (CI 0. 94 - 0. 98); no child, 0. 46 (CI 0. 23 - 0. 90); highly educated, 0. 88 (CI 0. 73 - 1. 07). [Conclusions] Of the study subjects, 70% responded that they wanted to know the findings. The response indicating they want to know the findings was associated with genomic literacy score, age and the presence of children.

2187T

Clinical Annotation of Sequence Variants in a Population-based Study of African Americans: Implications for report of incidental findings. *A. A. Adeyemo, F. Tekola-Ayele, A. Bentley, A. P. Doumatey, J. Zhou, G. Chen, D. Shriner, The. MH-GRID Consortium, CN. Rotimi.* NHGRI/CRGGH, National Institutes of Health, Bethesda, MD.

The widespread adoption of whole genome and whole exome sequencing (WES) in population genetic association studies has spurred great interest in how to handle incidental findings that could be of medical significance. In contrast to clinic based sequencing which is motivated by clinical indications, genetic association studies are implemented to identify variants that may not necessarily be the causal variant for the disease or trait. . We present our findings from WES of 555 unrelated African Americans from the Washington DC metropolitan area. The subjects were enrolled as part of the Howard University Family Study, a population based genetic epidemiology study with participants that were not ascertained for any specific phenotype. After appropriate technical filters, variants that had a call rate >= 95% (n=12,449) were annotated for clinical significance using the NCBI *ClinVar* database (20150504 version, n=111332 entries). Based on g the *ClinVar* classification of clinical significance, 471 (3. 8%) of the variants were classified as “pathogenic”, of which 56 were nonsense mutations leading to a stop codon. These represented a wide range of complex or monogenic disorders including: breast cancer susceptibility (e. g. *BRCA2*), drug toxicity (e. g. *TPMT* deficiency) and hematologic disorders (e. g. *G6PD* deficiency). Over 90% have entries in OMIM. While most of the disorders associated with the mutations are recessive or complex, eight of the mutations were for autosomal dominant disorders. There were 132 singleton and 1 doubleton individuals carrying pathogenic variants (in contrast to 1926 and 14, respectively for the other variants). Genes harboring these pathogenic mutations showed enrichment for the PANTHER pathways: blood coagulation (5. 87 X 10-6) and *TGF*-Beta Signaling pathway (8. 54X10-3). These findings illustrate how population based association studies have the potential to generate incidental findings of potential clinical significance. It also highlights the urgent need for sequencing studies to have in place at the onset a plan for handling incidental findings especially in study populations that lack adequate access to genetic counseling and medical genetics services.

2188F

Intentions to receive individual genetic research results and incidental findings from genomic studies among research participants in a familial cancer research program. *J. T. Loud, R. Bremer, P. Mai, M. H. Greene, B. P. Alter, S. A. Savage.* Clinical Genetics Branch, National Cancer Institute, Rockville, MD.

Purpose: To describe the frequency with which adult clinical research participants from high-risk families enrolled in a familial cancer research program give consent to receive their personal, clinically-validated research genetic test results (RR) and incidental findings (IF). **Methods:** Consents from 629 adult (≥18 years old) participants enrolled in three longitudinal studies within NCI's Clinical Genetics Branch's Family Cancer Research Program (Inherited Bone Marrow Failure Syndromes, Li Fraumeni Study and Familial Testicular Cancer Study) were obtained between January 2012 and June 2014. Genetically affected and unaffected individuals, including those with and without cancer were enrolled. Separate consents were obtained for the longitudinal study and for specific clinical genetic testing. A cross-sectional analysis was performed on the choices indicated on their most recent longitudinal study consent regarding receipt of RR and IF. **Results:** Nearly all research participants (97. 2%) indicated their preference to receive RR and IF. The participants who intended to decline either RR, IF, or both included: 3 cancer survivors (1=RR, 1=IF and 1=both), 9 who knew their primary mutation status (4=RR; 4=IF; 1= both), 3 were non-bloodline (1=RR; 2= both), 1 was untested but with the phenotype of interest (1=IF), and 2 were parents of an affected child (2=both). We speculate that the individuals who knew their mutation status (mutation positive=8; mutation negative=1) and declined RR, IF or both already had sufficient information or were not prepared to learn more. Both parents who refused both RR and IF had a child affected with a dominantly inherited syndrome, yet they chose not to learn which lineage harbored the mutation. The three cancer survivors were >50 years old, two were childless, and perhaps felt that the information wouldn't be useful for their personal healthcare decision-making. **Conclusions:** In this well-defined, highly-motivated population of individuals from families at high genetic risk of cancer, adult research participants overwhelmingly indicated their preference to receive both their RR and IF. Future research will seek to identify the underlying reasons for declining RR and IF in the small numbers of those who did so, and to study the impact of receipt of RR and IF on personal medical decision-making in individuals from families at high genetic risk of cancer.

2189T

Whole Genome Sequencing Analysis and Reporting for Ostensibly Healthy Individuals in a Research Study. *M. D. Linderman[1,2], G. A. Diaz[2], S. A. Suckiel[1,2], R. E. Zinberg[2], N. S. Abul-Husn[2,3], S. C. Sanderson[1], M. Wasserstein[2], E. Sarapata[4], A. Kasarskis[1,2], E. E. Schadt[1,2].* 1) Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Genetics and Genomics Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 3) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Harvey Mudd College, Claremont, CA.

Personal genome sequencing (PGS) is increasingly employed for pre-dispositional screening for disease risk in ostensibly healthy individuals. The shrinking cost of whole genome sequencing (WGS) makes it possible to apply this technology to increase both the scope and precision of pre-dispositional PGS. However, the increased scope and complexity of this data creates corresponding challenges for the implementing laboratory and reporting provider. As part of the HealthSeq project, a longitudinal cohort study in which unselected ostensibly healthy participants received a variety of genetic results from WGS performed on the Illumina HiSeq platform, we developed a comprehensive sequencing analysis and interpretation protocol and report. We concurrently reported both health and non-health-related genetic findings to 40 individuals (5 in a pilot , 35 in the HealthSeq study), including monogenic disease variants, polygenic disease risk, pharmacogenomic phenotypes, physical traits and ancestry. A mean (SD) of 48. 9 (15. 9) potential monogenic disease variants per participant were extracted for review by genetics professionals, with 1. 65 (1. 5) and 0. 63 (1) ultimately determined to be pathogenic and likely pathogenic respectively, according to the then draft ACMG variant interpretation guidelines. Predicted absolute genetically informed risk for Type 2 Diabetes, Coronary Artery Disease and Age-related Macular Degeneration were in the ranges [35%, 84%], [43%, 63%] and [9%, 72%], respectively. No correlation was observed between predicted risk and participant-reported family or health history. Metabolizer phenotypes for clopidogrel and simvastatin, genotypes used for warfarin dosing, ancestry information, and genotype-based predictions for lactose intolerance and other non-health-related traits were also reported. We expect the breadth of the pre-dispositional PGS result return to grow over time to satisfy recipient interests and maximize test utility. Our experience concurrently reporting a broad spectrum of health and non-health-related PGS results will provide valuable data to inform the future development of much needed standards and procedures for predispositional WGS.

2190F

What is the psychological impact of personal genome sequencing on ostensibly healthy individuals? The HealthSeq project at six-month follow-up. *S. C. Sanderson[1,2], M. D. Linderman[1], S. A. Suckiel[1], R. Zinberg[1], G. A. Diaz[1], M. Wasserstein[1], A. Kasarskis[1], E. E. Schadt[1].* 1) Icahn School of Medicine at Mount Sinai, New York, NY; 2) University College London, London, UK.

There is growing interest in the potential value of whole genome sequencing (WGS) for seemingly healthy individuals. WGS in this context could potentially lead to improved health and wellbeing, advances in disease prediction and prevention, and even diagnoses of previously unsuspected disorders. The HealthSeq project is a longitudinal cohort study in which ostensibly healthy individuals were offered personal WGS results. Qualitative and quantitative interviews were conducted at multiple timepoints including 1 week ("T3") and 6 months ("T4") after personal results were returned. Psychological measures included an adapted version of the Multidimensional Impact of Cancer Risk Assessment (MICRA) questionnaire. Participants were asked how often they had experienced a range of positive and negative feelings about their personal genome sequencing (PGS) results over the past week (never/rarely/sometimes/often). Twenty-nine (82. 9%) of the 35 participants completed the T3 and T4 follow-up interviews. Scores on the 6-item MICRA Distress Subscale (0=no distress, 40=high distress) decreased from T3 to T4, although most participants' scores were low at both timepoints (mean 1. 69 vs. 0. 48 respectively: Z=-2. 68, p=0. 007). At T3, most participants reported they had never "regretted" having their genomes sequenced (27/29, 93. 1%); most that they had sometimes or often felt "relieved" (27/29, 93. 1%) and "happy" (28/29, 96. 6%) about their results; and one fifth (6/29, 20. 7%) that they had sometimes or often felt "concerned about insurance. " One participant reported having sometimes regretted having PGS. In the qualitative portion of the interview, this participant discussed their feelings about their PGS results which included a rare 'pathogenic' Brugada syndrome variant, and increased risk of Alzheimer's disease (APOE e4/e4). At T4, the majority of participants again reported never having regretted having PGS (28/29, 96. 6%); two thirds reported having felt relieved (18/29, 62. 1%) and happy (20/29, 68. 9%) about their results; and a quarter (7/29, 24. 1%) that they had felt "concerned about insurance". The majority, but not all, of HealthSeq participants reported positive experiences, and little distress or regret, about having PGS. Insights from the HealthSeq project regarding the psychological impact of PGS on healthy people will be useful for clinicians, researchers and policy-makers concerned with maximizing the potential benefits and minimizing the potential harms of PGS.

2191T

Developing a provider facing genome sequencing report: Results of key informant interviews. *J. L. Williams[1], A. K. Rahm[1], A. L. Fan[1], H. Stuckey[2], D. T. Zallen[3], J. Green[1], M. Bonhag[1], L. Feldman[4], M. Segal[4], M. S. Williams[1].* 1) Genomic Medicine Institute, Geisinger Health System, Danville, PA; 2) Department of Medicine, Penn State Hershey College of Medicine, Hershey, PA; 3) Department of Science and Technology in Society, Virgina Tech, Blacksburg, VA; 4) SimulConsult, Inc. , Chestnut Hill, MA.

Background: Genome sequencing is rapidly entering into clinical practice often without adequate education for clinicians to understand implications of the results that are returned. Rapid detection of new genes and mechanisms leave even seasoned geneticists scrambling to synthesize the resulting information into meaningful communication for patients. We propose to develop genomic results reports with advanced functionality including point of care education and clinical decision support. Providers associated with a whole genome sequencing research study were invited to participate in semi-structured interviews aimed at soliciting their feedback on a draft genome results report. **Methods:** A draft provider report was developed by the research team based on the CDC recommended genetic test report format. In addition to clinicians, the team includes a patient investigator and several experts in patient engagement and communication. The draft report was then presented to participants in a clinical research project exploring the use of genome sequencing for undiagnosed cognitive disability and autism. Semi-structured interviews were used to elicit prior experience with genetic test result communication and feedback about the draft report. Interviews were transcribed and analyzed using interpretative phenomenological review. **Results:** Participants endorse the importance of having a report created for providers and a report designed for patients. Analysis of provider comments resulted in recognition of three constructs around communication of genome sequencing results. Universally, the providers found genetic results complex and expressed concern about their ability to adequately understand the genomic finding and implications. Second, providers expressed relief at having diagnosis-specific information readily available within the context of the patient visit that guided clinical care for the patient. Finally, providers accepted and appreciated a resource that could support communication of genome results with families. **Conclusion:** Participants value a report that will offer explanation of genome findings in a way that increases communication of the result with their patients. The final genome results reports will be used in the comparative effectiveness portion of this PCORI project.

2192F

Parental Views of Newborn Screening Using Next Generation Sequencing Technologies: Implications for Policy. *G. Joseph[1], F. H. Chen[2, 3], J. Harris-Wai[3, 4], J. M. Puck[3, 5], C. F. Young[6], B. A. Koenig[2, 3].* 1) Anthropology, UCSF, San Francisco, CA; 2) Institute for Human Genetics, UCSF, San Francisco, CA; 3) Institute for Health and Aging, UCSF, San Francisco, CA; 4) Division of Research, Kaiser Permanente, Oakland, CA; 5) Pediatrics, UCSF, San Francisco, CA; 6) Genetics, CSU Stanislaus, Turlock, CA.

Parental views about the translation of next-generation sequencing (NGS) into the public health newborn screening (NBS) setting are not well elucidated. One goal of NSIGHT, a NICHD/NHGRI funded program exploring the use of NGS in newborns, is to examine stakeholder views, perspectives and values about the possible future expansion of NBS to include NGS. To address this goal, we conducted five 2. 5-hour-long focus groups between November 2014 and February 2015 with: parents of children with primary immune disorders recruited from an academic immunodeficiency clinic (1 group; n=5); and women 18-30 weeks pregnant recruited from prenatal clinics at two urban California hospitals, an academic medical center and a public hospital (3 English; n=22 and 1 Spanish; n=5). Participants represented a socio-economically and ethnically diverse sample. To elicit preferences about the incorporation of NGS into state-mandated public health screening, we provided information about case examples of pharmacogenomics (PGx) and adult onset disorders such as hereditary breast cancer, and explored issues of parental consent, return of results, and data privacy. Focus group recordings were fully transcribed in English. Transcripts were coded by multiple team members and analyzed using qualitative methods: thematic analysis of coded segments of text and constant comparison across demographic and disease categories. Several policy-relevant findings emerged: 1) a clear preference for more comprehensive NBS education in the prenatal period; 2) mixed views on whether a formal informed consent process was needed for current NBS or NBS expanded by NGS, but with expressed desire to be informed of downstream research using blood spots as a matter of respect and trust; 3) a preference for return of PGx results but no consensus on return of variants for adult onset disorders; and 4) expressions of both hope and concern about the medical system's capacity to safely store and make available NGS results when they are needed across the lifespan. Taken together, these results indicate that parents and prospective parents desire greater inclusion in the NBS process, suggesting the need for iterative review and revision of existing policies as technology evolves.

2193T

Ethical, legal and social implications of the application of genomics to infectious disease prevention, treatment and outbreak control. *R. Dvoskin[1], T. Bailey[1], P. Duggal[2], M. H. Lewis[1], D. Salmon[2], A. Sutherland[2], C. L. Thio[2,3], J. P. Kahn[1], G. Geller[1,2,3].* 1) Johns Hopkins Berman Institute of Bioethics, Baltimore, MD; 2) Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; 3) Johns Hopkins Medicine, Baltimore, MD.

Advances in genomics are contributing to the development of more effective, personalized approaches to the prevention and treatment of infectious diseases. Sequencing technologies are furthering our understanding of how genomic signatures in humans, pathogens and vectors—and their interactions—contribute to individual differences in immunologic responses to infections, vaccines, and drug therapies. With the potential for tailored interventions for individuals or subpopulations, ethical, legal and social implications may arise for clinical and public health practice. We conducted 12 key informant interviews with international experts in genetic epidemiology, immunology, infectious disease surveillance, vaccine development, public health ethics, law and policy. We also conducted four two-hour focus groups on this topic, two with a diversity of Johns Hopkins faculty in relevant fields, and two with graduate students in the schools of medicine and public health. Themes identified included: host genomics will be important for developing better and safer treatments and vaccines; among the most important infectious disease applications of genomics are immunogenomics, vaccinomics, and pathogen genomics; and genomic information is already used in clinical decisions for infectious disease—without much regard for downstream (ethical) implications. We will present examples of genomic discoveries with current or potential implications for infectious disease management along with ELSI issues that could arise as a result. For example, given known associations between human gene variants and susceptibility to influenza infection or response to vaccination, might individuals' genetic information be used to influence workforce restrictions on healthcare personnel during a flu pandemic? Or, given associations between gene variants and non-response to hepatitis B vaccine, might genotype inform decisions about prioritizing access to therapy, particularly in resource-limited settings?We will highlight the importance of anticipating such ELSI issues in advance of new scientific discoveries. Additionally, we will ask attendees how they see genomics figuring in to infectious disease research and management in the next five to ten years; what ethical, legal, or societal concerns they might foresee; and how they would want these issues to be addressed and analyzed.

2194F

RWANDAN PATIENTS WITH DISORDERS OF SEX DEVELOPMENT: DIAGNOSTIC CHALLENGES AND BEHAVIOR PROBLEMS IN AFRICAN CONTEXT. *L. Mutesa[1], J. Hitayezu[1], A. Uwineza[1,2], S. Murorunkwere[1], J. Ndinkabandi[1], A. Ndatinya[3], F. Rutagarama[3], O. R. Karangwa[3], A. Gasana[4], J. H. Caberg[2].* 1) Center for Medical Genetics, University of Rwanda, KIGALI, Kigali, Rwanda; 2) Center for Human Genetics, University of Liege, Belgium; 3) Department of Pediatrics, Rwanda Military Hospital, Rwanda; 4) Department of Urology, Rwanda Military Hospital, Rwanda.

BackgroundDisorders of sex development (DSD) comprise a variety of congenital diseases with anomalies of the sex chromosome, the gonads, the reproductive ducts and genitalia. The most common DSDs result from disruption of androgen levels and activity that affect later embryonal development, such as congenital adrenal hyperplasia and androgen insensitivity syndrome (AIS). DSDs are always challenging and very difficult to manage. Socio-economic and cultural aspects have a great impact on decision making regarding the management of these conditions. The situation is more complicated in resource-poor settings like in Africa, where access to education and medical care is limited in both quantity and quality of infrastructure, diagnostic tools and medical professionals. In addition, traditional values and beliefs are also very strong in various cultures and sexual issues are taboo subjects. MethodsThe present study is a 5-year prospective descriptive cohort of patients with suspicion of DSDs referred to our genetic clinic between January 2010 and December 2014 for genetic investigations and counseling. All patients underwent abdominal ultrasound or MRI and hormonal analysis before genetic testing including karyotype and molecular tests. ResultsIn total,A series of 49 patients aged between 1 and 39 years were clinicaly and geneticaly diagnosed with DSDs associated with behavior problems in most of cases. The majority were diagnosed either with sexual ambiguity and hypospadias, or micropenis, or primary amenorrhea, or poor development of secondary sexual characteristics, or primary infertility. In most of female cases, the ultrasound and MRI revealed absence of uterus and ovaries. The FSH, LH or testosterone hormons revealed major abnormal values in more than 60% of patients. The AIS and Rokitansky syndrome were observed in the majority of these patients. The choice of gender identity after karyotyping raised several psychological and ethical issues in the majority of adults patients. The outcome of surgery was successful for the social integration of some of these patients. ConclusionThe present study showed that patients with DSDs have major behavior problems in African context. They suffer from extreme anxiety and very high psychological behavior disorders related to their gender identity. Our data suggest that gender assignment has to be avoided before expert and multidisciplinary evaluation especially in young patients.

2195T

Incidental pathology findings in the Genotype Tissue Expression (GTEx) program. *L. Carithers[1], S. Koester[2], R. Little[2], N. Lockhart[3], A. Rao[1], J. Struewing[3], S. Volpi[3], C. Weil[1], H. Moore[1].* 1) National Cancer Institute, Rockville, MD; 2) National Institute of Mental Health, Rockville, MD; 3) National Human Genome Research Institute, Rockville, MD.

GTEx is a National Institutes of Health (NIH) program that collects tissues from over 30 different tissues from hundreds of deceased donors and performs molecular analysis to evaluate the relationship between genetic variation and gene expression in normal human tissues. These deceased donors may have also donated organs and tissues for transplant purposes. Pathology review is conducted on all GTEx tissues to ensure that the correct target is sampled and to evaluate autolysis, however there have been rare observations which may have potential medical significance to either organ transplant recipients or to recipients of banked tissues (skin, bone, tendon, etc). The GTEx Executive Committee, made up of lead Program Directors from the National Human Genome Research Institute, the National Institute of Mental Health, and the National Cancer Institute, requested consultations with representatives of the NIH Department of Bioethics, relevant clinical experts and policy experts from the Organ Procurement and Transplantation Network (OPTN) to determine if an ethical obligation exists to return these incidental findings to the biospecimen source sites in order to allow notification using the proper medical and regulatory channels. They also examined the literature related to the risk of disease transmission to organ and transplant recipients. It was determined that there may be an ethical obligation to inform biospecimen source sites when incidental pathology findings could have potential medical significance to transplant recipients. A formal group of experts, the GTEx Incidental Findings Evaluation Group, composed of experts in oncology, pathology, and bioethics was established to evaluate incidental findings identified by GTEx pathologists. When the Evaluation Group determines that a pathology incidental finding may have potential medical significance to a transplant recipient, a pathology evaluation is requested from a CLIA certified pathology laboratory. The Evaluation Group then reviews the surgical pathology report from the CLIA lab and makes a recommendation to return or not return a finding to the corresponding biospecimen source site. Over 850 GTEx cases have been reviewed and only a select few cases have had incidental findings that were determined to be of potential medical significance to transplant recipients. Those findings have been returned to the biospecimen source sites for transmission to treating physicians and tissue banks.

2196F

Next-generation counseling: increased efficiency and high patient satisfaction utilizing web technology and telephone for post-test counseling and education in carrier and inherited cancer screening.
M. Gilats, G. A. Lazarin, K. Ready, E. W. Denne, I. S. Haque. Counsyl, South San Francisco, CA.

Post-test patient followup is integral to genetic screening programs. In cancer screening, an increasing number of patients is being tested outside traditional genetics clinics; in reproductive genetics,increasing use of expanded carrier screening (ECS) and its high test-positive rate also poses a challenge. The efficient allocation of genetic counselors' (GCs) time is a necessary step to large-scale implementation of genetic testing in these novel contexts.

To address this need, our laboratory has offered three models of an online results delivery system with access to post-test telephone genetic counseling. Here we compare the models' effects on GCs' time spent: V1 (every patient speaks to a GC), V2 (results are systematically triaged, and most patients are given results online, with web education and immediate GC consultation option), and a "traditional" model (the clinic delivers results and GC consultation is available when initiated by the patient).

We analyzed consultation uptake, duration, and satisfaction for 410,668 carrier screening results and 1,232 inherited cancer screening results issued by our laboratory, identifying the following trends:

1. In all versions, more serious results were associated with longer consultation times.
2. In V2 and the traditional model, more serious results were also associated with higher consultation uptake rates.
3. Consults were longer on average for V1 results than for V2. E. g. , an individual carrier consult was 12 min (9 - 17 min, interquartile range) for V1 and 10 min (7 - 15 min) for V2. The shorter consult times may derive from pre-consult web education issued through V2.
4. Illustrating optimal GC allocation, the V2 model increased consultation rates up to 5. 7x versus the traditional model for positive inherited cancer screen results and the V2 website education provided for negative results enabled lower consultation rates compared with V1.
5. Patient satisfaction was high (4. 9 / 5. 0 rating) in all models.

We demonstrate time utilization benefits by a results-delivery model that combines web education and counseling. Time saved by efficient delivery of simpler results offsets time used due to increased positives. In one example, 2 min per consult were saved. As a laboratory issuing multiple thousand results per week, with concomitant GC availability, this translates to hours saved enabling feasibility of post-test follow up for widespread gene-panel screening in both the reproductive and oncology contexts.

2197F

Design, implementation and outcomes of a "Psychiatric Genetics for Genetic Counsellors (PG4GC)" workshop in the UK and creation of the International Society for Psychiatric Genetic Counselling (ISPGC).
K. McGhee[1], A. Inglis[2], J. C. Austin[1,2]. 1) Translational Genetics Group, Dept. of Life & Environmental Sciences, Bournemouth University, Poole, Dorset, United Kingdom; 2) University of British Columbia, Departments of Psychiatry and Medical Genetics & British Columbia Mental Health and Addictions Research Institute, Vancouver, BC, Canada.

Knowledge about the roles of genetic and non-genetic contributors to the development of psychiatric disorders is rapidly accumulating. Concurrently, the importance and urgency of translating this knowledge into better health provision is increasing. The first specialist psychiatric genetic counselling (PGC) service of its kind opened in Vancouver in 2012, generating interest from healthcare professionals from around the world who wished to develop similar services. In responding to this interest, we noticed that for many of these individuals, the barriers to their provision of high quality specialist services for individuals with psychiatric disorders and their families related to four core issues: lack of familiarity with specialist (psychiatric and/or genetic) terminology, feeling out of touch with the state of the art research used in psychiatric genetics, lack of confidence in discussing psychiatric disorders – both in general and their aetiology (in lay language) specifically, and uncertainty regarding how to discuss the risks for illness recurrence amongst family members of affected individuals. Therefore, with a target audience of genetic counsellors in mind, we developed an intensive two-day workshop (PG4GC) that aimed to address these four core issues and form the foundation for the development of a strategy for delivering PGC both within the UK and globally. The workshop, which was delivered in the UK in Feb 2015, incorporated didactic content, small and large group discussion, a problem based learning case, and individual reflective work. In total, 23 participants from seven countries attended and completed pre- and post- workshop surveys, which revealed that participants were more confident with respect to all four core issues after the workshop. At the conclusion of the workshop, participants founded the International Society for Psychiatric Genetic Counselling (ISPGC, which interested individuals are welcome to join). Members of this fledgling organization met at the 2015 World Congress of Psychiatric Genetics and will facilitate an ASHG workshop based upon our earlier model. Plans are underway for a second PG4GC workshop in 2016 and for expansion of the ISPGC.

2198F

The Individualized Medical Genetics Center: Facilitating Systematic Integration of Genetic Testing into Patient Care. *L. Medne[1], E. Bedoukian[1], A. Wilkens[1], M. Deardorff[1,2], I. Krantz[1,2].* 1) Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

The implementation of genetic testing is becoming progressively more complex yet it is being used in individual patient care at an increasingly rapid rate. Challenges encountered by those ordering genetic testing include: comprehension of diagnostic options, selection of the appropriate genetic test, authorization by insurance, communication with patients and families, promotion of patient autonomy, interpretation of genomic variants, assessment of variants with diagnostic utility and their potential phenotypic correlation, as well as many others. There is thus an overwhelming need for collaborative integration of genetics professionals within any medical specialty that orders complex genetic testing. This recognition led to the development of the Individualized Medical Genetics Center (IMGC) at The Children's Hospital of Philadelphia (CHOP). The IMGC is currently composed of two medical geneticists and three genetic counselors whose mission is to facilitate access to state-of-the-art appropriate genetic testing and targeted medical management for children, families, and clinicians across CHOP. Since July 2014, the IMGC has been involved with over 450 patients. IMGC team has done 56 inpatient consults and 313 outpatient evaluations. Referrals to the IMGC have included requests for test selection, obtaining insurance authorizations, evaluation with pre- and post-test counseling. Specialties utilizing IMGC services have included: rheumatology, endocrinology, immunology, neurology, and hematology. Perhaps surprisingly, geneticists have also referred patients to the IMGC for the primary indication of insufficient time in clinic to facilitate complex testing like exome sequencing. IMGC facilitated tests have included single gene testing, panel testing, targeted familial testing, and exome sequencing. There has been positive feedback from both families and clinicians. IMGC has tracked several metrics: insurance approvals and denials, confirmed molecular diagnoses, variants of uncertain significance, secondary findings, and additional testing needed for result interpretation. The IMGC team provides clinical expertise towards exome analysis offered by CHOP's Division of Genomic Diagnostics. The IMGC is an unprecedented and valuable initiative that facilitates complex genetic care for patients across the hospital. The success of its unique model for the facilitation of genetic testing may benefit similar endeavors at other hospitals in the US and abroad.

2199F

Examining Preparedness Among Hispanic Women Undergoing Genetic Cancer Risk Assessment. *B. Nehoray[1], C. Ricker[2], K. Yang[1], J. Weitzel[1], K. Blazer[1].* 1) City of Hope, Duarte, CA; 2) USC Norris Comprehensive Cancer Center, Los Angeles, CA.

Background: Despite recognition of Genetic Cancer Risk Assessment (GCRA) as a standard-of-care service, there are significant disparities in access to GCRA among at-risk underserved Hispanics. This study examines the preparedness of patients participating in a randomized controlled study examining the effect of a culturally-appropriate pre-GCRA telephone intervention on attendance, preparedness, knowledge, and satisfaction with *pro-bono* GCRA services. **Methods**: Participant eligibility was adult Hispanic women who met NCCN criteria for *BRCA* referral. Participants were randomized to a motivational interview intervention (MI), time and attention control (TAC), or usual care (UC). A 10-item preparedness survey was administered verbally immediately pre-GCRA. Items included understanding about GCRA, information sought about cancer genetics, communication with and engagement of family. **Results:** Of 281 participants, 88, 105, and 88 were in the MI, TAC, and UC arms, respectively. Most (77%) had a history of breast or ovarian cancer. Most understood the purpose of GCRA, describing it as: to learn if cancer in the family is hereditary (43%), for gene testing (23. 5%), and to learn about family risk (17%). Over half (52%) came with questions, including risk to family members (13. 5%), cancer etiology (8. 5%), and how genetic testing is performed (4. 5%). Of 19% who sought information before GCRA, the Internet was the most frequently cited source (9%); <2% sought information from a clinician. Overall, 46% were accompanied to the appointment; stated reasons were for emotional support (27%), to understand information (22%), or for transportation (12%). Participants were often accompanied by a partner (17%), daughter (12%), or sister (5%). Fifty-two percent did not contact family members for history, 15% stating that they already had family history information. Participants often contacted female relatives: mothers (20%), sisters (17%), and aunts (7. 5%). **Conclusions:** No significant differences in responses were noted between study arms; however, 90% knew the purpose of GCRA. This contrasts with the literature, which reports limited understanding of cancer genetics among Hispanics. Moreover, many came to the session accompanied by a family member, and with questions about GCRA. Further analysis of the other components of the study will better elucidate the potential value of culturally-appropriate pre-GCRA communication to enhance uptake and experiences with GCRA.

2200F

Exploring the process of decision-making about participation in genetic research on mental illness. *J. Austin[1,2], H. Andrighetti[1], A. Semaka[1].* 1) Dept of Psychiatry, UBC, Vancouver, BC, Canada; 2) Dept of Medical Genetics, UBC, Vancouver, BC, Canada.

Introduction: There is a range of barriers to recruitment for research on mental illness (MI), including distrust of researchers and social stigma. Among individuals who do participate in MI research, little is known about how and why they decide to participate. This study explored the process of decision-making around participation in genetic research on MI, including motivating factors and perceived benefits of participation, and expectations regarding return of genetic research results. **Methods:** This qualitative study utilized grounded theory methodology. Open-ended interviews were conducted with 16 individuals who had either completed participation or had recently made a decision about participation in a genetic research study on MI, led by genetic counselors. Interviews were transcribed and analyzed using the constant comparative method and open, axial, and theoretical coding procedures. **Results:** Illness acceptance and establishment of trust with the research team and institution were foundational elements required for individuals to consider participating in genetic research on MI. Main motivators for participation included perceived personal relevance, anticipated benefits, a desire to "give back", and accessible study procedures. Perceived benefits of research participation included access to support and resources via the research team, the opportunity to learn, and improved self-worth. Return of personal genetic research results did not appear to be a major factor in the decision-making process regarding participation. **Conclusion**: Our data suggest that participation in genetic research on MI helps make meaning of individuals' illness experience and empowers them to adopt positive health management strategies. Our results support the value of genetic counselors in research. Genetic counselors possess a unique skill set that enables them to build trusting relationships that facilitate recruitment and retention of participants. These findings may inform strategies that improve participation rates, decrease attrition, and maximize participant benefits.

2201F

Genetic counselling in a wound care unit: Unexpected contribution to conventional diagnosis strategies by drawing a pedigree to diagnose hereditary sensory and autonomic neuropathy type 2 (HSAN2) in female patient. *M. O Cevik[1], F. DOGAN[2].* 1) MEDICAL GENETICS, ADIYAMAN UNIVERSITY SCHOOL OF MEDICINE, ADIYAMAN, TURKEY; 2) DEPARTMENT OF PLASTIC and RECONSTRUCTIVE SURGERY ADIYAMAN UNIVERSITY SCHOOL OF MEDICINE, ADIYAMAN, TURKEY.

Here,we report a 30 year old woman who had never obtained a genetic counselling service in her previous life and whose disease could not be diagnosed with classical analytical approach previously for 20 years. Our collaboration enabled clinicians to recognize the hereditary nature of the disease and changed the course of therapy. According to her previous medical documentation available, patient's very first symptoms had started as early as she was 6 years old. She had started feeling tingling and nubness on her both feet the gradual loss of pain feeling. Then, in time, she started having skin injuries-ulcerations that would never heal again and frequently got infected. At the age of ten, to stop these ulcerations, amputation had been applied to her left big toe. Amputation almost stimulated a rapid osteolysis response in her left foot to proximale to the knee and she lost her left tibia and fibulae. Soon after amputation, right big toe had also started ulcerating, then came another amputation. However, starting from the age of 23, she had started experiencing dysplasia in her fingers and hand as well. As the years adveanced, her bones became so weak that she had to had supportive operation that bone allografts were added to support tibiae. She also had two two operations due to pelvis fractures. Now, she can only walk with the help of ortheses. Her big toes had been amputated. She has a deepening overall osteopenia thus increased fracture risk (DEXA -2,2 Z , 25% total bone mineral density) in all of her body. Shehad syndromic axonal polyneuropathy- gangliopathy at her both feet, osteopenia-osteoporosis osteolysis and osteonecrosis in the MR findings at her both femurs, adenomatosis liquids at her tibiae. In addition, she had recurrent osteomyelitis at her feet and hands. Neuronal desensitiziation was obvious that her wounds and bone fractures were totally painless. As we drew her pedigree, we realized that among her first degree cousins and secondary degree cousins whose ages vary between 9 and 16, there were very similar symptoms. Gathering these information, we have utilized Simulconsult®- and pedigree to reach a conclusion. As a result, we concluded that she had HSAN2 disease that went undiagnosed for 25 years. An unexpected contribution from us to wound healing center raised great awareness among other medical professionals.

2202F

An automated drawing tool of pedigree, intending to use for genome-cohort studies. *A. Shimizu[1], K. Yamamoto[1,2], T. Tokutomi[2], K. Sobue[3,4], A. Fukushima[2,5].* 1) Division of Biomedical Information Analysis, Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Shiwa-gun, Iwate, Japan; 2) Department of Clinical Genetics, School of Medicine, Iwate Medical University, Shiwa-gun, Iwate, Japan; 3) Executive Director, Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Shiwa-gun, Iwate, Japan; 4) Department of Neuroscience, Institute for Biomedical Sciences, Iwate Medical University, Shiwa-gun, Iwate, Japan; 5) Division of Innovation and Education, Iwate Tohoku Medical Megabank Organization, Iwate Medical University, Shiwa-gun, Iwate, Japan.

ObjectiveIn clinical genetics, genealogical information is extremely important for obtaining accurate genetic diagnoses. However, the creation of pedigree charts requires time and labor as they are typically created manually through interviews held on an individual basis and require specialized knowledge of clinical genetics, in addition to chart-creation skills. Although numerous large-scale studies, such as genome-cohort studies, have been conducted in recent years, accurate genealogical information remains necessary for the analysis of polymorphisms and diseases in the future. However, with the traditional method of pedigree chart creation, it is not possible to create pedigree charts using large amounts of genetic information obtained through interviews at a single recruiting event. We therefore developed novel software named “f-tree” that can support a user to collect genealogical information accurately and create pedigree charts, and has a variety of uses from large-scale genome-cohort studies to everyday medical care. Methodsf-tree creates pedigree charts based on a multiple-choice questionnaire regarding genealogical information. Pedigree charts are created in the following manner. 1) A multiple-choice questionnaire on genetic information is filled out by the participant or by an interviewer. 2) A person responsible for data entry inputs information from the questionnaires into f-tree. 3) f-tree automatically creates a pedigree chart. The process of chart creation is displayed on the f-tree screen in real-time, which allows for inspection during the creation process. 4) The created pedigree charts can be stored on electronic media or be exported into databases. Summary-When considering the future health of citizens, it is necessary for the general public to understand the principles of genetics. For this reason, the importance of basic genetics education has been pointed out. In such education, a proper understanding of pedigree charts is essential. f-tree is a new type tool that automatically create pedigree charts which enables even a less-experienced doctor and medical practitioner. We will soon release f-tree on our website (http://iwate-megabank. org/en/genetic/) as a freeware program and a free mobile app for the general public, so that it can be popularized and utilized not only in clinical and medical research, but also in educational settings.

2203F

The unintended birth of a second child with a genetic disorder in the same family – potential culprits and solutions. *M. Bembea[1,2], K. Kozma[1,2], C. M. Jurca[1,2].* 1) Genetics Department, Municipal Clinical Hospital dr. Gavril Curteanu Oradea, Oradea, Bihor, Romania; 2) Faculty of Medicine and Pharmacy Oradea.

BACKGROUND: Typically, in a family, only one child is affected by an unanticipated genetic disorder. After the first affected child, genetic counseling usually helps families to make decisions regarding a next pregnancy. In our setting, the unintended birth of a second child affected with the same condition is not uncommon. We hypothesize that the quality and certainty of the diagnosis of the first affected child, and the presence and effectiveness of genetic counseling, could be potential contributors to what the parents self-describe as the unintended birth of a second affected child. OBJECTIVES: To determine potential reasons for the unintended birth of a second child with the same genetic disorder. PATIENTS AND METHODS: This is a retrospective cohort study of 4,177 patients with genetic disorders in the evidence of a single county’s centralized genetics service, from 1984-2014. RESULTS: There were 148 (3. 5%) patients from 59 families who had at least one sibling affected with the same genetic disorder. Diagnosis of the first affected child in the family was full and clear for 19/59 (32%) families, partial or incomplete in 23/59 (39%) families and was not yet made for 17/59 (29%) families. Genetic counseling after the birth of the first affected child was done for 28/59 (47%) families. Of these, the staff ascertained good comprehension in only 8/28 (29%) families and poor comprehension in 20/28 (71%) families. Thirty-three families (56%) had only affected children (i. e. , one or all children affected). Thirteen families (22%) had a normal child born before the affected one, and other 13 families (22%) had a normal child born after the birth of an affected child. The mean time period between the first affected and the next affected child was 4. 8 years (range, 1 year - 26 years). DISCUSSION: We discuss possible explanations for these results and potential solutions to avoid this undesirable situation. CONCLUSIONS: The birth of a second affected child in a family poses special challenges to providers (to improve the process for timely diagnosis and for genetic counseling), to families (who require psychological and educational support) and to the community (social and economic issues). KEY WORDS: genetic counseling, affected siblings,.

2204F

"It's been a rollercoaster": Evaluating prospective father's experiences with abnormal prenatal microarray results. *L. Pilchman[1], K. J. Stumm[1], L. Conway[1], A. Werner-Lin[3], F. Barg[3], K. Kellom[4], B. A. Bernhardt[2].* 1) Department of Genetic Counseling , Arcadia University , Philadelphia, PA; 2) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) School of Social Policy and Practice, University of Pennsylvania, Philadelphia, PA; 4) PolicyLab, Children's Hospital of Philadelphia, PA.

Background: Chromosome Microarray (CMA) has become a first tier diagnostic test in pediatrics, and is rapidly being introduced in the prenatal setting. However, interrogation of the entire genome leads to the discovery of copy number variants of unknown or variable clinical significance thereby complicating pregnancy decisions. The uncertainty associated with pathogenic or uncertain variants causes considerable anxiety for pregnant women undergoing testing, but little is known about the experiences of their male partners. The purpose of this study was to evaluate how prospective fathers experience abnormal prenatal microarray results. Methods: Twelve men completed semi-structured interviews addressing their experiences of receiving abnormal prenatal CMA results. Interviews were digitally recorded and transcribed verbatim. The transcripts were imported into NVivo 10 for coding and analyzed using directed content analysis. Results: Many men followed their wives' lead in deciding to undergo CMA with the intention of being supportive, respectful of her autonomy, and because there was no increased risk of miscarriage or cost. Men collaborated with their wives in the decision of whether or not to continue the pregnancy. To facilitate decision-making, partners relied on rational thought processes (as opposed to emotional or moral reasoning) considering factors such as the probability of clinical involvement, the fetus' expected quality of life and an anticipated parenting experience. Most did not anticipate the possibility or understand the implications of a variant result, and many expressed worry about the fetus' future. Men varied on whether they sought additional information elsewhere; many trusted their wives' robust research efforts, and to cope, a few sought information online. The impact of uncertain results combined with the excitement of becoming a father caused many to feel like they were on an "emotional rollercoaster. " Participants made recommendations for improving the CMA process, including the provision of educational media outlining the purpose, risks, benefits, and possible results of CMA. Discussion: These findings indicate that men take on varying roles in prenatal testing and care during pregnancy. Additional recommendations for counseling include encouraging partner involvement in prenatal care and educating couples to the possibility of identifying variants of uncertain or unknown significance.

2205F

Individual perceived health status and health values: Factors influencing self-initiated genome testing, risk perception and health actions in ill versus healthy persons. *C. Farrell[1], K. Eng[2], B. Holaday[1], J. Evans[3], R. Green[4], S. Roberts[5] for the PGen Study Group.* 1) Healthcare Genetics, Clemson University, Clemson, SC; 2) Roswell Park Cancer Institute, Buffalo, NY; 3) Department of Genetics, University of North Carolina School of Medicine, Chapel Hill, NC; 4) Harvard Medical School, Boston MA; 5) University of Michigan, School of Public Health, Ann Arbor, MI.

Purpose: Factors influencing self-initiated genome testing are unclear. This study aims to explore how the interplay between individuals' perceived health status and their actual medical diagnoses are related to the appraisal of, and response to, direct-to-consumer (DTC) genetic testing. **Methods**: 1464 DTC genome testing users who participated in the 2012 Impact of Personal Genomics (PGen) study were classified into 4 mutually exclusive health status groups: 1) Diagnosed ill [DI] participants reporting formal medical diagnoses and poor to fair health; 2) Medically managed [MM] had medical diagnoses but reported their health as good to excellent; Health in question [HQ] had no medical diagnoses but reported health as poor to fair; and Healthy [H]) had no medical diagnoses and good to excellent self-reported health. Analyses (Chi-square, t-tests, ANOVA and correlations) assessed differences across these groups in terms of perceived risk (of disease), motivations for testing, and values and interests specific to genome testing. **Results:** The majority of DTC consumers were classified as MM (65%), with H (20%), DI (14. 6%) and only 0. 4% in the HQ group. The DI group had a greater number of medical diagnoses than the MM group (M = 4. 4 vs. 2. 75), with both of these groups older on average than the H group (M = 51 vs. 50 vs. 38). Interest in health information was high (98-99%) in all groups. Each group differed significantly ($p<0.001$) from the other groups on two items (interest in learning about pharmacogenomic test results and learning about risk for other diseases), with the DI group having the highest percent of interest, MM an intermediate, and H the lowest percent. The DI and MM groups also differed significantly ($p<0.035$), while the MM and H groups did not (by Scheffe homogeneity test), on 5 of 7 items related to motivation for testing or personal value of genomic information (e. g. , creating a better plan for the future; reducing the risk of getting sick). Perceived risk was weakly correlated (. 272) with medical diagnosis status alone. In contrast, perceived health status alone, or combined with medical diagnosis status, demonstrated moderate correlations (. 301 and . 326, respectively; $p=0.01$). **Conclusions**: Assessing the concordance or discordance between individuals' perceived health status and actual medical diagnostic status can help understand how consumers, spanning ill and healthy persons, are likely to value and respond to DTC genomic testing services.

2206F

Families At Risk: Long-term Impact of Huntington's Presymptomatic Genetic Testing. *D. JH. Mathews[1], J. Bollinger[1], R. Dvoskin[1], J. Brandt[2].* 1) Berman Institute of Bioethics, Johns Hopkins University, Baltimore, MD; 2) Division of Medical Psychology, The Johns Hopkins Hospital, Baltimore, MD.

For decades, Huntington's disease has served as a model for how we think about genetic testing, and its benefits and risks for at-risk individuals and their families. In 1983, the gene for Huntington's disease (HD) was the first genetic disease mapped using DNA polymorphisms. Shortly thereafter, linkage tests were developed to allow for presymptomatic genetic testing for HD, and two clinical trials began to provide and study such testing. The Baltimore Huntington's Disease Project began in 1986 and enrolled 180 individuals in the first 10 years of the program. Early experience with this cohort influenced collective thinking about key issues related to genetic testing and the return and communication of test results; these challenges are still widely discussed and debated. As increasing numbers of genetic tests are being used to both directly and incidentally assay genes for adult-onset neurodegenerative disease, and as large-scale genetic testing is increasingly integrated into clinical care, it is critical that we understand the implications of presymptomatic testing for at-risk individuals and their family members, not just in the near-term, but over the course of the lives of individuals and their families. However, due to the relatively recent introduction of presymptomatic genetic testing into clinical care, few such studies exist. To address this gap, we are conducting a follow-up study with people who were enrolled in the original Hopkins presymptomatic testing study, their study partners, who accompanied them to counseling and testing/reporting sessions as a requirement of the study, and family members of tested individuals. Here we report the initial results from the first phase of our research, involving one-hour semi-structured interviews with original study participants and study partners. Interview topics include their testing experience, reasons for testing, communication of test results, and impact of testing on their mental health, relationships, and employment. These data will be used to guide focus groups with spouses and adult children of original study participants. Our findings will begin to fill an important gap in our knowledge about long-term implications of presymptomatic testing. Furthermore, they will have direct implications for guidelines regarding when and how to conduct genetic testing for the expanding list of adult-onset neurodegenerative diseases.

2207F

"Something extra on chromosome 5": Couples' understanding of positive prenatal chromosomal microarray analysis (CMA) results. *B. A. Bernhardt[1], A. Russell[2], S. A. Walser[1], R. J. Wapner[2].* 1) Translational Medicine and Human Genetics, Univ Pennsylvania, Philadelphia, PA; 2) Department of Obstetrics and Gynecology, Columbia University Medical Center, New York, NY.

CMA identifies copy number variants (CNVs) associated with known syndromes, risk for neuro-cognitive disorders, an uncertain phenotype, or normal development. When diagnosed prenatally, couples' understanding of the nature and consequences of CNVs impacts decision-making and coping. To assess understanding of positive prenatal CMA results, we conducted phone interviews with 28 women (2 terminating pregnancies and 26 continuing) and 13 male partners 4-8 weeks after receiving positive CMA results from genetic counselors. Transcript analysis assessed how CNVs were described, interpreted and understood phenotypically. We compared this with laboratory interpretation (pathogenic, likely pathogenic, variant of uncertain significance (VOUS), or likely benign). We found nearly all the women were able to describe the chromosome involved, whether the CNV was a deletion or duplication and the range of possible phenotypic effects. Conversely, the men used vaguer language and were much less clear about phenotypic involvement. All men and women know if the CNV was inherited or *de novo*. Most participants initially interpreted results as abnormal or uncertain. For couples who learned through parental testing that a parent carries the same CNV as the fetus, previously uncertain or abnormal results frequently were reinterpreted by the couple as meaning the baby would be unaffected, even if the CNV was classified by the lab as likely pathogenic or VOUS. Some couples receiving VOUS results were reassured by lack of definitive information suggesting a problem, while others were distressed by having limited information about the severity and possible phenotype associated with the CNV. When considering possible clinical involvement, couples tried to imagine the best and worst case scenario they might encounter as parents. Since most were reassured through ultrasound that the fetus was anatomically unaffected, concern focused on learning disabilities and autism risk. These findings suggest that women, but less so male partners, generally understand the nature and clinical implications of prenatal CMA results. Couples feel reassured, perhaps sometimes falsely so, when a CNV is inherited from a "normal" parent. When a CNV is *de novo* without ultrasound findings, even when some prognostic information is available, couples experience considerable uncertainty. Research is needed to identify counseling approaches to aid understanding and adjustment to uncertain results.

2208F

A double edged sword: being a *BRCA1/2* mutation carrier and a man with prostate cancer. *E. K. Bancroft[1,2], C. Moynihan[2], A. Ardern-Jones[1], A. Mitra[2,3], E. Castro[2,4], E. Page[2], N. Taylor[1], R. Eeles[2,1].* 1) Oncogenetics Team, Royal Marsden NHS Foundation Trust, London, United Kingdom; 2) Oncogenetics Team, Institute of Cancer Research, London, United Kingdom; 3) Department of Uro-oncology, University College Hospitals, London, United Kingdom; 4) Prostate Cancer Clinical Research Unit, Spanish National Cancer Research Centre (CNIO), Madrid, Spain.

An increased risk of breast and prostate cancer is observed in men with *BRCA1/2* mutations. No work has reported on the psychosocial impact of having both a *BRCA1/2* mutation and prostate cancer. Methods: Men identified as *BRCA1/2* mutation carriers and with prostate cancer were invited to participate in a qualitative interview study. Data were analysed using a Framework Approach. Masculinity theory was used to explore the impact on men who had been diagnosed with a *BRCA1/2* mutation, perceived as a 'female' condition and having prostate cancer, a uniquely male tumour. Results: 15 men were identified as eligible, of which 12 were interviewed. The umbrella concept of a 'Double Edged Sword' wove its' way through men's responses. Men responded in stereotypical masculine ways as well as exhibiting so called 'feminine' reactions. Despite men's awareness of carrier status, their BRCA mutation experience was described, as 'on the back burner'. The presence of the mutation was only relevant to men in terms of a 'bonus' enabling a pathway to early diagnosis of prostate cancer and responsibility towards family members, mainly daughters. Embodiment of prostate cancer took centre stage indicating a stain on men's biographical timelines, stoicism, control and 'normality'. However men appreciated certain aspects of care that did not 'fit' with masculine stereotypical responses. Conclusion: We suggest that health workers take a reflexive approach that is void of ideal male/female assumptions while respecting differences and similarities between males (and women) together with an exploration of men's needs that are sometimes invisible.

2209F

A UK-wide national audit of genetic testing in retinoblastoma suggests that transition care should be a priority for medical genetics. *A. C. Foster[1], I. Boyes[1], L. Burgess[1], S. Carless[1], V. Bowyer[1], S. Parkes[2], M. Olarinde[3], H. Jenkinson[4], M. Parulekar[4], J. Ainsworth[4], J. Kingston[5], E. Rosser[5], M. A. Reddy[3], T. Cole[1].* 1) Clinical Genetics Unit, Birmingham Women's Hospital NHS Foundation Trust, Birmingham, West Midlands, United Kingdom; 2) UK Cancer Registry; 3) Barts Health NHS Trust, London, United Kingdom; 4) Birmingham Children's Hospital NHS Foundation Trust, Birmingham, West Midlands, United Kingdom; 5) Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom.

Retinoblastoma is the most common malignant tumour of the eye in childhood, with all bilateral tumours and 14% of unilateral tumours due to an autosomal dominant mutation in the RB1 gene present in the germline. Genetic testing should be carried out in all cases to enable accurate risk assessment and inform appropriate medical management and genetic counseling for the affected individual, their siblings, and their future offspring. We carried out the first UK wide national audit of knowledge of genetic testing in patients with retinoblastoma and their families. Questionnaires were sent to contactable individuals in the UK between ages 16 and 45 to assess whether they had the genetic information required for their own management and that of their families. 292 individuals were included, 100 (34%) with bilateral disease and 192 (66%) with unilateral disease. Only 72% of patients with bilateral disease and 32% of patients with unilateral disease understood the genetic implications of retinoblastoma for themselves or family members. In both groups, individuals who did not recall having had a genetic test or did not know the results of testing were significantly less likely to correctly assess recurrence risk for their relatives (bilateral group $p<0.001$, unilateral group $p<0.001$). These individuals were also significantly less likely to have children (bilateral group $p=0.012$, unilateral group $p<0.001$), and yet were significantly more likely to request an appointment to discuss recurrence risks (bilateral group $p=0.005$, unilateral group $p=0.001$), indicating they wanted this information when made available. The results of the audit suggest that individuals with retinoblastoma may be making life choices based on inaccurate or incomplete information regarding recurrence risks and reproductive options. As retinoblastoma services are often seen as a paradigm for genetic medicine, these findings are likely to be representative of many rare genetic diseases, and indicate that a focus on transition care is urgently needed in the current era of mainstreaming genetics.

2210F

Disclosing genetic risk for coronary heart disease: perception of self-responsibility and reaction to risk disclosure in a randomized clinical trial (from the MI-GENES investigators). *S. Brown, H. Jouni, E. Austin, T. Marroush, I. Kullo, MI-GENES Investigators.* Division of Cardiovascular Diseases, Department of Medicine, Mayo Clinic, Rochester, MN.

Whether disclosing genetic risk for coronary heart disease (CHD) to individuals influences psychosocial parameters related to perception of self-responsibility and reaction to risk disclosure is unknown. The myocardial infarction genes (MI-GENES) trial randomized participants aged 45-65 years who were at intermediate risk for CHD based on conventional risk factors and not on statins, to receive their conventional risk score (CRS) or their CRS plus a genetic risk score (GRS) based on 28 susceptibility variants. CHD risk was disclosed by a genetic counselor and then discussed with a physician. Surveys to assess parameters related to self-responsibility (including intention to change and measure of locus control) and reaction to risk disclosure (including impact of events, reaction to risk, and decisional regret) were completed before and three and six months after risk disclosure. We assessed whether these behaviors differed by GRS disclosure. Results were reported as the mean difference (with standard error) in the score for each survey response; significance was determined by regression analysis. Over follow-up, GRS participants felt impacted by stressful life events more than CRS participants (1. 81 ± 0. 77,p=0. 02). Compared to CRS participants, GRS participants embraced a relatively predetermined view of their CHD risk (0. 48 ± 0. 20, p=0. 02), while recognizing their responsibility to appropriately care for themselves in order to avoid increasing their CHD risk (0. 37 ± 0. 17, p=0. 02). Intention to change, reaction to risk, and decisional regret did not differ significantly between the GRS and CRS groups. Disclosure of GRS for CHD resulted in greater measures of self-responsibility (particularly elements of locus control) and reaction to risk disclosure (including the stressful impact of events) in study participants. Self-responsibility and reaction to risk disclosure may serve as psychosocial determinants of the impact of disclosing CHD genetic risk in precision medicine.

2211F

Family advocacy and support in ultra-rare genetic diseases: Example of*OPHN1* mutations. *M. C. Towne[1], T. S. Schwartz[1], R. C. Pelletier[1], A. H. Beggs[1,3], I. A. Holm[1,3], P. B. Agrawal[1,2,3].* 1) Division of Genetics and Genomics and The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA; 2) Division of Newborn Medicine, Boston Children's Hospital, Boston Children's Hospital, Boston, MA, USA; 3) Department of Pediatrics, Harvard Medical School, Boston, MA.

Patients with ultra-rare genetic disorders often feel isolated as little is known about their disorder. Connecting patients with the same disorder provides opportunities to learn more about the condition and offers families the chance to network and bond, giving them emotional support. Here we discuss an 18-year old male, referred to the Gene Discovery Core of The Manton Center for Orphan Disease Research (GDC) for molecular diagnostic evaluation of "Cohen-like" syndrome, and who was negative for *COH1* mutations responsible for most cases of Cohen syndrome. Whole exome sequencing identified a novel variant in *OPHN1* determined to be the likely cause of disease. However, the family wanted to know more and wanted to be connected with other people carrying *OPHN1* mutations. To support the desire of the family to find other individuals with an *OPHN1* mutation, we reached out to corresponding authors on published *OPHN1* case studies, however we had no success. We then approached commercial and academic laboratories offering clinical testing for *OPHN1*. A connection to another individual with an *OPHN1* mutation was made through one of those laboratories. Subsequently, through outreach efforts made by the affected families, the network expanded. We enrolled these additional families in GDC study and gathered their phenotypic information. Within a year, six families with *OPHN1* mutations were identified from around the world through patient outreach including: creating a disease-specific webpage, modifying existing open-access internet resources on the gene, and creating an email address monitored by family advocates. Not only were families connected through these efforts, but less than a year after the first two families were connected, a conference call to discuss drug trials on this condition with published authors, international experts on *OPHN1*, families, and patient advocates was held. Uncovering the genetic etiology for families is only the first step in providing complete care to a patient. Outreach efforts have connected six families with *OPHN1* mutations has connected six families. In addition, collaboration between families, researchers, clinicians, and the clinical diagnostic labs will potentially make a huge difference towards management of this ultra-rare condition.

2212F

Considerations for genetic testing of the partners of individuals with autosomal dominant cancer susceptibility syndromes that cause a severe recessive phenotype. *R. L. Bennett, L. Amendola, W. Raskind.* Dept Med/Div Med Gen, Univ Washington, Seattle, WA.

There are several autosomal dominant (AD) cancer susceptibility syndromes that are known to cause a severe recessive phenotype when two pathogenic variants (PVS) are inherited. Constitutional mismatch repair deficiency (CMMR-D) is caused by biallelic mutations in one of the Lynch syndrome genes (*MLH1*, *MSH2*, *MSH6*, *PMS2*). Individuals who are biallelic for PVS in *BRCA2* have Fanconi anemia, and severe phenotypes in offspring who are biallelic for *BRCA1* PVS and *PALB2* PVS have been described. Heterozygotes for an *ATM*pathogenic variant (PV)have an increased risk of breast and possibly other cancers, and those who are biallelic for PVS in *ATM*have ataxia-telangiectasia (AT). Increasingly, individuals heterozygous for PVS in *FH* are recognized to have an increased risk of renal cancer (HLRCC), and biallelic offspring have the severe metabolic disorder fumarate hydratase deficiency. It is standard practice to offer genetic counseling to couples when one of the partnersis heterozygous for a PV causingan autosomal recessive condition, even when the risk to have an affected childis less than 1%. For individuals who carry aPV in a gene associated with anAD cancer susceptibility syndrome where there are known founder mutations, the possibility of identifying a mutation in the partner may be high (such as for the Ashkenazi founder mutation in *BRCA2*, or the American founder mutation in *MSH2*). The carrier frequency of *ATM* is estimated to be as high as 1 in 100 in some populations, thus a person identified with a PV in *ATM*through testing for hereditary cancer syndromes would have as high as a 1 in 400 chance to have a child with AT. In this context, testing the partner of an individual who carries an ATM PV for mutations in *ATM* may be indicated. In contrast, for rare conditions like HLRCC genetic testing of partners may not be recommended given the low likelihood of having an affected child. The clinical evaluation of the partner including a detailed family history and consideration of the age and sex related penetrance of the condition should be taken into account. We discuss populations that may have higher risk to inherit biallelic muations in the AD cancer susceptibility genes and address genetic counseling issues including discussions of reproductive decision making, age at testing, insurance coverage for testing, identifying consanguineous relationships, and approaches to germline genetic testing.

2213F

Population Screening for Hereditary Cancer: Opinions from the General Public. *L. A. Cross[1], E. Lawrence[1], R. E. Grubs[1], A. Durst[1], R. Jankowitz[2], D. Thull[1,2], D. Stephan[1].* 1) Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Hematology and Oncology, University of Pittsburgh Medical Center, Pittsburgh, PA.

Approximately 10% of cancers are due to a mutation in a single gene. Testing technologies, which detect genetic mutations associated with hereditary cancer, continue to improve to be more accurate and affordable. With this, population screening to identify high-risk individuals and manage them appropriately becomes feasible. This study aims to examine how individuals in the general population view screening for hereditary cancer syndromes without a family history, as well as potential psychosocial impacts. We developed a 57-question survey administered through the Qualtrics Survey System to students at the University of Pittsburgh, University of Northern Iowa, and will be delivered to Amazon Mechanical Turks. The study population currently consists of individuals 18–50 years old (89 female, 150 male, 1 unspecified). Participants were initially asked if they would want to participate in a screening program for hereditary cancer syndromes. 63% (152/240) wanted to participate now, 31% (74/240) would consider screening later in life and 6% (14/240) would never consider screening. The 226 participants interested in screening were presented with 5 scenarios in the context of population screening to evaluate the psychosocial impacts using both quantitative and qualitative means. The scenarios posed involved receiving results that increased the cancer risk to 15-25%, 50%, and 80%, as well as 2 scenarios with a 0% increased cancer risk, one with a diagnosis of cancer 15 years later. At the 50% level, 40% of respondents stated it would affect daily life, 45% stated it would affect mental health, 14% stated it would affect the ability to find a partner, and 9% stated it would affect the ability to find/keep a job. After the scenarios, 5% (11/226) indicated they would no longer want to participate in population screening. This data suggests the majority of participants are interested in population screening for a hereditary cancer predisposition. Recent FDA guidance is unclear if mutation carrier status falls under the loophole allowing “carrier tests” to be delivered with no physician intermediary. Our data suggests that the need for genetic counseling and other support is important. The recent launch of low cost *BRCA1/2* tests in the marketplace with initial utility and cost savings of population-based screening are exciting, but should be coupled with counseling support for all consumers and downstream clinical connectivity for mutation positive consumers.

2214F

Measuring Awareness and Identifying Misconceptions About Genetic Counseling Services and Utilizing Television to Educate. *D. Goldberg[1], M. Bocian[1], K. Singh[1], K. Osann[1], W. W. Grody[2], J. Austin[3]*. 1) University of California Irvine, Irvine, CA; 2) University of California Los Angeles, Los Angeles, CA; 3) University of British Columbia, Vancouver, BC, Canada.

Understanding awareness and perceptions of genetic counseling (GC) is important in identifying and overcoming potential barriers to GC services. However, there are relatively few empirical data regarding these factors among US-based populations. To address this, we attended various community events for the general public, disability community, and new parents and recruited participants for a survey-based study comprising demographic questions, closed-ended knowledge-based and awareness questions, and open text sections. We applied descriptive statistics to responses about demographics, awareness of GC, purposes of GC, and perceptions of GC practice. In total, 320 individuals participated, including 69 from the general public, 209 from the disability community, and 42 from the new parent community. Slightly more than half of respondents (n =173, 54%) had heard of GC. Risk assessment and counseling were among the most frequently cited activities attributed to genetic counselors; a few felt that GC was related to eugenics. Respondents thought that GC aims to prevent genetic disorders (n=82, 74%), helps people find their ethnic origins and understand their ancestry (n=176, 55%), advises people whether to have children (n=140, 44%), and helps couples have children with desirable characteristics (n=126, 39%). Our data showed the majority of participants preferred to watch a medical thriller involving genetic counseling, followed by documentary series; comedy was rated the lowest. These data revealed gaps in awareness of GC and misperceptions about its purpose and can be useful in devising targeted interventions by developing entertainment-based education to improve public knowledge of genetic health and the roles of GCs.

2215F

Evolution of Genetic Counseling in Genomic Medicine-Experiences from Mayo Clinic Center for Individualized Medicine. *K. Hunt[1], B. McLaughlin[1], T. Kruisselbrink[2], K. Guthrie[3], K. Lazaridis[2], T. McAllister[2], M. Ferber[2], E. Klee[2], E. Winkler[2], K. Johnson[4], K. Schahl[4], E. Wieben[2].* 1) Hem/Onc, Mayo Clinic, Scottsdale, AZ; 2) Center for Individualized Medicine, Rochester, MN; 3) Center for Individualized Medicine, Jacksonville, FL; 4) InformedDNA, St. Petersburg, FL.

Statement of Purpose: The role(s) of genetic counselors need to evolve at a pace concomitant with the integration of next-generation sequencing (NGS) into clinical care. Genomic testing (multi-gene panel tests, exome and genome sequencing) is associated with: more complex informed consent discussions, scalability issues surrounding genomic data, challenges with variant classification and healthcare management implications (limited or no guidelines), ethical issues associated with database collection and return of research participants' results and the ongoing need to educate medical providers and the public about the latest genomic advances. These challenges have resulted in the development of new roles for genetic counselors and the advancement of genetic counseling techniques and skills. **Methods Used:** The Center for Individualized Medicine (CIM) designed and opened an Individualized Medicine (IM) Clinic in 2012 with the goal of integrating personalized medicine into clinical care. The initial phase of this project was to implement an Individualized Medicine (IM) Clinic with three distinct services: 1. Oncology Genomics, 2. Diagnostic Odyssey and 3. Healthy Genome Sequencing. Another core component of CIM is the Biobank Program. **Summary of Results:** Four new genetic counselors were hired between three Mayo Clinic locations to support the advancement of genomic testing into clinical practice. In addition, existing genetic counselors at the three sites were recruited to contribute to the development of the IM clinics. The roles and responsibilities of the genetic counselors within the IM clinic were expanded to support the unique elements and challenges of genomic medicine. Expanded functions of the genetic counselor include: content development for participant interactions within the Biobank and developing plans with investigators on return of results, variant interpretation assistance, integration of genomic results into clinical care, expanded genomic counseling sessions that incorporate discussions about incidental findings, implications of a variant of uncertain significance (VUS) and potential need for long-term follow-up and re-testing and developing patient and physician educational resources. As a result of these efforts, the roles and responsibilities of genetic counselors are evolving to meet the needs of the advancing genomic technologies and the subsequent challenges incorporating these technologies into patient care.

2216F

NCBI Hackathons: Community Driven User Centered Design of Software to Interface with Genomics Datasets. *B. Busby.* NCBI, Bethesda, MD.

After we established that genomics professionals from the community could build deployable software for genomic analysis in an educational hackathon setting, (http://dx. doi. org/10. 1101/018085) we wanted to go farther with implicitly user-centered software prototyping. Therefore we identified three general categories where available open-source software tools to interface with NCBI datasets were lacking, and set out to build deployable software prototypes, guided by members of the genomics community. The three categories we identified were RNA-seq, translational genomics, and education/democratization of access. We built nine teams of volunteers, three for each category, drawing on the previous experience of these individuals, as well as their motivations for attendance. Drawing from our experiences in the January 2015 hackathon, we understood that the major goal of attendees is to finish deployable software, and therefore we endeavored to enable August attendees to do so by breaking software production into smaller pieces, to enable attendees to work through them in three days and then asking three veteran hackathoners (who are experienced Genomics professionals) to help attendees stitch the work together, and write tests for the resultant software from the other two categories. The resultant software from NCBI hackathons is freely available under a creative commons license at github. com/DCGenomics.

2217F

Best Practices in Genetic Cancer Risk Assessment: Delivery Models and Barriers in Community Settings. *I. Solomon, B. Nehoray, M. Niell-Swiller, C. Rybak, K. Yang, J. Weitzel, K. Blazer.* Clinical Cancer Genetics, City of Hope National Medical Center, Duarte, CA.

Purpose: Genetic cancer risk assessment (GCRA) is an interdisciplinary clinical service that incorporates genetics, oncology, and patient/family counseling skills to identify and prescribe personalized screening and preventive care for individuals with increased cancer risk. Advances in genetic technologies are fueling demands for GCRA services in community settings. This study explored practices and needs of 7 cohorts of community-based clinicians who participated in an interdisplinary program of CME-accredited GCRA training and practice support from 2010-2015. **Methods:** A two-part survey comprised of check boxes, ratings scales and open-ended prompts addressing GCRA practice issues was administered to course participants after distance learning and prior to face-to-face workshops. Descriptive and comparative analyses were performed with SAS v. and qualitative analysis was conducted on open-ended responses. **Results:** Surveys were completed by 277 participants: 130 MDs (47%), 48 genetic counselors (GCs) (17%), 93 nurses (33%), 6 other (2%); 78% dedicate <50% of their time to GCRA, and 59% planned to expand these services. A majority (84%) conducted initial GCRA sessions in-person; 62% conducted in-person results disclosures; 32% reported using a multidisciplinary team approach to GCRA. More than half (58%) spent >1 hour with patients for the initial GCRA session; 80% spent >30 minutes for results disclosure counseling. Compared to MDs, GCs more often reported that they adequately reviewed basic genetics, family history, inherited syndromes, and interpretation/communication of test results during GCRA sessions; GCs and Nurses more frequently provided resources for GCRA patients (p<. 05 for all comparisons). No significant differences were observed by cohort. GCRA practice workshop outcomes revealed lack of time, financial compensation and administrative support, and need for continuing education as the most significant barriers to expanding best practices in GCRA and research collaboration across all practice settings. **Conclusion:** Our findings suggest that motivated community clinicians are dedicating significant resources to developing best practices in GCRA. This demonstrates the value of interdisciplinary team-based training, and points to the need for continuing education and support for practice and research collaborations for community-based providers.

2218F

The GEC-KO Website: Genetics education for primary care providers. *J. C. Carroll[1,2], S. Morrison[3,4], J. Allanson[4,5].* 1) Family and Community Medicine, University of Toronto, Toronto, ON, Canada; 2) Family and Community Medicine, Mount Sinai Hospital, Toronto, ON, Canada; 3) Genetics Education Canada Knowledge Organization, Ottawa, ON, Canada; 4) Children's Hospital of Eastern Ontario, Ottawa, ON, Canada; 5) University of Ottawa, Ottawa, ON, Canada.

Background: Without integration into primary care, the promises of genomic medicine for patients are unlikely to be fully realized. Internationally, primary care providers have expressed lack of knowledge of and confidence in core primary care genetics competencies. Needs assessments have indicated that relevant, brief, reliable, evidence-based, up-to-date primary care specific resources are essential. The GEC-KO website (Genetics Education Canada-Knowledge Organization www. geneticseducation. ca) had a soft launch in late 2013 as an accessible platform for resources and tools designed to increase genetics literacy in healthcare professionals. It is one component of a larger educational program facilitating translation of research and enabling development, collection, dissemination and evaluation of genetics educational materials. Educational materials on the website are designed to meet the needs of Canadian primary care providers and are informed by previous work evaluating primary care genomics knowledge translation interventions. Purpose: To describe the types of educational materials on the GEC-KO website and provide analytics which assess the pattern of use of the website and demographics of users. Methods: Google analytics were used to assess website use comparing three 6 month periods since the soft launch of the 2013 website. Results: GEC-KO educational materials include point-of-care tools, summaries of genetic disorders and referral and testing guidelines, and contact information for genetics clinics. In the 6 month period from Nov 1 2014 to April 30 2015, 2,021 individuals accessed the site, 82% of them new users. This was an increase from 465 and 860 users in the previous two 6 month periods. There were 5,333 page views, with an average of 2. 18 pages/session. Users included those from Canada (50%), USA (13%), Brazil (8. 5%), and France (2%). Users were directed to the site by referral from another site (39%), organic search using search engines such as Google (33%), direct URL address (25%) and a small number through social media (2. 5%). Top pages accessed have varied over time but include hemochromatosis, non-invasive prenatal testing and genetics clinics contact information. Conclusions: Early results show that users are accessing these relevant, evidence-based up-to-date, concise, accessible educational resources. More work is needed to evaluate these resources and whether they change practice and patient outcomes.

2219F

The Role of Graduate Year and Training in Cancer Genetics Knowledge in Primary Care Physicians. *V. Chan[1, 4], D. Tegay[2], W. Blazey[3], S. Koehler[2], B. Krishnamachari[4].* 1) Academic Medicine Scholars Program, NYIT College of Osteopathic Medicine, Old Westbury, NY; 2) Department of Medicine, NYIT College of Osteopathic Medicine, Old Westbury, NY; 3) Office of Academic Affairs, NYIT College of Osteopathic Medicine, Old Westbury, NY; 4) Division of Research, NYIT College of Osteopathic Medicine, Old Westbury, NY.

INTRODUCTION Knowledge about cancer genetic testing can aid cancer screening and prevention in high-risk patients. In 1996 and 2003, the American Society of Clinical Oncology (ASCO) respectively issued the original and updated ASCO Statement on Genetic Testing for Cancer Susceptibility, setting recommendations for cancer genetic testing and clinical practice. Physicians graduating before 1996 or without cancer genetics training may not have adequate exposure to current standards of care for hereditary cancer syndromes. We hypothesized that physicians who graduated after the ASCO guidelines and reported training would have greater cancer genetics knowledge. **METHODS** A survey was administered to predominantly primary care physicians evaluating their graduation date and cancer genetics training, practice patterns, and knowledge. Age-adjusted odds ratios (ORs) and their 95% confidence intervals (CIs) were calculated using unconditional logistic regression. **RESULTS** 159 physicians completed the survey. Graduates after 1996, compared to graduates before, were more likely to respond correctly to the items, "For cancer syndromes affecting most women, maternal history is more important than paternal history" (OR=5. 79, 95% CI=2. 60–12. 9) and "The BRCA mutations found in Ashkenazi Jewish individuals are generally one of three specific mutation" (OR=3. 29, 95% CI=1. 42-7. 60). In addition to the above items, graduates after 2003, compared to graduates before, were more likely to correctly respond to "Genetic sequencing will detect all types of genetic mutations" (OR=2. 64, 95% CI=1. 04–6. 75). While graduates after 1996, compared to graduates before, were more likely to receive cancer genetics training in medical school (OR= 3. 70, 95% CI=1. 74-7. 84) and residency (OR=2. 57, 95% CI=1. 27-5. 14), they were also likely to be less likely to receive continuing medical education (OR=0. 16, 95% CI=0. 07–0. 34). Nearly 90% of physicians in both groups expressed interest in participating in future training on hereditary cancer risk assessment. **CONCLUSION** To our knowledge, this study is the first to analyze physician graduation date and cancer genetics training and knowledge. This study suggests that there is a need for more genetics training in physicians who graduated before the original and revised ASCO guidelines.

2220F

Blended Learning for Primary Care Providers in Cancer Genetics Management. *K. Reed[1], E. Edelman[1], T. Ingram[1], M. Dougherty[2].* 1) Clinical and Continuing Education Program, The Jackson Laboratory, Bar Harbor, ME; 2) Education Dept, ASHG, Bethesda, MD.

Changing provider behavior, as is necessary for the integration of genetics into clinical care, requires more than a single educational intervention. Educational programs that focus on competencies, multiple instructional strategies and repeated exposures to key messages have been associated with improved outcomes in continuing education. The Jackson Laboratory and the American Society of Human Genetics developed an educational program to improve primary care providers' (PCPs) clinical skills in assessing and managing genetic risk for cancer. This blended learning program consists of a 6 ½ hour in-person workshop and 12 months of digital communications. The program addresses competencies related to assessing a targeted family history; assessing the clinical utility of testing; facilitating patient decision-making; and using guidelines to manage care. The program was piloted with 21 early adopter PCPs in November 2014. Participants completed knowledge, attitudes, and confidence assessments pre, post and 4 months post, which were analyzed by Fisher exact tests. Immediately post-workshop, PCPs demonstrated improved confidence in their ability to determine appropriate referrals, discuss the benefits and limitations of genetic testing, provide counseling about management related to genetic information, and improved overall confidence in genetics of hereditary cancer (all $p \leq 0.01$), as well as improved knowledge on interpreting genetic test results ($p=0.02$). The participants' confidence and knowledge levels at 4 months post-workshop returned to nearly pre-workshop levels. Of the 4 monthly communications received post-workshop, 19-71% participants visited the associated webpage and 11-63% of those engaged with the respective educational intervention. These results indicate that our approach for the workshop was effective in improving confidence and knowledge to incorporate genetics. However, we did not observe sustained gains 4 months post-workshop, which we attribute, in part, to the lack of participant engagement with the reinforcing activities. We need to identify effective approaches to engage participants with digital resources. As we implement the workshop more broadly, we will refine and test different communication approaches for ongoing engagement to maximize the impact of blended education - interactive in-person education and reinforcing digital communications - on practice change.

2221F

A co-ordinated national approach to workforce transformation: Health Education England's Genomics Education Programme. *A. Seller[1, 2], M. Bishop[1], JP. Maytum[1, 3], V. Davison[1, 3], S. L. Hill[1, 3], Genomics Education Programme Team.* 1) Genomics Education Programme , Health Education England , Birmingham , United Kingdom; 2) Oxford Regional Genetics Laboratories, Oxford University Hospitals NHS Trust, Oxford, United Kingdom; 3) NHS England, Skipton House, 80 London Road, London SE1 6LH, United Kingdom.

England aims to become the first country to introduce genomic technologies into routine healthcare, and has launched the 100,000 Genomes Project, led by Genomics England, to sequence 100,000 genomes from patients with cancer and rare diseases by Dec 2017. This ambitious project, combined with the integrated nature of England's National Health Service (NHS), provides a unique opportunity to implement a co-ordinated approach to workforce transformation to successfully embed genomic medicine into mainstream clinical practice. The Health Education England Genomics Education Programme (GEP) was established with a threefold purpose: to support those NHS staff directly working on the 100,000 Genomes Project through NHS Genomic Medicine Centres; to support the wider transformation of NHS services through genomic medicine and to upskill existing NHS staff. Our strategic approach provides a range of tailored education and training that meets the learning and professional needs of staff, based on their interface with genomic medicine. Key achievements to date include:1. Establishing an active web platform (www. genomicseducation. hee. nhs. uk), supported by social media (@genomicsedu)2. Developing a multiprofessional Masters in Genomic Medicine to be delivered by a network of 9 universities, with individual course modules available for continuing professional development. 3. Direct funding of 550 Masters places & 1000 course modules4. Development of free online courses, including introductions to genomics & bioinformatics and a course supporting the consent requirements of the 100,000 Genomes Project5. Increasing workforce capacity & capability by commissioning 27 doctoral-level scientist training positions in Genetics and in Molecular Pathology of Acquired Disease6. Funding a national network of genetic diabetes nursesEarly indications that the workforce is engaging with our programme is evidenced by the number of registered learners (1,382) and rapid growth on social media (c3,100 followers). Preliminary findings indicate our resources meet learner needs: 95. 2% of participants state staff should complete the consent module before recruiting patients to the 100,000 Genomes Project. The GEP's work will continue to be underpinned by this evidence-based approach to ensure the strategy for workforce transformation is effective. While the results are facilitated by our integrated NHS, this structured approach is transferable to other healthcare systems.

2222F

Pharmacogenomics Assessment in a Current Internal Medicine Practice. *C. R. Vitek, T. B. Curry, P. J. Caraballo, J. Wilson, J. Giri, C. M. Formea, W. T. Nicholson.* Center for Individualized Medicine, Mayo Clinic, Rochester, MN.

Background: Pharmacogenomics is the union of individual genetics and pharmacology. Current expansion for use of pharmacogenomics at the bedside is underway, accelerated by initiatives in personalized medicine. Many providers that were unexposed to pharmacogenomics concepts during training are now required to apply this science to therapeutic decisions. It is essential that health care providers are knowledgeable in these concepts as medicine advances. Methods: As part of an institutional educational grant, an evaluation was conducted to assess self-perceived confidence in pharmacogenomic knowledge and the requirement for additional education for practicing internist. An e-mail-based assessment which consisted of a 15-question, multiple-choice survey was provided to 145 internal medicine practitioners. These included physicians, nurse practitioners, and physician assistants within the Division of General Internal Medicine at Mayo Clinic. The assessment tool employed a Likert scale and was administered using REDCap® program software. Results: A total of 43. 4% (63/145) surveys were completed. Over 90% (58/63) of the respondents agreed pharmacogenomics was relevant to their current clinical practice. These 90% (58/63) also agreed they should be able to provide information to patients on appropriate use of pharmacogenomics testing. However, when self-perceived knowledge was assessed, only 24% (15/63) agreed they could identify medications that require pharmacogenomic testing. Additionally, 84% (53/63) disagreed they could apply the results of pharmacogenomic tests to drug therapy selection, dosing, or monitoring,Conclusions: Overall, internal medicine practitioners believed that pharmacogenomics was relevant to their current clinical practice and they should be able to provide information to patients. However, most practitioners did not feel confident in their knowledge of applicable medications or their ability to apply the results clinically. These outcomes indicate an opportunity to develop and provide clinically applicable pharmacogenomics education to improve practitioner knowledge.

2223F

Workshop in Applied Genomic Medicine. *A. D. Gilbert[1,2], C. R. M Marshall[1,5,6], P. Ray[1,3,4,5,6], S. W. Scherer[1,3,4,5], R. D. Cohn[1,2,3,4], S. Bowdin[1,2].* 1) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto , Ontario, Canada; 2) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; 3) Program in Genetics and Genome Biology, Research Institute, The Hospital for Sick Children, Toronto, ON, Canada; 4) Department of Paediatrics and Molecular Genetics, University of Toronto, Toronto, ON, Canada; 5) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada; 6) Genome Diagnostics, Department of Paediatric Laboratory Medicine, The Hosptial for Sick Children, Toronto, ON, Canada.

Background: The completion of the Human Genome Project and the subsequent introduction of clinical genome and exome sequencing (CGES) have placed clinical genetics healthcare professionals (CGHP) at the forefront of genomic medicine. These new diagnostic modalities are changing not only the scope of medicine but the overall scope of practice for the CGHP. Providing accessible venues for geneticists and other health care providers to acquire the skills necessary to integrate CGES into patient care is of paramount importance. Over the last two years, a team from SickKids, The Centre for Applied Genomics (TCAG) and the University of Toronto has been developing a concept of introducing genomic medicine into paediatric healthcare. With this experience, the team has designed a workshop that addresses the evolving educational needs. **Goals:** At the end of this workshop, participants will be able to: integrate CGES into their clinical practice, assess phenotypic and bioinformatics data, debate counseling and ethical considerations, and describe translational and research methodologies. **Method:** We have designed a 2. 5 day applied genomic medicine case-based Workshop combining a didactic component and interactive breakout sessions, focusing on the key components of performing clinical genomic sequencing and returning results. This is hosted by SickKids Hospital with experts from Canada, the US and the UK comprising the team of instructors and facilitators. A target audience of clinical genetics trainees and practicing clinicians; molecular, biochemical, and cytogeneticists; genetic counsellors, and bioinformaticians working in the field of genomics, nationally and internationally, will be encouraged to attend. To ensure the participation of both geneticists in training and geneticists in underprivileged nations, scholarships will be available. **Outcome Measures:** Workshop participants will complete a pre-workshop questionnaire to capture current knowledge, methods and practices regarding CGES. A post-workshop survey will be administered to determine whether or not the attendee has achieved their goals. Finally, a 6 month follow-up evaluation will be sent to all attendees to obtain information regarding potential changes and/or improvements in the application of CGES. If successful the workshop will be offered every year in Toronto or at other international venues.

2224F

GTR, ClinicalTrials. gov, and ClinVar: NCBI resources that improve access to information about research testing. *W. S. Rubinstein, A. J. Malheiro, D. R. Maglott, B. L. Kattman, V. Hem, B. Gu, M. Ovetsky, K. S. Katz, S. Chitipiralla, G. Song, M. J. Landrum, J. M. Lee.* National Institutes of Health, National Library of Medicine, National Center for Biotechnology Information, Bethesda, MD.

Scientists, clinicians, patients and families all have an interest in access to information related to research-based genetic testing. This information is dispersed among the many institutions supporting such studies. The patient advocacy community has become a driving force in connecting participants and genetic research studies, and the research community has organized around collection of standardized data elements. ClinicalTrials. gov has redoubled efforts to gather study results, but does not organize around genetic topics. NCBI integrates information about genetic tests, studies, and research results to enhance access to and enrollment in research studies and to promote discovery and human health. The NIH Genetic Testing Registry (GTR®; http://www. ncbi. nlm. nih. gov/gtr/) is a free online resource supporting unrestricted access to comprehensive genetic test details and contextually relevant information. To our knowledge, GTR is the only website that supports access to genetic research tests performed worldwide with detailed information about the tests and associated studies. As of June 2015, GTR has 261 research tests covering 384 phenotypes offered by 58 laboratories in 11 countries. In addition to the test description, research tests include study information such as eligibility criteria, participation requirements, study site location, investigators, and study contact information. Test providers can offer details about whether or not participants have an opportunity to learn about the results of the research test and whether confirmation may be done in a CLIA-certified lab, with information about possible associated costs. Test providers can provide a downloadable study consent form and a ClinicalTrials. gov identifier that supports access to further study details. When results are obtained, submitters to ClinVar (http://www. ncbi. nlm. nih. gov/clinvar/) can document the methods used to obtain data by citing the GTR test accession. The purpose of research tests is categorized by submitters as either a test performed to contribute to generalizable knowledge or for a laboratory to generate data to make technical improvements to a test. 'Contribute to generalizable knowledge' is applicable if the intent is to publish the study results and/or the study protocol is approved by a research ethics committee. The integration of GTR, ClinicalTrials. gov, and ClinVar thus advances public health and research into the genetic basis of health and disease.

2225F

Knowledge and attitudes of nurses about a population-based education and carrier screening program for recessive diseases in Cree communities of Northern Quebec. *J. Le Clerc-Blain[1], V. Gosselin[2], A. Bearskin[3], J. E. Torrie[2], G. A. Mitchell[1,4,5], B. J. Wilson[6], A. Richter[1,4,5], A-M. Laberge[1,4,5].* 1) Research Center, CHU Sainte-Justine, Montreal, QC, Canada; 2) Cree Board of Health and Social Services of James Bay, Chisasibi, QC, Canada; 3) Eeyou Awaash Foundation, Chisasibi, QC, Canada; 4) Medical Genetics Division, CHU Sainte-Justine, Montreal, QC, Canada; 5) Department of Pediatrics, University of Montreal, Montreal, QC, Canada; 6) School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, Ottawa, ON, Canada.

BACKGROUND: A population-based educational and carrier screening program (ECSP) was started in 2006 for Cree encephalitis (CE) and Cree leukoencephalopathy (CLE), two neurodegenerative autosomal recessive conditions with high carrier rates in James Bay Cree communities (Canada). Developed by local health authorities in collaboration with the Eeyou Awaash Foundation (EAF; a community family support group), the CE-CLE ECSP informs the population about CE-CLE, availability of carrier screening, and reproductive options for at-risk couples. **OBJECTIVE:** To describe knowledge and attitudes of nurses working in Cree communities about the CE-CLE ECSP. **METHODOLOGY:** Surveys were handed out to health professionals attending the Cree Board of Health and Social Services of James Bay's (CBHSSJB) Annual Nurses Training in November 2014. Data was collected on demographics, knowledge of CE-CLE (clinical features, myths, inheritance, carrier risk), and attitudes about the program. Descriptive statistics are reported. **RESULTS:** Of 86 meeting attendees, 80 nurses answered the survey. The majority (75/80) was aware that the CE-CLE ECSP is offered on the territory. Participants scored high marks on the disease knowledge questions, but less so on inheritance patterns, prenatal diagnosis, and program activities. Most considered that their role is to refer individuals with positive family history (66/80), to refer pregnant women (63/80), and to give information about the CE-CLE ECSP to individuals of reproductive age (62/80). One third of participants (26/80) felt they did their job when clients chose to do carrier testing. A few felt they did not do their job when someone refused carrier testing (6/80), or perceived some pressure to screen all those eligible (17/80). **FURTHER STUDIES:** We are continuing to administer surveys to other local health professionals (nurses, community workers, community health workers, physicians) who were not present at the Nurses Training. **CONCLUSION:** Participants are aware of the program. General knowledge is satisfactory. Some participants equated success with clients choosing to be tested, or felt pressured to screen all those eligible. This suggests that a subset of nurses are not fully informed about the mission of the CE-CLE ECSP, which is to make carrier screening available on a voluntary basis to those wanting to know their carrier status for reproductive decision-making. Thus, clearer communication of the program's mission is needed.

2226F

Residents as Teachers and Leaners: Transitioning Traditional Seminar Series to Team-Based Learning to Prepare Genetics Residents to be Teachers. *D. S. Regier[1], D. H. Hadley[2], P. S. Hart[2], M. Muenke[2].* 1) Childrens National Health System, Washington, DC; 2) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD.

Geneticists need to be not only experts in genetics and genomics but need to be able to share this expertise with students and other health care professionals. This is a very lofty goal for a group of trainees without formal training in educational theory, and many trainees may have never participated in emerging education styles, such as team-based learning (TBL) and the flipped classroom (FC). When these trainees enter their first faculty role, they are often unprepared to participate in educational programs using these styles, and may be less willing or enthusiastic about teaching. By improving trainee's confidence in their ability to educate using emerging educational styles, we propose that this may improve their ability to serve as educational leaders. A TBL seminar series was developed for the 2013-14 academic year. Residents and fellows led sessions designed to teach genetic disease or techniques. Learners took turns being the Learner Leader for each session and created the curriculum for the TBL. During the second year of intervention, 2014-2015 academic year, no additional changes were made to allow for all learners to have an opportunity to implement this teaching style. Participants increased their confidence in leading both participating in and leading a TBL session. Atbaseline, learners were asked to quantify their comfort level with leading sessions using a Likert scale. This was scored after participating in a session with an experienced TBL presenter. Thirty-three percent of learners reported that they were comfortable with leading a TBL session with some help; none willing to mentor another trainee. After one year of participating in TBL classrooms, 80% of participants reported themselves as confident enough with the teaching style to either facilitate a TBL session alone or be a mentor for new trainees preparing a TBL session. Learners reported increased engagement and knowledge retention, based on free text comments regarding the class changes. After two years, learners were unwilling to return to a traditional didactic-style series due to the improved learning environment of the TBL sessions. Overall, this intervention has increased learner's confidence in leading PBL sessions. Further studies are needed to determine if these interventions improve learner's ability to communicate with others in the medical fields, foster genetic and genomic literacy and/or improve recruitment to the field of genetics.

2227F

Developing Online Learning Modules for Medical Genetics Residency Training. *T. Wang[1,2], H. Vernon[1,2,6], M. Gunay-Aygun[1,2], V. Corson[1,3], J. Axilbund[4], G. MacCarrick[1], C. Applegate[1], A. Gherman[1], P. Sosnay[1,5], J. Bodurtha[1,2,6], D. Valle[1,2].* 1) Inst Gen Med, Johns Hopkins Univ, Baltimore, MD; 2) Department of Pediatrics, Johns Hopkins University, Baltimore, MD; 3) Department of OBGYN, Johns Hopkins University, Baltimore, MD; 4) Department of Oncology, Johns Hopkins University, Baltimore, MD; 5) Department of Medicine, Johns Hopkins University, Baltimore, MD; 6) Kennedy Krieger Institute, Baltimore, MD.

With rapid advances in genomics technology and information from the Human Genome Project, medical genetics has progressed from the study of rare Mendelian diseases to a rapidly expanding appreciation of the genetic contribution to all diseases in all areas of clinical medicine. To meet the challenges of training physician medical geneticist in the 21st century, the American College of Medical Genetics (ACMG) has published 65 core competency areas (>300 specific competencies) that are required for medical genetics residents to achieve during their clinical training. The performance of each resident in these competency areas will be evaluated as a part of the ACGME milestone program. However, no systematic learning and evaluation tools are readily available to assist individual residency programs to achieve these training objectives. We have developed online, case-based learning modules to cover all core-competency areas outlined by the ACMG during the 18 months of required clinical residency training in medical genetics. Each learning module is focused on 1-2 clinical competency areas consisting of relevant clinical cases, key teaching points, practice guidelines, core clinical and/or research articles, and pre- and post-tests. These modules are housed at our residency education website (https://igm. jhmi. edu/residency-main) with password-protected access. Each resident is required to complete one learning module during a relevant clinical rotation (2 weeks) and participate in a focused teaching and evaluation session with an attending geneticist at the end of the rotation. Progress of the residents toward learning objectives is measured by pre- and post-tests, feedback from attending physicians, and in-training examination (ITE). These data are collected as a part of milestone-based evaluations for individual resident during his/her medical genetics residency training. We are piloting these learning modules for medical genetics residents during their first and second years of clinical rotations at the Johns Hopkins Hospital and affiliated training sites. We anticipate that these learning modules will be a valuable training tool for medical genetics residency programs to achieve and evaluate clinical competencies for their trainees.

2228F

Development of a competency-based genomic education resource for physicians. *K. W. Weitzel[1], E. Edelman[2], R. Roberts[3], B. Korf[4], M. Murray[5], J. Jenkins[6], ISCC Education Products Working Group.* 1) Personalized Medicine Program, University of Florida College of Pharmacy, Gainesville, FL; 2) Genomic Education, The Jackson Laboratory, Bar Harbor, ME; 3) University of Ottawa Heart Institute, Ottawa, ON; 4) Department of Genetics, University of Alabama, Birmingham, AL; 5) Genomic Medicine Institute, Geisinger Health System, Wilkes Barre, PA; 6) Genomic Healthcare Branch, National Human Genome Research Institute, Bethesda, MD.

Background: Faced with an increasing number of clinically relevant genomic applications, physicians have expressed a need for improved knowledge/skills in genetic risk assessment, testing, and management. While some groups are developing physician educational resources, there has been limited coordination of these efforts, and no single resource addresses this gap. **Objective**: We describe the development and initial assessment of a novel repository of peer-reviewed genetic/genomic educational resources for physicians. **Methods**: The Genetics/Genomics Competency Center (G2C2; http://g-2-c-2. org/) is an online repository created to facilitate development, access, and dissemination of competency-based educational resources for healthcare providers and educators. In collaboration with G2C2, we developed competency-based criteria and a peer-review process to identify and evaluate educational resources aligned with physician needs. Resources were solicited from member organizations of the Inter-Society Coordinating Committee for Practitioner Education in Genetics (ISCC), evaluated for G2C2 inclusion based on ISCC-developed physician competencies in genomic medicine, and mapped to these competencies. Resources became available on G2C2 in June 2014; usage data for physician resources was collected for a 6-month period post-launch. **Results:** A total of 87% (77 of 89) of submitted resources were accepted, which addressed all physician competency domains (family history, genomic testing, patient treatment based on results, and somatic and microbial genomics). The most common reason for declining a resource was limited scope or patient applicability. Most resources addressed genomic testing, and microbial genomics was addressed the least. Accepted resource types included websites (32%), guidelines and policies (17%), courses (12%), documents (12%), books (6%), articles (5%) and other (18%), with 13% providing CME credit. From January to June 2015, 191 users accessed these resources in 261 sessions and viewed an average of 9 pages/session. **Discussion:** Using a robust quality assessment process, we developed a peer-reviewed, competency-based repository for curating physician genomic educational resources and identified needs for additional materials (e. g. , microbial genomics). Future efforts will seek to grow the available resources and expand the number of clinicians and educators who use G2C2 to access genetic/genomic education.

2229F

Touching Triton: building student understanding of common complex disease risk. *A. Hott, K. East, M. Loftin, N. Lamb.* Educational Outreach, HudsonAlpha, Huntsville, AL.

Life science classrooms often emphasize the exception to the rule when it comes to teaching genetics, focusing heavily on rare single gene and Mendelian traits. In contrast, the vast majority of human traits and diseases are caused by more complicated interactions between genetic and environmental factors. Research indicates that students have a deterministic view of genetics, generalize Mendelian inheritance patterns to all traits and have unrealistic expectations of genetic technologies (Mills Shaw et al. , 2008). The challenge lies in how to help students analyze complex disease risk with a lack of curriculum materials (Lewis and Wood-Robinson 2000, Lewis and Kattmann, 2004). Providing open access to both content resources and an engaging storyline can be achieved using a serious game model (Gee, 2005, Wong et al. , 2007, Annetta et al. , 2009). Touching Triton was developed as a serious game where students are asked to analyze data from a medical record, family history and genomic report in order to develop an overall lifetime risk estimate of six common complex diseases. Evaluation of student performance shows significant learning gains in key content areas along with a high level of engagement.

2230F

Considering Genetics Education Through the Lens of Cognitive and Developmental Psychology. *M. Peffer, M. Renken.* Georgia State University, Decatur, GA.

relationship between student's spatial reasoning skill and genetics misconceptions and (2) assessment of a genetics problem-based learning (PBL) unit with elementary school students. As students enter adolescence, their understanding of genetics necessitates additional focus on the mechanistic basis of genetics, which requires high demands on the learners' spatial reasoning. Previous work indicated that spatial reasoning skills are essential to success in STEM fields. To determine if spatial reasoning ability is related to genetics misconceptions in high school students, we assessed spatial reasoning aptitude using the Project Talent items and genetics misconceptions using questions from the AAAS Project 2061 database. Our preliminary work indicates students who score higher on abstract reasoning and 3D, but not 2D visualization have fewer genetics misconceptions. Our current work examines if improving students' spatial reasoning skills alongside standard genetics curricula will result in an enhanced understanding of genetics. This work offers a novel approach for improving genetics education in adolescents by focusing on the cognitive processes underlying genetics understanding. Very little is known about how students in early grades learn genetics. We designed and implemented a PBL unit for first through fifth graders. Working in multi-age groups, students completed a unit that required students to learn the principles of heredity and evolution and apply them to make predictions about how a fictitious species would evolve. We found that engaging young students in a multi-age genetics PBL unit resulted in improvement in students' understanding of common genetics misconceptions as shown by pre and post unit tests. Consequently, genetics PBL instruction may be effective in situating younger students' genetics learning as to minimize the development of common genetics misconceptions during the early years.

2231F

An Educational Program on Human Genetics in Japan: KYOTO MODEL ; by developing a MANGA cartoon media that can enhance importance of awareness of family health history. *Y. Hiraoka, M. Torishima, N. Akiyama, T. Wada, S. Kosugi.* Medical Ethics and Medical Genetics, Kyoto Univ. Graduate Sch. Medicine, Kyoto, Kyoto, Japan.

[Background] Today genetic technology has been rapidly advancing and genetic testing is largely available in medical practice. In Japan, little chance to learn human genetics at school, and we, genetic counselors and medical geneticists, must play a vital role to promote human genetics literacy of the public for providing optimal genetics service. “Family health history” (1) is an accessible tool that can captures genetics information, lifestyle and environmental factors, (2) allows health care providers to diagnose conditions and to assess risks and (3) increases health and genetics knowledge for the individual and the family. In this study, we developed a MANGA booklet, as a new educational implement of human genetics and family health history, and measure the effectiveness of this new tool for improving human genetic literacy of the public. Japanese MANGA, or a book in cartoon fashion, has been a widely spreading subculture in the world now, and therefore, could be one of the useful ways in human genetics education. [Study] We developed a new type of tool, the MANGA booklet, for improving human genetic literacy. The booklet has basic information about (1) human genetics including diversity of genes and environmental risks, and about (2) family health history, including instructions for gathering detailed medical information for individual family members and instructions to create a pedigree chart, which helps medical providers to predict and identify risks for many common diseases. These were assessed by certified specialists in medical genetics and genetic counselors. And we discussed with cartoonists to remake the MANGA style to be a more accessible and useful tool. As a preliminary study, we are now interviewing the undergraduates at our university to validate the booklet. [Discussion] Family health history is “the first genetic test” in medical practice or genetic counseling. And it is important for both healthy and sick persons to have correct information of their familial information for their own health. The MANGA tool that we have developed seemed to be a practical tool to interest the public in human genetics, and help them learn familial and own health status. We expected it to be accessible, easy to comprehend human genetics and family health history. We are also planning to translate the MANGA textbook into other languages to be available in many countries.

2232F

Genetics Home Reference: Educating the General Public about Precision Medicine. *H. Collins[1], S. Calvo[2], K. Greenberg[1], L. Forman Neall[3], S. Morrison[2].* 1) ICF International, Fairfax, VA; 2) U. S. National Library of Medicine, NIH, Bethesda, MD; 3) Medical Science and Computing, LLC, Rockville, MD.

Clinicians and researchers use precision medicine to assess the impact of genetic, environmental, and lifestyle factors as part of individual variability. This approach targets specialized treatment and prevention strategies for diseases. While much work is needed to build the tools necessary for clinicians and researchers to use a precision medicine model, the general public also needs a resource that explains the concept of precision medicine and elucidates the details of the genetic variables involved. For the past 12 years, Genetics Home Reference (GHR; http://ghr. nlm. nih. gov/) has become a trusted consumer website of the U. S. National Library of Medicine. GHR conveys complex genetic information to the public through brief, plain language summaries of genetic conditions and associated genes with related links to outside resources. GHR is intended as a resource for patients, families, and the general public needing information about genomic medicine. Many of the more than 1,000 genetic condition summaries provided on the GHR website are relevant to precision medicine, including cancers, immune deficiencies and dysfunctions, enzyme deficiencies, multifactorial conditions that involve gene—environment interactions, and drug sensitivities. GHR also has detailed information on more than 1,200 genes, describing their normal function and how their variations contribute to changes in health. Chromosomal changes and their contribution to human disease are also discussed on the GHR website. GHR provides a primer called *Help Me Understand Genetics*, which has chapters discussing pharmacogenomics, genetic testing, diagnosis and treatment, and genomic research. The primer also provides information about precision medicine, including the President’s Precision Medicine Initiative and the potential benefits and challenges of this approach. This presentation describes the resources that GHR provides to engage the public in various aspects of precision medicine. It also includes important context that consumers need to fully understand precision medicine, such as information about genetic conditions, gene function and variability, and overviews of genetic concepts. We discuss how GHR can involve healthcare consumers in the precision medicine conversation by giving them easy-to-understand tools that cover both the genetic causes of disease and the factors that play a role in disease variability.

2233F

Personal Genetics Education Project: A multifaceted and interdisciplinary approach to public engagement about personal genetics. *M. Gelbart, D. Waring, L. Tomaselli, M. Shea, T. Wu.* Personal Genetics Education Project, Harvard Medical School, Boston, MA.

Through its direct engagement with public audiences, the Personal Genetics Education Project (pgEd. org) at Harvard Medical School experiences on a near-daily basis the growing gap between what genetics is making possible and what many people realize. pgEd sets out to catalyze a spark, striving within as little as five minutes to inspire people’s imaginations about the possibilities and complexities of where genetics is heading. The three overarching goals of pgEd are (1) to decrease fear of, or reluctance to engage on, the topic of genetics, (2) to instill confidence in individuals so that they will feel comfortable asking questions about genetics, and (3) to promote acceptance of diversity. We focus on instilling curiosity, seeking to lower the barriers that inhibit nonscientists from learning more about genetics. We engage audiences through the entertainment industry, an on-going series of Congressional briefings, social media, as well as in high schools and most recently libraries and faith communities. We rely on topics that fascinate people, the ones that they cannot get out of their minds – woolly mammoths, space travel, personal stories, and the stickiest of ethical dilemmas surrounding reproductive technologies, privacy and consent, and the uses of DNA in the criminal justice system. We findthat individuals are captivated by these topics and interested to learn more. We also find that individuals are impressed and grateful that scientists feel a responsibility to engage in these conversations. Though pgEd does not set out to bring people into science, this is a most fortunate unintended consequence.

2234F

Public attitude towards governance of consumer targeted genetic testing and secondary use for research in Japan. *K. Muto[1], A. Nagai[2], Z. Yamagata[2].* 1) The University of Tokyo, Tokyo, Japan; 2) Yamanshi University, Chuo, Japan.

Background: In 2014, several Japanese industries announced the launch of DTC (direct-to-consumer) personal genetic testing services in Japan. The Ministry of Trade and Industry (METI) has encouraged an industrial association, Council for Protection of Individual Genetic Information (CPIGI), to prepare an accreditation system based on their best practice guidelines. On the other hands, the Ministry of Health, Labour and Welfare started to discuss the definitions of "diagnosis" to regulate DTC personal genetic testing services. Currently, the Japanese Diet is revising the Act on Protection of Personal Information in which personal genome data hasn't been covered. However, we still do not have enough data on public attitudes towards these services and regulation. Methods: Cross-sectional and anonymous online surveys were administrated to men and women in Japan in March 2014 and March 2015. Participants of these studies have been extracted so as not to overlap from the survey panel. The questionnaires included questions concerning genetic knowledge, attitudes towards genetic testing and regulations governing DTC personal genetic testing services. We compared the data of these surveys and past relevant dataset. Results: The mean age of the respondents was 45. 7 ± 14. 0 years in 2014 and 45. 8 ± 14. 0 years in 2015. The percentage of the respondents who haven't known the term "genome" was lower than "gene", "DNA" and "iPS Cell". The percentage of the respondents who knew about DTC personal genetic testing services was 32. 9% in 2015, which is higher than in the 2014 survey by 10 points. The percentage who were willing to purchase or who had purchased these services was almost the same level. The results of willingness to undergo 6 types of genetic testing by scientific evidence. More respondents in 2015 (35. 4%) allowed the secondary use of genomic data for medical research than in 2014 (28. 6%). The percentage of respondents who wanted to receive personal results were 34. 3% in 2014 and 33. 6% in 2015. Discussions: The Japanese public showed slightly more positive attitudes towards DTC personal genetic testing services and the secondary use for medical research. However, more than 50% of our respondents couldn't understand the meaning of scientific evidences. There may be a possibility of affecting the attitude about the future of the services.

2235F

Documentary on thalassemia: a public awareness initiative. *I. Panigrahi[1], A. Chaudhary[2].* 1) Pediatrics, PGIMER, Chandigarh, Chandigarh (UT), India; 2) Pathology, DMCH, Darbhanga, Bihar.

Thalassemia is a public health problem in India with carrier rates from 3-17% in several communities especially in north India. The documentary has been made to increase awareness on thalassemia especially for the public and the availability of preventive measures. During a CME on thalassemia at Darbhanga in April this year, this was planned, and assistance was taken for cartoons/puppets also from the Puppet Theatre Chandigarh. This documentary is available on DVD and also incorporates a thalassemia song that describes the pain and problems of thalassemia patients. The draft of the song is available on YouTube also. The short film of around 12 minutes narrates the basics, management and complications in thalassemia in simple lucid language. Photographs, illustrations, videos and puppet characters like animated red cell are used in the film. The film is narrated by a voice over artistein Hindi, a language understood by most Indians, but subtitles are in English, including for the song. Key messages are given at the end of the film.

2236F

Development of an interactive online clinical whole genome sequencing educational and engagement tool. *E. Ramos, S. Waleetorncheepsawat, R. Maile, K. Sherman, L. Nguyen, K. Rhodes, S. Coon, V. M. Raymond.* Illumina, Inc. , San Diego, CA.

The medical, psychosocial, and social promise of clinical whole genome sequencing (cWGS) is sometimes overshadowed by the vast amounts of complex data, difficulty in translating data into actionable information, and the need to communicate information in lay language. Existing websites allow genomic exploration but are neither consumer friendly nor routinely include educational or interactive components. We developed a free, web-based, Understand Your Genome® Community where adults who are presumed to be healthy and who have undergone a cWGS Predisposition Screen can explore their genome using the MyGenome application. Those who have not had cWGS can also join and explore a demonstration genome. The goal is to create an interactive, engaging workspace to increase individual genomic literacy and expand the personal utility of cWGS data. Individuals learn about medical conditions, including clinically significant variants in Mendelian disorders and pharmacogenomic indications identified through their screening and can link to educational web resources. Individuals can access variant information (e. g. allele frequency, gene callability, disease inheritance patterns and amino acid changes). As individuals become aware of other variants of interest through academic or lay literature, they can use the chromosome browser feature to navigate to these areas within their genome and identify their variation at that position. Participant feedback has been positive. Website statistics demonstrate frequent and return visits. Participants overwhelmingly request the ability to follow genomic discoveries, research their variants, share genomic information with family, friends and health care providers, and donate their genome to research. Participants are also interested in networking and identifying individuals within the community with similar genomic findings. The development of a free, online educational community has been received positively by individuals engaged in cWGS. As we continue to explore the utilization of genomic information in a healthy adult population and redefine "value" in terms of genomic information, understanding the motivations for sharing genomic information and the opinions on the key components and features of the this community will be important in maintaining active engagement by members.

2237F

People's attitude toward genetic testing for children and informed consent/assent in Japan. *I. Ishiyama[1], Z. Yamagata[2], J. Minari[3], G. Yoshizawa[3], K. Kato[3].* 1) Tokoha University, Fuji, Shizuoka, Japan; 2) University of Yamanashi, Chuo, Yamanashi, Japan; 3) Osaka University, Suita, Osaka, Japan.

[Purpose] We've been conducting nationwide questionnaire surveys since 2005 to 2014 in order to explore attitude change on genomic studies of the general public in Japan. The aim of this study was to clarify the attitude change of Japanese general public toward genetic testing of disease susceptibilities for children, to assess people's opinion on the children's appropriate age of consent/assent in research collecting human biological specimens, and to examine the factors related to these attitudes. [Methods] In 2005, 2009 and 2014 surveys, peoples (age, 20-69) were selected from the Japanese general population by a stratified two-phase sampling method. They were queried about the following topics by mail: (1) attitude toward genetic testing for disease susceptibilities of children on common diseases which will be able to be prevented or controlled, (2) appropriate age of children to be sought consent/assent by researchers participating research providing biological specimens, (3) awareness of self-condition on health, (4) level of genomic literacy, and (5) demographic and socioeconomic background. Datasets were examined using χ^2test, one-way ANOVA and logistic regression models. [Results]The response rate was 54. 3% (2,171/4,000) in 2005, 52. 5% (2,009/4,000) in 2009, and 56. 4%(1,354/2,400) in 2014. The genetic testing for disease susceptibilities of children for common diseases was favored by 55. 5% people in 2005; 58. 8% in 2009; and 50. 3% in 2014. A difference in the attitudes was observed. Having high level of genomic literacy showed positive relation with the favorable attitude, having a child (or children) actually showed negative relation, and awareness of self-condition on health showed no relation. Concerning literacy and favorable attitude, there was a strong relation in males but a partial relation in females in 2014 survey. In regard to questionnaires of appropriate age of children to be sought consent/assent for providing biological specimens in 2014 survey, 1. 3% people answered 5years, 3. 3%:7years, 15. 1%:10years, 25. 4%:13years, 24. 9%:16years, 12. 9%:18years, 4. 4% answered collecting biological specimens should not be allowed in the cases of minors, and 12. 7% answered I don't know. Peoples with high level of literacy tended to make a response of younger age. Assessing the competence of a child to consent/assent and the best process of seeking consent/assent should be discussed in Japanese society.

2238F

Overcoming Technical Barriers to Incorporate Molecular Structures in Teaching with Molecule World™ on the iPad and iPhone. *TM. Smith, SG. Porter.* Digital World Biology, Seattle, WA.

It is impossible to grasp fundamental concepts of genetic variation without understanding the relationship between sequence, structure, and function. While an abundance of data resources can be used to help students' understand these relationships, they are underutilized because easy-to-use tools are not yet commonplace. Filling the gap between the embarrassment of data riches and practical classroom use requires user-friendly tools, content that demonstrates specific applications with interesting stories, and packages that combine instruction, assessments, and inquiry-based investigations. Finally, in our experience teaching bioinformatics, we have observed that, when given a choice, many students prefer to use their mobile devices instead of desktop computers. Consequently, to acquaint students with bioinformatics, the educational tools and materials need to be operable on these devices. In response to nearly two hundred interviews with K-12 and college teachers and students, we created the Molecule World™ iPad and iPhone apps to display 3D-data from multiple structure databases (MMDB, PDB, and PubChem). Molecule World employs a novel rendering engine that allows us to uniquely highlight chemical properties and sequence orientation. The ability to display and highlight sequences within molecular complexes enables exploration into the relationships between sequence, structure, and function in new ways. In comparison to desktop programs, students and teachers have communicated that Molecule World overcomes the obstacles associated with installing software and understanding complicated user interfaces. Other data, collected in professional development workshops, and bioinformatics courses, along with summer technical camps with high-school students, support the hypothesis that being able to view and simultaneously interact with data improves teaching capabilities and student engagement. Work supported by NSF grant IIP 1315426.

2239F

Teaching meiosis brings together concepts from four different areas. *D. L. Newman[1], K. DeOca[1,2], C. M. Catavero[1], E. Zajicek[1], L. K. Wright[1].* 1) Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, NY; 2) Department of Biology and Marine Science, Jacksonville University, Jacksonville, FL.

Although the topic of meiosis is brought up numerous times during a typical undergraduate biology program, most undergraduates have only a superficial and/or deeply flawed model of this process. Biology education research has been documenting this phenomenon for decades and researchers have identified numerous misconceptions that students hold about meiosis. Many of these findings are presented as isolated areas of confusion, not tied in with holistic conceptual understanding. Based on results from an assessment instrument that we developed and validated using methodology from the DBER literature we have developed a framework that we feel explains the gaps in students' knowledge. Deep analysis of textbooks, interviews, and classroom artifacts revealed that meiosis is neither a single concept nor is it based on a single biological principle. In fact, it brings together several disparate broader topics: chromosome structure, ploidy, homology, and molecular mechanism. We have validated the importance of these themes through alignment with published concept inventories and interviews with biology experts. Using this framework we analyzed 17 college-level biology textbooks and found that information related to these four major themes are often missing, hidden (perhaps obvious to an expert but not a novice), or are found in a chapter that follows the discussion about meiosis. We suggest a new framework for teaching about meiosis based on linking foundational knowledge of these four topics to discussions of cell division. Future work will investigate the effectiveness of this approach.

2240F

Quick Medical Genetics: A YouTube channel for clinical genetics education. *P. M. Boone.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Purpose: Online videos are a potential resource for clinical genetics education. Yet, such videos lag behind textbooks, review articles, and online encyclopedias (e. g. OMIM, GeneReviews) in comprehensiveness and availability. Thus, I sought to establish a collection of clinical genetics lectures on YouTube, the most utilized source of free online videos. **Methods:** A YouTube channel, Quick Medical GeneticsTM (https://www. youtube. com/c/quickmedicalgenetics), was created and nine video lectures were piloted. Videos were short (< 15 min), focused on a single genetic disorder (e. g. ataxia-telangiectasia) or a family of genetic disorders (e. g. disorders of galactose metabolism), and were created with an anticipated audience of both trainees and medical professionals. **Results:** The videos were viewed 1,132 times over a four-month pilot phase. The average view time was 4 min 14 sec, for a total of 4,736 minutes watched. Individuals in 96 countries watched the videos, the top 5 viewer origins being the United States, United Kingdom, India, Saudi Arabia, and Egypt. Thirty-four individuals subscribed to the channel. **Conclusions:** The Quick Medical GeneticsTM YouTube channel is a collection of free, easily accessible educational videos about genetic disorders. In a pilot phase, viewers from dozens of countries watched thousands of minutes of content. In order to be scalable to the large number of known genetic disorders, input from the genetics community will be necessary. Thus, I welcome videos from genetics trainees and professionals; please inquire at quickmedicalgenetics@gmail. com.

2241F

Learning across the lifespan: a team-based approach to genomics education for high school students to practicing professionals. *E. Edelman, W. Barter, J. Corrigan, T. Ingram, L. John, K. LaRue, M. McKernan, W. Nakonechy, K. Reed, C. Wray, T. Litwin.* Genomic Education, The Jackson Laboratory, Bar Harbor, ME & Farmington, CT.

Realizing the promise of precision medicine will require a medical and scientific community conversant in genomics. While efforts exist to educate specific audiences about genetics and genomics, these activities often happen in isolation from one another and may not support the cross-discipline discussion and understanding critical for broad genomic literacy. Furthermore, training programs that focus on a single audience or delivery approach diminish the capacity to leverage successful models and efficiently adapt existing resources. We describe a unique model for lifelong genomics learning, from high school to continuing education. The Jackson Laboratory (JAX) has created an ecosystem in which a multidisciplinary educational team, embedded within a research institute, seeks to empower learners to join the shared quest to improve human health through education, research and collaboration. Under the aegis of JAX Genomic Education (GE), 5 core programs address the needs of diverse learners. (1)The STEM and Undergraduate Education program provides authentic research experiences to students as well as professional development and inquiry-based classroom modules for high school STEM teachers. (2)The Genomics Workforce Consortium delivers professional development for college faculty, expertise in curriculum review for community colleges and research and internships for students. (3)The Pre- and Postdoctoral Education program provides resources for 64 postdoctoral associates and 25 Ph. D. candidates, including a program focused on skill development in communication, management and entrepreneurship to help themsucceed as independent scientists. (4)The Courses and Conferences program leverages the scientific expertise within JAX and delivers courses in medical and mammalian genetics and genomic techniques to scientists. (5)The Clinical and Continuing Education programempowers healthcare professionals to integrate genomics into their clinical practice by providing competency-based CME and clinical tools. GE Online & Digital supports digital collaboration and learning tools in all JAX programs, facilitating access and engagement through scalable online courses. Being housed within the same office and working collaboratively allows JAX GE to leverage internal expertise and learner feedback to develop cross-disciplinary opportunities. This approach to genomics education enables us to better meet the needs of our learners and prepare the workforce for genomic medicine.

2242W

Costs of clinical decision support for genomic tests may be greater than the cost of initial genomic testing. *P. C. Mathias[1], P. Tarczy-Hornoch[2], B. H. Shirts[1].* 1) Laboratory Medicine, University of Washington, Seattle, WA; 2) Biomedical Informatics and Medical Education, University of Washington, Seattle, WA.

For optimal personalized medicine a physician will need to be aware of genetic results from testing performed years earlier, as rare genomic findings may alter clinical care for a fraction of patients. Well-designed active clinical decision support (CDS) alerts within an electronic health record are important for incorporating genomic data into clinical workflows. Via mathematical modeling we evaluated potential costs of design, implementation, and maintenance of CDS for genomic medicine in a variety of scenarios. For our modeling we made a number of assumptions. We assumed the costs of designing, building, and implementing a single genetic CDS rule is about $7. 5K We assumed that CDS rules would need to be in place 20 to 30 years before they would be used for the average individual. CDS rules must be maintained over this time, assumed to be 15-20% of implementation costs. The lifetime costs of maintaining CDS rules and support systems for precision medicine were thus assumed to be 4 to 6 times the cost of initial rule implementation. As an example of one scenario we looked at genomic testing in 2% of the population, and local CDS implementation for 10 actionable variants improving care for 1% of the individuals tested. With assumptions above we found that the cost of communicating genomic information to a physician at the point of care could be about $26K per patient benefiting from the genomic information. If additional variants lead to actionable findings in a higher fraction of patients, efficiently implementing multiple CDS rules will decrease CDS costs per patient receiving benefit. Efficiencies of scale from sharing CDS rule implementation across many institutions or implementation of genomic testing in a greater proportion of patients could reduce per patient genomic CDS costs. We modeled scenarios across a range of assumptions and found that in many scenarios per-patient cost of the electronic communication of genetic information using CDS was greater than the cost of initial genomic testing and annotation. Only in scenarios that include mobility of genomic data across institutions, cooperation between institutions for CDS, and widespread genomic testing will precision medicine become cost effective. Our modeling suggests additional inter-institutional efforts to build efficient, scalable systems for communicating genetic test results and actionable interpretations at the point of care are needed to decrease cost of genomic decision support.

2243W

A cost effective in house method for HIV genotyping – circumventing the use of costly kits for developing countries. *M. Mitne-Neto, M. G. S. Barth, C. M. Moreira, P. R. Sacramento, C. P. Reys, M. T. dos Santos, A. M. Fraga.* R&D, Fleruy Group, São Paulo, Brazil.

Statement of purpose: More than two hundred thousand Brazilians carry the Human Immunodeficiency Virus (HIV). Brazil offers freely treatment for infected patients, which costs US$427 million per year for the Federal government. The proper patient management requires HIV genotyping, aiming to identify mutations promoting resistance to each of the effective drugs. The Trugene HIV genotyping assay (Siemens) was one of the main assays available in the market, however, its worldwide discontinuity surprised private laboratories and those working for the Brazilian healthcare system. Methods: In order to circumvent this situation we developed an in house method that allows the sequencing of the Protease and Reverse Transcriptase regions and compared its performance with the Siemens assay. Primers for One-Step RT-PCR were designed to amplify a 1300bp region, followed by a nested PCR and traditional Sanger sequencing. Resulting sequences were submitted to the HIV Drug Resistance Database – Stanford University. In order to validate the method fourteen samples (subtypes B / C / F) that were previously genotyped by the Trugene system were evaluated in the new workflow. Results: Susceptibility profile, classified as High Level of Resistance, Susceptible, Potential low-level resistance and Low level resistance showed 100% concordance on all samples, through the 20 evaluated drugs, between the two platforms. Main and minor resistance mutations to protease inhibitors, nucleoside reverse transcriptase inhibitors and non-nucleoside reverse transcriptase inhibitors also showed 100% concordance between the two platforms over all the 14 samples. The lower limit of detection was established as 2000 copies/mL for the in house system by the evaluation of 5 separated aliquots of the same sample. The implementation of this methodology has a great economy impact, since it reduces reagents costs by 40% and the analyst hands on time by 50%. Additionally, our assay design allows the immediate transfer of amplified regions to next generation sequencing, which is still under evaluation, in order to identify mutations that are under the 20% allele burden cutoff. In summary, we developed a cost effective in house method for HIV genotyping, which is cheaper and faster than the previously offered kits in the market. This data is of special interest of developing countries due to the highly costs related to HIV proper treatment.

2244W

Attitudes towards receiving genetic information on susceptibility to the effects of arsenic exposure in rural Bangladesh. *B. Pierce[1], M. Eunus[2], M. Rahman[2], B. Barmon[2], A. Ahmed[2], R. Hasan[2], T. Islam[2], H. Ahsan[1].* 1) Public Health Sciences, University of Chicago, Chicago, IL; 2) U Chicago Research Bangladesh Ltd. , Dhaka, Bangladesh.

Exposure to arsenic, a known carcinogen, is a serious global health issue that has been linked to risk for cancer, cardiovascular conditions, and overall mortality. In Bangladesh, >50 million people consume water from naturally-contaminated wells with arsenic exceeding >10 µg/L. Our group and others have demonstrated that genetic variation near the arsenic methytransferase gene (*AS3MT,* 10q24. 32) is associated with arsenic metabolism efficiency (measured as urinary arsenic metabolites) and is the major common determinant of arsenic metabolism in the genome. Individuals who are slow metabolizers due to 10q24. 32 genotype have higher risk for arsenical skin lesions -- the classical sign of arsenic toxicity. In light of this emerging knowledge and its potential for identifying high-risk individuals, we conducted a survey of 200 participants in the Health Effects of Arsenic Longitudinal Study (HEALS) to assess attitudes towards receiving genetic information related to arsenic toxicity risk. Returning genetic information has the potential to increase the efficacy of an ongoing exposure-reduction program in HEALS by further motivating highly-susceptibility individuals to avoid contaminated water sources. Among 200 consecutive participants visiting our primary care clinic who agreed to participate in this study, 100% were interested in receiving genetic information, and 91% indicated that such information would motivate them to reduce their exposure. Over 99% of respondents indicated that they would disclose their genotype status to family, and 59% reported they would be comfortable disclosing their status to non-family members. Individuals with no formal education were less likely to be concerned about arsenic exposure than individuals with some education (p=0. 02), but low- and high-education groups reported similar levels of motivation to reduce exposure in response to genetic information. To our knowledge, this is the first assessment of attitudes towards genetic testing in Bangladesh, and the first such assessment related to susceptibility to arsenic toxicity. Our results indicate that a) individuals living in arsenic-exposed regions are interested to know if they are at increased risk and b) sharing genetic information with high-risk individuals may provide additional motivation to reduce exposure. Additional research is needed to estimate the impact of returning genetic results on health behaviors and outcomes and to assess potential harms.

2245W

Cost-Benefit Analysis of Genomic Testing Vs. Standard Genetic Testing in Inpatient Population. *A. Khromykh[1], B. Solomon[1,3], F. Moeckel[1], D. Bodian[1], R. Baveja[2,3], J. Vockley[1], J. Niederhuber[1].* 1) Inova Translation Medicine Institute, Inova Fairfax Hospital, Falls Church, VA; 2) Fairfax Neonatal Associates at Inova Hospital for Children, Falls Church, VA,; 3) Inova Children's Hospital, Falls Church, VA,.

PurposeAs the cost of WGS is rapidly decreasing and analytic capabilities evolve, WGS and related modalities are becoming increasingly usable. In order to harness the potential of clinically-based WGS, we are conducting a research study to detect novel genomic etiologies as well as assessing the application of WGS as a diagnostic tool in real-time clinical practice. **Methods**We utilize trio-based WGS, RNA expression, methylation, and miRNA characterization in order to investigate potential molecular etiologies of the congenital disease. Genetic workup data was obtained through thorough review of probands' electronic medical records. Cost averaging was calculated based on actual reimbursements to the hospital (these negotiated rates are ~40-60% lower than market price). **Results**The average cost of standard genetic/etiologic laboratory tests was $3,500 per patient. We were not only able bioinformatically identify the causative genetic/genomic mutations in multiple probands in our cohort population utilizing WGS, but also that cost of clinical genetic/etiological testing was 2-3 times greater than research-based cost of sequencing the whole genome. **Conclusion**Our preliminary data confirms that trio-based WGS can provide answers to patients and families with undiagnosed disorders, as well as the novel finding that research-based WGS may already be considered a cost-effective alternative to the standard genetic work-up. Further, WGS can act as an immediate diagnostic tool, and can also provide additional clinical utility related to a wide range of genomic-related health information.

2246W

Annotating rare diseases with HPO terms: a new Orphanet service to refine phenotype-genotype correlations. *A. Olry[1], C. Prévot[1], C. Jaramillo[1], C. Gonthier[1], S. Janmaat[1], V. Lanneau[1], S. Gérard[1], M. Hanauer[1], PN. Robinson[2], A. Rath[1].* 1) INSERM, US14 - ORPHANET, Paris, France; 2) Institute for Medical Genetics, Charité-Universitätsmedizin Berlin, 13353 Berlin, GERMANY.

Detailed phenotype data, combined with ever-increasing amounts of genomic data, have an enormous potential to accelerate the identification of disease-associated genes and clinically actionable variants. The harmonization of phenomic information including disorders and phenotype traits that are stored in different systems (patient records, databases, registries) in a non-standardized way, is a cornerstone for the data integration needed for translational research. In order to contribute to the harmonization of phenotype and rare disease (RD) data, two resources that are now considered as standard are being integrated: the Orphanet RD nomenclature and its ontological version ORDO, and the Human Phenotype Ontology (HPO), by annotating the Orphanet nomenclature with HPO terms. The interoperable phenotype core terminology produced by the International Consortium for Human Phenotype Terminology (ICHPT) was used to align the Orphanet thesaurus of clinical signs, to which RD were already annotated, to HPO terms. The annotations were manually reviewed and refined, and further annotation with HPO was conducted. The 2689 RDs initially annotated with the clinical signs of the Orphanet thesaurus are now annotated with HPO. Additional information was also produced:- Estimated frequency of the annotated HPO terms in the patient population (very frequent, frequent and occasional),- Whether the annotated HPO term is a major diagnostic criterion or a pathognomonic feature of the RD. This work improves both of these standard resources. HPO annotations increased considerably the phenotypic description of RD, with 4,508 HPO terms now used compared to the 1,273 clinical signs previously used in the Orphanet thesaurus of signs and symptoms. On the other hand, HPO will be enriched as around 370 missing phenotypic terms have already been identified. Furthermore, Orphanet provides genes associated to RDs annotated with HPO terms, opening new avenues for a better interpretation of phenotype-genotype correlations. Finally, interfacing these two standard resources will allow the integration of clinically well-defined RD in repositories already using HPO and/or other terminologies already mapped with the Orphanet nomenclature, including OMIM. These annotations will be freely available at Orphadata (www. orphadata. org). This pilot study will be expanded in the future by the collaborative development of tools for medical biocuration as well as interoperable RD resources for the community.

2247W

Thalassemia in Egypt: Paving the way for prevention and cure. *G. Y. El-Kamah[1], K. Amr[2], A. AbdelNeam[2], E. Bayoumy[2], M. Ghandour[1], H. Ahmed[1].* 1) Clinical Genetics department, Human Genetics and Genome Research division, National Research Centre, Cairo, Egypt; 2) medical molecular genetics department, Human Genetics and Genome Research division, National Research Centre, Cairo, Egypt.

Introduction: Beta thalassemia is one of the most common human genetic disorders, and represents a major public health problem**.** In Egypt, beta-thalassemia mutations occur at high frequency with a carrier incidence of about 9-10%. **Objective: D**efining the spectrum of mutations among Egyptian beta-thalassemia patients. Studying whether suggested genetic determinants can predict phenotypic severity and its impact on genetic counseling as an important step in providing proper counseling & disease control. **Methodology:** This is a prospective cross-sectional study performed from 2005 till 2015 in the Hereditary Blood Disorders Clinic. 500 Egyptian beta thalassemia cases were included. Clinical data were recorded followed by molecular analysis & mutations detection. **Results:** Twenty-six mutations are identified till date. The five most prevalent mutations were as previously reported IVS-I-110 (G>A), IVS-I-6 (T>C), IVS-I-1 (G>A), IVS II-745 (C->G) and IVS II-848 (C->G) accounting for more than 80% of the independent characterized chromosomes. The five mutations responsible for 80% of our cases are all Mediterranean in origin except IVS II-848, which is considered Egyptian. Comparing Egyptian mutations' spectrum to previous reports 21 other (novel & known) mutations were characterized. Type of beta-thalassemia mutation &/or determinants was not persistently correlating with the phenotypic outcome. No intra-familial heterogeneity was detected among our included family members helping genetic & prenatal counseling. **Conclusion:** the study helped in providing better counseling however**,** further mutations characterization and genetic modifiers studies are needed for the clarification of phenotypic heterogeneity. Also, immunological determinants, their prognostic values & psychosocial impact of the disease on thalassemia patients and their families are in progress.

2248W

Views of patients regarding universal Lynch Syndrome screening. *T. Hyams[1], D. Bowen[1], S. Shiovitz[2,3], M. Miller[1], W. M. Grady[3], W. Burke[1].* 1) Department of Bioethics and Humanities, University of Washington, Seattle , WA; 2) University of Washington, Division of Oncology, Seattle, WA; 3) Fred Hutchinson Cancer Research Center, Clinical Research Division, Seattle, WA.

Approximately 2-5% of the annual 150,000 colorectal cancer (CRC) cases diagnosed in the United States are attributable to Lynch Syndrome (LS), a familial cancer syndrome. . The lifetime CRC risk in individuals with LS is estimated to be ~40- 82%, which is much higher than the population's CRC risk of 5%. In 2009, the Evaluation of Genomic Applications in Practice and Prevention working group recommended universal screening for newly diagnosed CRC patients using microsatellite instability (MSI) and mismatch repair immunohistochemistry (IHC) to provide appropriate management for identified patients and their family members. This would include genetic testing and, if appropriate, offering early cancer detection interventions. . The purpose of this abstract is to describe the experiences of CRC patients with universal screening. We approached and interviewed 10 patients diagnosed with CRC in the Multidisciplinary CRC Clinic at the Seattle Cancer Care Alliance, a collaborative outpatient facility of the University of Washington and the Fred Hutchinson Cancer Research Center. The purpose of the interviews was to explore patient experiences with universal testing for MSI and IHC and family interactions about CRC risk and testing. We sent a letter, and then called patients with an invitation to participate in the telephone interview. Interviews lasted 60 minutes and were recorded and then transcribed for analysis. In general, patients did not remember providing consent for MSI/IHC testing, nor did they remember receiving results. No patients objected to being tested, even those that did not remember the process. Trusting the provider to make important recommendations was a theme in the interviews, as was being overwhelmed with procedures and activities after diagnosis. Few patients had talked to their family members about familial cancer risks since the diagnosis. Most patients reported that they would talk to family members if it were important to the family members' health, and most wanted assistance with family communication. All patients were offered contact with a genetic counselor to answer questions after the interview was over. Identification of key themes in discussion the LS testing process has important implications for future universal testing. Our evaluation has helped to identify ways to assist patients and families with issues surrounding universal LS screening.

2249W

A SMARTer way to set genomic test reports on FHIR. *M. S. Williams[1], J. A. Billet[2], G. J. Moore[2].* 1) Genomic Medicine Institute Geisinger Health System, Danville, PA; 2) Institute for Advance Applications, Geisinger Health System, Danville, PA.

As genome sequencing moves into the clinic, innovative approaches to present the results to patients and providers are needed given the complexity and dynamic nature of the results. More control by the patients is desirable as the genomic data has persistent relevance to the patients' health care. We have worked with patients and families to develop a patient-facing genomic test report. In this paper we describe implementation of an interactive patient facing genomic testing in the electronic health record (EHR). Most EHRs do not have the capability to support the advanced functionality of our report. We utilized an internally developed application, COMPASSTM, to develop the dynamic genomic test report. COMPASSTM employs its own application program interface (API) and other service-based architectures to seamlessly integrate into the EHR. This architecture is being utilized to put COMPASSTM on the path of an emerging open standards approach for sharing healthcare information called Fast Interoperable Health Resources (FHIR®). In order to present information to patients and family within the patient portal tethered to the EHR, an API called SMART will be utilized upon industry adoption. SMART on FHIR is an emerging open standards-based technology stack that allows integration of clinical, genomic and related data for both the patient and provider report. The genomic report has been integrated with our Epic EHR and is being tested in a prospective randomized comparative effectiveness trial. The primary outcome of the study is the impact of the dynamic and interactive patient and provider reports on communication, patient and provider satisfaction and patient engagement. Feedback on the functionality and usability of the report is used for iterative improvement. This approach to reporting genomic reports using a standards-based approach has the potential to be generalizable across certified EHRs and has the potential to transform communication of genomic results.

2250W

Evaluation and Optimization of a Universal Lynch Syndrome Screening Program at Geisinger Health System. *A. Kulchak Rahm, N. S. Kip, S. Guha, A. Fan, H. G. Kaspar, J. L. Williams, R. Gogoi, Z. Chen, D. Davis, M. S. Williams.* Genomic Medicine, Geisinger Health System, Danville, PA.

Background: Universal Lynch Syndrome screening (ULS), a genomics objective of Healthy People 2020, is defined as tumor screening of all individuals newly diagnosed with colorectal (CRC) or endometrial cancer (EC) and has been variably implemented in multiple health systems. ULS was instituted at Geisinger Health System in 2009 for CRC and optimized to include immunohistochemical (IHC) screening of all CRC resected tumors, with reflexive *BRAF* and *MLH1* promoter hypermethylation (PHM) testing through pathologists in May, 2014. EC tumor screening started in 2012, but was limited to age ≤60 or age >60 with specific tumor histology; without reflexive PHM testing. **Methods**: Two genetic counselors and a molecular pathologist evaluated all EC and CRC patients with IHC testing since 2012. Results and care gaps were presented to pathologists, genetics services, gynecologic oncologists and ethics. **Results**: Since optimization, 132 CRC biopsies and resections were sent to pathology to be screened for mismatch repair proteins (MMRs) by IHC, and for *KRAS, BRAF* mutations by PCR. 17 (13%) patients had MLH1/PMS2 loss and 1 (0. 8%) had PMS2 loss. Of those with MLH1 loss, 15 (88%) were reflexed to *BRAF* and PHM testing, with 3 (18%) found to be *BRAF* negative. Of those, 2 (66%) had PHM; the patient without PHM was referred to genetics for evaluation, as was the patient with PMS2 loss by IHC. For EC, 161 EC patients were screened by IHC (n=152 ≤60 and n=19 >60) since 2012; of whom 30 (18%) had loss of MLH1/PMS2 and 11 (7%) had MSH2 or MSH6 loss. Only 10 patients with MLH1 loss (33%) had PHM testing (*BRAF* testing is not indicated in EC). Of the 20 remaining EC patients, 6 (30%) were referred to genetics prior to PHM testing and 14 (70%) had no additional follow-up. Of patients with MSH2 or MSH6 loss, 10 (91%) were referred to genetics. **Conclusion**: After comparing the optimized CRC program and the limited EC program, we determined that reflex testing by pathologists appropriately specified patients at risk for LS and reduced over-referral to genetics. It was also determined that the limited EC tumor protocol was resulting in incomplete screening and over-referral to genetics. Engagement of multiple stakeholders and clarification of policy goals resulted in agreement to add EC tumor screening to the existing ULS program with reflexive PHM testing by pathologists to improve LS patient identification and facilitate appropriate care.

2251W

Qualitative Assessment of the Veterans Health Administration Telegenomic Clinic Implementation. *B. Lerner[1], N. McIntosh[1], M. Meterko[1], S. Wiltsey-Stirman[1], V. Venne[2].* 1) Center for Healthcare Organization & Implementation Research , Veterans Health Administration, Boston, MA; 2) Genomic Medicine Service, VA Salt Lake City Health Care System, Veterans Health Administration, Salt Lake City, UT.

ObjectivesSince 2010 the Veterans Health Administration's (VA) Genomic Medicine Service (GMS), based in Salt Lake City, has been working to create a VA-wide Telehealth Genomics program to increase Veteran access to genomic services. Participating VA facilities must complete a complex process that requires engaging several clinical and administrative departments and giving GMS providers laboratory ordering privileges and access to patient records. Experience with the first 40 adopters showed significant variation in a facility's ability to expeditiously implement this program. The goal of this study was to evaluate the implementation process and identify best practices to facilitate more efficient adoption of Telegenomics as the program expands to the 153 VA facilities and outlying clinic locations. **Methods**We ranked 40 facilities on implementation efficiency based on subjective and objective criteria, and selected three "high" and two "low" facilities for this formative evaluation study. We interviewed personnel involved in the implementation including project leaders (Telehealth Coordinators (TCs)), technical staff, chief medical officers, pathology directors, and referring providers at each facility about their implementation experiences. Several GMS staff were also interviewed. We analyzed data using deductive coding based on the Implementation Science conceptual framework, Consolidated Framework for Implementation Research (CFIR). Concepts and themes related to implementation facilitators and barriers were identified, as were recommendations to improve the process. **Results**We conducted 43 interviews. At the efficient facilities, TCs had strong project management skills and understood the sequence of steps and roles of people needed for clinic implementation. They also relied heavily on well-established relationships and good communication with the relevant departments. Also, clinicians and the laboratory director at these facilities were advocates for Telegenomics. TCs from inefficient facilities had limited prior inter-department relationships and communication skills, and/or lacked an overarching understanding of the implementation process. **Conclusions**The efficiency of implementing Telegenomics may be improved by providing TCs tools to better manage the implementation process, including guidance for the development of inter-departmental relationships and channels of communication.

2252W

Genomics and personalized medicine in primary care: a professional engagement study using a novel, theory-informed workshop approach. *B. J. Wilson[1], C. A. Catley[2], J. Little[1], S. G. Nicholls[1], J. C. Carroll[3], H. Etchegary[4], M. Taljaard[5], CIHR Emerging Team in Genomics in Screening.* 1) School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, ON, Canada; 2) Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada; 3) Department of Family & Community Medicine, Mount Sinai Hospital, University of Toronto, ON, Canada; 4) Craig L. Dobbin Centre for Genetics, Memorial University, NL, Canada; 5) Centre for Practice Changing Research, Ottawa Hospital Research Institute, Ottawa, ON, Canada.

Background:Studies consistently suggest that primary care professionals see great potential in genomics and personalized medicine, but hold reservations about its current utility for their own practice. These observations have changed little in fifteen years; methods which go beyond surveys and common qualitative methods are required if we are to capture insights useful for designing effective primary care implementation strategies for such emerging technologies. We designed a workshop approach, adaptable for online implementation, to address this concern. Objective:To evaluate the utility of a deliberative workshop approach as a method to identify issues that primary care physicians and nurses perceive as salient to the successful adoption and implementation of genomics and personalized medicine in practice. Methods:We developed a prototype professional engagement approach informed by Kirkpatrick's educational hierarchy (relating to educational goals about genomics as a new technology) and the Theoretical Domains Framework (relating to the comprehensive behavioral aspects of using a new technology). It comprised (a) a standardized background on genomics and personalized medicine, (b) hypothetical case studies chosen to reflect the congruence between some potential uses of genomics and current primary care activities (e. g. , disease risk assessment), (c) pauses for facilitated deliberation or reflection, and (d) embedded self-completion surveys. We recruited family physicians and primary care nurses working in Ontario, Canada, to participate (in-person or online formats). We conducted descriptive analyses of the quantitative and qualitative data. Results:Twenty physicians and 44 nurses participated (34 in-person, 30 online). Baseline analyses showed no differences by professional group or workshop format in perceptions of genomics or assessments of clinical utility. Pre-post quantitative comparisons indicated shifts to more positive judgments of likely utility for personalized medicine in general, and for the systematic use of family history information (one of the case studies), but not for genomic profiling for CRC risk or type 1 diabetes risk prediction (the other case studies). Dominant themes in qualitative analyses indicated both positive and negative beliefs about the consequences of using genomics in practice, and perceived challenges with integration into practice. We noted more extensive qualitative data from online than in-person participants.

2253W

Optimizing the Design of a Population-Based Reflex Testing Program for Lynch Syndrome in Ontario: Health Care Providers' Perspectives. *Y. Bombard[1, 2], L. Rozmovits[3], A. Sorvari[2], C. Daly[2], J. C. Carroll[1,4], E. Kennedy[1,4], S. Gallinger[1,4], L. Rabeneck[5], N. N. Baxter[1,2].* 1) University of Toronto, Toronto, Ontario, Canada; 2) St Michael's Hospital, Toronto, Ontario, Canada; 3) Independent Qualitative Researcher, Toronto, Ontario, Canada; 4) Mount Sinai Hospital, Toronto, Ontario, Canada; 5) Cancer Care Ontario, Toronto, Ontario, Canada.

Background: Lynch syndrome (LS) is a relatively common genetic syndrome, which increases the risk of colorectal and endometrial cancers. Universal IHC or MSI testing (reflex testing) of colorectal tumours can identify at-risk relatives but there is variation in access and debate about providing informed consent or an opt-out. We explored health care providers' (HCPs) experiences managing LS to inform the design of Ontario's reflex testing program. **Methods:** We conducted qualitative interviews with Ontario HCPs with experience managing LS patients or an interest in colorectal cancer (CRC) screening. Participants included CRC surgeons (n=6), general surgeons (n=5), genetic counsellors (GC;n=7), medical oncologists (n=2), primary care providers (n=3) and a gastroenterologist. Interviews focused on information provision, opt-out, result notification and support for subsequent options. Participants constituted a convenience sample recruited through referrals from the research team supplemented by snowball sampling. Qualitative data were analyzed using content analysis and constant comparison techniques. **Results:** HCPs experiences and perceptions of current practice pointed to inconsistent management of suspected LS patients, variation in IHC testing practice and challenges in GC service delivery. Limited resources, active treatment and complex family history questionnaires presented barriers to accessing GC. Providers supported a reflex testing program due to the lack of coordinated care, under-ascertainment of LS patients and opportunity to standardize the increasing use of genomic tests in practice. While most supported an opt-out, two positions emerged as when to provide an opt-out. Most endorsed an opt-out after reflex testing because they felt IHC testing is akin to other pathology, which is not optional; but some preferred the opt-out before testing out of concern for patients experiencing distress, insurance discrimination or a diagnostic odyssey that may turn out to be inconclusive. **Conclusions:** Current practice supports the need for a reflex testing program, which can reduce inconsistencies in referrals and improve coordinated care for LS. Results highlight opportunities to optimize delivery of care, which include increased genetics resources and provider awareness of LS. However, how to support meaningful information provision to enable an opt-out without jeopardizing uptake of screeningand its public health benefit remains a larger policy challenge.

2254W

Cohort profile: Taiwan Biobank (TWB) for the future of next generations. *C. N. Hsiung[1,2], C. Y. Shen[1,2].* 1) Taiwan Biobank, Academia Sinica, Taipei Taiwan; 2) Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.

To understand the relationship between genetics, environmental exposure, diet, and the etiology/progression of disease, the Taiwan Biobank is establishing a scientific infrastructure accessible to biomedical researchers. Through the recruitment and follow-up of a cohort of 200,000 individuals from the general population with no history of cancer and another of 100,000 patients with chronic diseases of public health importance from medical centers, the Taiwan Biobank aims to improve the health of future generations and facilitate genomic research in the post-genomic era in Taiwan. Currently, more than 38,000 participants from different parts of Taiwan had been recruited, and more than 700,000 of biospecimens, including blood, urine and DNA, had been collected. The structured questionnaire had been administrated to each participant to collect information regarding risk factors, dietary patterns, life style, and family history of diseases. Physical examinations, hematology, and biochemistry measurements were measured. Whole-genome genotyping information of 7,203 individuals using TWBv1. 0 chip (653,291 SNPs, specifically for the Han Chinese in Taiwan) were obtained. Whole Genome Sequencing of 1,000 individuals had been completed. By using these health-related information and the results of genome-wide studies, a preliminary prediction model to type 2 diabetes was proposed with good accuracy. These information and specimen have been released for application in the nation. The rationale for this huge and challenging effort is the development of personalized medicine, in which the progressive elucidation of risk factors and the molecular pathogenesis of disease will both improve disease prevention and facilitate therapy development at the individual level.

2255W

Benefits and Limitations of a Multidisciplinary Approach to Individualized Management of Cornelia de Lange Syndrome and Related Diagnoses. *K. January[1, 5], S. Noon[1], K. Loomes[2], A. Harrington[3], M. Deardorff[1, 6], M. Pipan[4, 6], L. Conway[5], I. Krantz[1, 6].* 1) Division of Human Genetics, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Gastroenterology, Hepatology and Nutrition, The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Division of Physical Therapy, The Children's Hospital of Philadelphia, Philadelphia, PA; 4) Division of Child Development, The Children's Hospital of Philadelphia, Philadelphia, PA; 5) Arcadia University Genetic Counseling Program, Glenside, PA; 6) The Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA.

Given the clinical complexities of Cornelia de Lange Syndrome (CdLS), the Center for CdLS and Related Diagnoses at The Children's Hospital of Philadelphia (CHOP) and The Multidisciplinary Clinic for Adolescents and Adults at Greater Baltimore Medical Center (GBMC) were established to develop a comprehensive approach to clinical management and research issues relevant to CdLS. Little work has been done to evaluate the general utility of a multispecialty approach to patient care. Previous research demonstrates several advantages and disadvantages of multispecialty care. This research aims to better understand the benefits and limitations of a multidisciplinary clinic setting for individuals with CdLS and related diagnoses. Parents of children with CdLS and related diagnoses who have visited a multidisciplinary clinic (N=52) and who have not visited a multidisciplinary clinic (N=69) were surveyed to investigate their attitudes. 90. 0% of multispecialty clinic attendees indicated a preference for multidisciplinary care. However, some respondents cited a need for additional clinic services including more opportunity to meet with other specialists (N=20), such as behavioral health, and increased information about research studies (N=15). Travel distance and expenses often prevented families' multidisciplinary clinic attendance (N=41 and N=35 respectively). Despite identified limitations, these findings contribute to the evidence demonstrating the utility of a multispecialty approach to patient care. This approach ultimately has the potential to not just improve healthcare for individuals with CdLS but for those with medically complex diagnoses in general.

2256W

Integrating genomic medicine across a healthcare system. *S. L. Hill, R. Cashman, S. Chapman, L. Quayle, J. Fisher, JP. Maytum, J. Stewart.* NHS England, London, UK.

England has embarked on an ambitious programme to integrate genomic medicine across its integrated National Health Service (NHS) through the creation of a network of NHS Genomic Medicine Centres (GMCs). This programme has been elicited by the 100,000 Genomes Project – the nationwide initiative to sequence 100,000 genomes from patients & families with rare disease and cancer. NHS England committed to putting in place arrangements to enable the acquisition of samples and accompanying clinical data of sufficient quality to enable whole genome sequencing under the contractual arrangements for the Project. The GMCs work in partnership with national partners Genomics England (a company owned by the UK Department of Health, set up to deliver the 100,000 Genomes Project) and Illumina (the company Genomics England procured to deliver the sequencing and support the analysis of the resulting genome). A transparent and rigorous procurement exercise was conducted in 2014 to commission and designate the first ever Genomic Medicine Centres. This was underpinned by a common contract and service specification which sets out in detail the responsibilities and operation of all of the organisations delivering genomic services. The service specification detailed every element of the sample and data pipeline in participants with rare disease and cancer (from acquisition to validation of whole genome sequence findings to feedback to clinical teams and participants) and provides not just the contractual basis for service delivery but a baseline for measuring improvement and the quality of outcomes. The contract, service specification and networked structure for the GMCs serves as a catalyst to a broader transformation programme to introduce genomic technologies into clinical care across the NHS in England in a systematic and sustainable way for public and patient benefit, to support scientific endeavour and diagnostic delivery and industry collaborations. This approach has brought in a new era of innovation and co-operation between healthcare organisations and is already showing good results. NHS GMCs have adopted innovative patient and public co-production to design their genomics services, created genomics multi-disciplinary teams spanning a multitude of specialisms, driven creative workforce training and forward-thinking approaches to informatics.

2257W

Medical Transition Checklists for Mobile Applications. *J. A. Dixon[1], S. M. Dixon[2], J. Dehlinger[1].* 1) Department of Computer and Information Sciences, Towson University, Towson, MD; 2) Department of Pediatrics University of Maryland School of Medicine, Baltimore, MD.

Supervised medical transition, usually under the care of a genetics team, is now considered to be a best practice in the management of individuals with a chronic disease. In response, many states and organizations have created paper-based checklist guidelines to help individuals navigate this process. However, these paper-based systems are often inconsistent and incomplete. Individuals undergoing medical transition today are digital natives and would benefit from a transition process in a medium they are familiar with. As evidence of this, an online survey was posted to five online support groups for individuals with chronic conditions where participants were asked whether they would consider using a web-based application that would assist with medical transition. Nearly half of the respondents (49. 18%; n=30/61) reported that they would consider using a web-based application to help with the medical transition process.

To address this need, this work describes a fully-functional prototype mobile application that was developed to replace, and improve upon, the existing paper-based medical transition processes. The design of the application included: 1. examining existing literature to develop a comprehensive task and skills list based on medical transition guidelines; 2. analyzing six publically available, paper-based transition models to develop a common, core set of recurring themes and concepts; and, 3. verifying the tasks, concepts and categorized thematic skill sets through semi-structured interviews with three clinical geneticists for critique, completeness, organization and medical relevancy. The resulting mobile application incorporates these design considerations to provide a comprehensive list of tasks and important information to understand as well as an organized approach for how to best achieve them. Specifically, the mobile application provides dedicated functionality for documenting patient data, insurance information, healthcare team information, medical conditions information, medication/pharmacy/medical equipment information, network of support, care management, and transition assessment and completion checklists. The application is undergoing extensive usability testing with initial positive results.

This work is part of a larger effort to enable a cost effective and efficient medical transition solution that empowers individuals going through the medical transition process.

2258W

Assessing provider barriers to implementation of genetic testing for common disease risk in the GUARDD (Genetic testing to Understand and Address Renal Disease Disparities) study. *N. S. Abul-Husn[1,2], M. A. Ramos[3], R. Negron[4], M. Rodriguez[3], T. Sabin[3], R. E. Zinberg[2], K. Fei[3], D. Hauser[5,6], N. Calman[5,6], C. R. Horowitz[3], E. P. Bottinger[1].* 1) The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Yale Institute for Network Science, Yale University, New Haven, CT; 5) The Institute for Family Health, New York, NY; 6) Department of Family Medicine and Community Health, Icahn School of Medicine at Mount Sinai, New York, NY.

Although healthcare providers are enthusiastic about the potential for genomics to improve patient care, certain barriers may prevent them from effectively adopting genomic medicine. We intend to identify these barriers, to help address practical challenges to implementation and enhance uptake of this emerging field. The GUARDD (**G**enetic testing to **U**nderstand and **A**ddress **R**enal **D**isease **D**isparities) study, a project of the NHGRI IGNITE (**I**mplementing **G**e**N**omics **I**n prac**T**ic**E**) Network, integrates *APOL1* genotype and associated kidney disease risk into the primary care management of hypertension in African ancestry patients. A survey was developed incorporating established constructs of effective implementation from the Consolidated Framework for Implementation Research, including knowledge and beliefs, self-efficacy, access to information, and perceived advantage of implementation. This was distributed to primary care providers practicing in three federally qualified community health centers and two academic primary care facilities in New York City immediately prior to the study onset. Data from the first 253 providers (63% residents, 31% attendings or fellows, 6% physician's assistants or nurse practitioners) revealed that the majority of providers (68%) agreed that genetic testing for common diseases was clinically useful. Over a third (38%) felt that their training had prepared them to work with patients at high risk for genetic conditions. A quarter (27%) believed that *APOL1* genetic testing would be important for patient care, but only 7% were confident in their ability to use *APOL1* genetic test results. Few providers (22%) felt that they could find reliable information to communicate *APOL1* genetic risk while caring for patients. The majority thought it would be helpful to have patient handouts (75%) and information about managing high-risk patients (77%) available through the electronic health record. These findings suggest that, despite being optimistic about genetic testing for common disease risk, providers have a deficit in their familiarity and confidence with utilizing genetic risk information in the primary care setting. This indicates that continuing genomic education, readily accessible tools to help communicate genetic risk at the point of care, and specific guidelines for the management of high-risk patients are needed. These data highlight some of the barriers hindering genomic medicine uptake and will inform approaches for improved implementation.

2259W

Next-generation sequencing identifies miRNAs in peripheral blood of African Americans related to metabolic syndrome and abdominal obesity. *S. Y. Gebreab, G. Amadou, P. Riestra, R. J. Khan, S. K. Sen, A. R. Davis, S. K. Davis, G. H. Gibbons.* Metabolic, Cardiovascular and Inflammatory Disease, National Human Genome Research Institute (NHGRI), NIH, Bethesda, MD.

Metabolic syndrome (MetS) is a major public health concern in the African American population due to high prevalence of abdominal obesity, hypertension, diabetes mellitus, and hyperglycemia in this population. Furthermore, MetS is attributed to increased risk of cardiovascular disease, including coronary heart disease and stroke in African Americans. Although MetS is strongly influenced by lifestyle and genetic factors, little is known about the molecular underpinnings of MetS in African Americans. MicroRNAs (miRNAs) mediate posttranscriptional gene expression and play an important role in maintaining metabolic homeostasis; hence, it is possible that they contribute to MetS physiopathology and its components. In the present study, we identify differentially expressed miRNAs in peripheral blood from African Americans with MetS and abdominal obesity using massively parallel sequencing of miRNA (miRNA-Seq). We analyzed miRNAs from 110 African Americans (15 with MetS vs. 95 without MetS subjects, and 57 with obese vs. 53 normal subjects) obtained from the Minority Health Genomics and Translational Research Bio-Repository Database (MH-GRID) study. After adjusting for age, sex, smoking, socioeconomic status and alcohol consumption, we identified a total of seven miRNAs: hsa-miR-378i, hsa-miR-483, hsa-miR-193b, hsa-mir-215, let-7b, hsa-mir-874, and hsa-mir-22 that were significantly up-or down-regulated in MetS and abdominal obesity. These miRNAs were previously identified in animal and human studies linked to human adipogenses and MetS. Although our study is preliminary, our results demonstrate that miRNA-Seq has the potential to help elucidate the underlying molecular mechanisms of MetS and abdominal obesity and to identify novel biomarkers for therapeutic targets of Mets and obesity related diseases in the African American population. Future research is warranted to elucidate the functional consequences of these miRNAs in relation to MetS and abdominal obesity.

2260W

Family history taking in pediatrics: it's much more than just a checklist. *L. Tessier[1], J. C. Brehaut[2], B. K. Potter[3], P. Chakraborty[1,4], J. C. Carroll[5], B. Wilson[3] on behalf of the CIHR Emerging Team in Genomics in Screening.* 1) Newborn Screening Ontario, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Clinical Epidemiology Program, Ottawa Hospital Research Institute, ON, Canada; 3) School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa, ON, Canada; 4) Department of Pediatrics, University of Ottawa, ON, Canada; 5) Family Medicine, Mount Sinai Hospital, University of Toronto, ON, Canada.

Background: Family history (FH) is a risk factor for many conditions in pediatric practice; interest has increased in supporting systematic FH taking in this area by identifying core conditions for enquiry and developing point of care tools. There is little published about current practice to inform implementing such changes. Objectives: To inform future FH taking interventions by identifying pediatricians' perceptions, attitudes, beliefs, and current practices. Methods: The Theoretical Domains Framework (TDF) was used to develop a comprehensive interview scheme. Semi-structured interviews were conducted with pediatricians practising in the Ottawa, Canada region. Analysis was by the constant comparison method, using a thematic approach. Results: The data revealed FH taking to be a firmly embedded, but surprisingly complex, aspect of pediatric practice. Participants described FH as part of regular, holistic care, extending to social and interpersonal, as well as clinical purposes; in addition to disease risk assessment, FH information helped clarify diagnosis, and select medication; tailor the overall patient management based on family circumstances; and provide psychosocial support for parents. It was also used as a method to build a relationship with parents/carers. FH and social history information were inextricably linked and often appeared to be the same concept in participants' minds. Participating pediatricians generally expressed confidence in their FH skills and reported tailoring their approach according to experience, after initial training early in their career. While acknowledging some challenges in ensuring accuracy, they were attuned to nuances in their interactions with parents and children which would affect whether, and what, they were told about illness in the family. Most were not concerned about formal evidence, and would not change their practice except for "good reason". Conclusion: The use of the TDF helped ensure a comprehensive approach to FH taking in pediatric practice. The findings suggest that FH taking in this setting is a complex activity, embedded in routine care. Recommendations for systematic enquiry about specific conditions cannot be seen as a simple additional activity to current practice. Efforts to make FH taking more systematic may founder if they fail to take into account pediatricians' attitudes, perspectives, and practices. Further studies should explore and seek to confirm and expand our observations.

2261W

Mitochondrial Disease Patients' Motivations and Barriers to Participate in Clinical Trial. *Z. Zolkipli Cunningham[1], A. Stoddart[2,3], E. McCormick[2], R. Xiao[4], M. Falk[2,5].* 1) Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA 19104; 3) Arcadia University, Glenside, PA 19038; 4) Department of Biostatistics and Epidemiology, University of Pennsylvania, PA 19104; 5) University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19014.

Background We investigated the motivations and barriers for clinical trials participation in mitochondrial disease patients. **Methods** An initial survey was performed on pediatric and adult patients with definite or possible/probable mitochondrial disease from the Children's Hospital of Philadelphia (n=30/67 respondents). Validation was performed in a larger cohort of U. S. patients self-identified as having mitochondrial disease through the Rare Disease Clinical Research Network (RDCRN) contact registry (n=290/1119 respondents). **Results** In the survey of CHOP-enrolled subjects, the response rate was 45%, with completed surveys returned for 15 children and 15 adults. The top 3 symptoms for clinical trial participation were developmental delay, muscle weakness, and seizures for affected children, whereas adults prioritized muscle weakness, exercise intolerance, and chronic fatigue. In the RDCRN validation survey, the response rate was 26%, with completed surveys returned for 121 children and 169 adults. Chronic fatigue, muscle weakness and exercise intolerance were the top 3 motivating symptoms for clinical trial participation. Motivating factors in both surveys related to treatment characteristics included if the study drug was a vitamin, antioxidant, natural or plant-derivative; a pill; self-administered; administered once daily; received by all participants; and guaranteed provision to patients at the end of the study. Relative barriers to study participation included a trial of a new drug, requirement to discontinue current medications, and symptomatic progression. Indeed, trial drop-out upon disease progression was predicted by 40-47% of adults and 54-67% of children. Regarding study design, daily or weekly blood draws and a requirement to pay to participate were relative barriers. Not having to travel or only traveling within the same city, financial reimbursements, and participating in later-stage (phase 3) studies were motivating factors. Pediatric and adult populations did not differ in the relative influence of treatment type or trial design. Study limitations include self-reporting of disease status in the RDCRN cohort and potential completion bias. **Conclusions** As the mitochondrial disease patient community sits on the eve of having the opportunity to participate in a burgeoning array of clinical intervention trials, patient motivations and deterrents should be seriously considered to tailor trial design for this complex disease population.

2262W

Geriatrics disease-burden from genetics point of view. *B. Ganguly[1], K. Sen[2], N Kadam[2]*. 1) MGM Center for Genetic & Diagnosis, MGM New Bombay Hospital, New Mumbai, Maharashtra, India; 2) MGM New Bombay Hospital, New Mumbai, Maharashtra, India.

Stem cells are immortal, ever young and unspecified with high self-replication process. Hence stem cells are being increasingly considered in cell-based therapy in regenerative medicine. Despite having such stem cells in the body, every organism ages with bundles of ailments. In human, aging is directly proportional to increase in health problems involving almost all organs. Among all systems, the circulatory machinery stands first in all ages over 60 years. In India, 60 years is considered as the onset of aging while people retire from work. A survey on five-year hospital record on 46000 cases aged over 60 years revealed 25% affected with cardiovascular system followed by 20% with general complication of cold/cough/fever and GI disorientation. An equal damage was observed in nephrological and orthopedic systems, which measures ~9%; however, a difference of double was evidenced in males and females for the former and later systems respectively. Human aging is not merely a consequence of wear and tear, but rather it is the inefficient or lack of replacement of these spare parts that leads to aging. Externally, dictation of aging starts from changes in hair and skin, which are otherwise self-generating from the committed stem cells, and thus, this phenomenon declares death of stem cells. Similar incidences are noticed in wearing out of human teeth, and also remarkable reproductive senescence in females. The unsolved puzzle of aging and genetic programming is under extensive investigation including GWAS (Eg. Werner syndrome), GxE epigenetic impact (e. g. identical twins), copy number variation in the whole genome, etc. In Indian scenario, predominant susceptibility to cardiovascular disorder after 60 years is related to stress for transition from ‘income-to-noincome’ state (retirement from work). Significant osteoarthritis problem in females is undoubtedly associated with ageing of ovaries and its genetic program; however, nutritional status of Indian women aged over 60 years is expected to have epigenetic control over expression of genes responsible for releasing necessary hormones. In quest of death of stem cells in different organs, it has been reported that loss of miRNAs and disruption of BIM protein disarray the survival and death of stem cells. Investigation on death-resistance of cancer cells i. e. mutated stem cells, and understanding gene expression in elderly over 60 years would establish avenues for curbing age-related diseases and early aging.

2263T

Plasma Amino Acid and Urine Organic Acid Profiles of Filipino Patients with Maple Syrup Urine Disease (MSUD) and Correlation with their Neurologic Features. *M. D. Chiong[1], M. A. Tan[2], C. P. Cordero[3], E. D. Fodra[4], J. S. Manliguis[4], C. P. Lopez[4], L. M. Dalmacio[5].* 1) Clinical and Metabolic Genetics, Institute of Human Genetics, National Institutes of Health, Manila, Manila, Philippines; 2) Department of Pediatrics, Section of Pediatric Neurology, Philippine General Hospital; 3) Department of Clinical Epidemiology, University of the Philippines Manila, Philippines; 4) Biochemical Genetics Laboratory, Institute of Human Genetics, National Institutes of Health; 5) Department of Biochemistry, College of Medicine University of the Philippines Manila, Philippines.

Background: Maple syrup urine disease (MSUD) is the most common inborn error of metabolism in the country. The main cause of the neuropathology is not well established although the accumulation of branched chain amino acids (BCAA) and alteration in large neutral amino acids (LNAA) as well as energy deprivation have been suggested. **Objectives:** It is the aim of the study to determine the plasma amino acid and urine organic acid profiles of Filipino patients with MSUD and correlate the findings with their neurologic features. **Methodology:** Twenty six Filipino patients confirmed to have MSUD were studied and their results were compared with 26 age and sex matched controls. Their neurologic features were reviewed and correlated with the results of their plasma amino acid and urine organic acid profiles. **Results:** Majority of the patients with MSUD had developmental delay/intellectual disability (88%), speech delay (69%) and seizures (65%). The amino acid profile of MSUD patients revealed low glutamine and alanine with high levels of leucine, isoleucine, phenylalanine, threonine and alloisoleucine compared to controls ($p<0.05$). The urine organic acids showed significantly elevated excretion of the branched chain ketoacids and succinate ($p<0.05$), however other Krebs cycle metabolites were not found in significant amounts. There were no metabolite markers in the amino acids or organic acids that correlated significantly with the neurologic features. The most remarkable finding was the discriminant analysis done on 7 clinically and statistically significant important amino acids in the plasma wherein elevations in leucine, isoleucine, alloisoleucine, phenylalanine and threonine, and decreased levels of glutamine and alanine clearly defined the boundary between an MSUD case and control. **Conclusion:** The findings suggest that there could be altered LNAA metabolism among patients with MSUD when the BCAAs are elevated. A set of plasma amino profile comprising of 7 amino acids may be suggestive of MSUD if altogether present. The urine organic acid analysis showed elevated excretions of the branched chain ketoacids and succinate. However, the biochemical findings were not significantly correlated with the neurologic features of patients with MSUD thus, there could be other unknown factors that cause the neurologic impairments in MSUD other than the elevated BCAAs and corresponding ketoacids or elevated LNAAs. .

2264F

Whole exome sequencing identifies a novel frameshift variant causing familial hypomagnesemia with secondary hypocalcemia: a case report. *M. K. Azim[1], A. Mehnaz[2], J. Zahid Ahmed[2], G. Mujtaba[1].* 1) Jamil-ur-Rehman Center for Genome Research, International Center for Chemical and Biological Sciences, University of Karachi, Karachi, Pakistan; 2) Department of Pediatrics, Dow Medical College, Dow University of Health Sciences, Karachi, Pakistan.

Hypomagnesemia with secondary hypocalcemia is a rare autosomal-recessive disorder characterized by intense hypomagnesemia associated with hypocalcemia. Mutations in the *TRPM6* gene, encoding the epithelial Mg2+ channel TRPM6 have been proven to be the molecular cause of this disease. The purpose of this study was to search for causal mutations in a 2 months old male patient of hypomagnesemia from a consanguineous marriage. Biochemical analyses confirmed the diagnosis of HSH due to primary gastrointestinal loss of magnesium. Whole exome sequencing of the trio (i. e. proband and both parents) was carried out using Complete Genomics sequencing followed by standard bioinformatics analysis. Exome sequence coverage, number of variants, percentage of known variants and the Ti/Tv ratio were calculated. Mean coverage of the target regions ranged between 40-50X. ANNOVAR was used to annotate functional consequences of genetic variation from exome sequencing data. After variant filtering, 12 candidate variants were identified. Analysis of these variants showed several missense mutations and deletion of two base pairs (at 9:77377267) in exonic region of *TRPM6* which resulted in a novel frameshift mutation. These mutations in *TRPM6* were confirmed by Sanger sequencing. With these investigations in hand, the patient was managed with calcium gluconate and magnesium sulphate. The patient remained asymptomatic and was developmentally and neurologically normal till his last follow up.

2265T

the natural history protocol on congenital disorders of glycosylation. *L. Wolfe[1], C. Lam[1], C. F[1], w. G[1], D. Krasnewich[2].* 1) Office of the Clinical Director, NIH/NHGRI, Bethesda; 2) 2National Institute of General Medical Sciences, National Institutes of Health, Bethesda, MD, USA.

Background: Congenital disorders of glycosylation (CDGs) are a group of more than 100 inborn errors in the glycan modification of proteins, lipids, or other biomolecules. Due to the rarity of these disorders, their protean clinical manifestations, and absence of clinical protocols dedicated to their study, their natural histories remain largely unknown. The goal of our new protocol is to systematically characterize the different types of CDGs. Here we report the cohort of individuals who have been evaluated under our protocol thus far and discuss interesting findings in these cases. *Methods*: NHGRI study 14-HG-0071 “Clinical and Basic Investigations into Known and Suspected Congenital Disorders of Glycosylation” (identifier: NCT02089789) was approved in March, 2014 and enrolled its first subject in May, 2014. Under this protocol, subjects are admitted into the National Institutes of Health Clinical Center and undergo blood and urine collection, skin biopsy, clinical photography, radiographic evaluations, and clinical examinations. *Results*: Since the protocol’s approval, we have evaluated 28 individuals and have over 25 additional individuals awaiting admission. The specific subtypes we have seen include subjects with NGLY1-CDG, SDR5A3-CDG, ALG8-CDG, ALG12-CDG, ALG13-CDG, PMM2-CDG, PGM1-CDG, ALG1-CDG, ALG2-CDG, MOGS-CDG and GMPPB-CDG. Subjects pending evaluations include patients with NGLY1-CDG, PMM2-CDG, MOGS-CDG, ALG13-CDG, COG5-CDG and several individuals with single mutations in multiple CDG genes and phenotypes consistent with a CDG. Novel findings include: 1) increased antibody titers to rubella and rubeola vaccines in patients with SDR5A3-CDG, NGLY1-CDG, and ALG1-CDG; 2) low CSF protein, albumin, IgG and neurotransmitter concentrations in patients with NGLY1-CDG; 3) hypohidrosis in individuals with NGLY1-CDG; 4) normal peripheral hearing with consistently dysynchronous and delayed transmission through the brainstem on ABR exams in patients with NGLY1-CDG; 5) elevated urine glycosaminoglycans in individuals with NGLY1-CDG, and 6) low tear production in patients with NGLY1-CDG, ALG1-CDG, and GMPPB-CDG. *Conclusions*: We are actively recruiting patients from families willing to participate in our ongoing natural history protocol on CDGs. Our efforts to robustly characterize this group of disorders will enhance our understanding of the expected course in different subtypes of CDGs, and pave the way for future therapeutic trials.

2266F

Toward identification of the pathogenic cell type driving calcification in pseudoxanthoma elasticum. *S. G. Ziegler[1,2], C. R. Ferreira[3], T. J. Creamer[1], D. S. Warren[1], L. Goff[1], A. B. Pinkerton[4], J. L. Millan[4], W. A. Gahl[3], H. C. Dietz[1,2].* 1) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) HHMI, Chevy Chase, MD; 3) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD; 4) Sanford-Burnham Medical Research Institute, La Jolla, CA.

Biallelic mutations in *ABCC6*, an ATP-dependent transporter whose ligand remains unknown, cause pseudoxanthoma elasticum (PXE) characterized by adult-onset elastic fiber calcification in the eyes, skin, and vessels. Strong expression of ABCC6 in the liver supported the prevailing view that peripheral tissue calcification in PXE reflects failed liver secretion of an endocrine inhibitor of calcification. Contrary to this view, we found that fibroblasts cultured from patients with biallelic mutations in *ABCC6* had increased levels of tissue non-specific alkaline phosphatase (TNAP) and showed a strong tendency for induced calcification that can be prevented by a TNAP inhibitor. Breakdown products of ATP regulate TNAP activity, suggesting that PXE is caused by local defects in extracellular ATP metabolism. To determine the biologically relevant cell type driving PXE pathogenesis, we generated a conditional *Abcc6* mouse model. The initial site of calcification in PXE mice is the fibrous capsule surrounding the muzzle vibrissae and appears at 20 weeks of age when *Abcc6* is constitutively deleted using CMV-Cre or in a complete *Abcc6* knockout. However, robust liver-specific deletion of *Abcc6* using albumin-Cre failed to induce calcification at 20 weeks; the same was true for Cre drivers specific for vascular smooth muscle, vascular endothelium, bone marrow, renal tubular cells, pericytes, fibroblasts, adipocytes, and the craniofacial mesenchyme. Curiously, albumin-Cre mediated deletion of *Abcc6* associated with poorly penetrant and mild vibrissae calcification at 1 year of age. We speculated that delayed albumin-Cre induced calcification manifests local (as opposed to hepatic) allele recombination in peripheral calcification-prone tissues. Consistent with this hypothesis, a cross between mTmG reporter and albumin-Cre mice revealed postnatal (adult) local transition of a discrete cell population in the muzzle fibrous capsule to a GFP-positive state (indicating recombination). FACS and RNASeq analysis of GFP+ cells suggest that they represent a specialized multipotent progenitor lineage. These data suggest particular relevance and vulnerability of cells recruited to sites requiring chronic tissue maintenance or repair in the predisposition to calcification imposed by ABCC6 deficiency; this model reconciles multiple phenotypic aspects of PXE-associated calcification including age-dependence and spatial focality, and has the potential to inform novel therapeutic strategies.

2267T

Congenital Protein Losing Enteropathy: An inborn error of lipid metabolism due to *DGAT1* mutations. *H. Pri Chen[1,2,3], J. Stephen[1], T. Vilboux[1,4], Y. Haberman[5,6], B. Pode-Shakked[3,7,8], O. Barel[9], A. Di Segni[9], E. Eyal[9], G. Hout-Siloni[9], A. Lahad[5], T. Shalem[5], G. Rechavi[3,9], B. Weiss[3,5], Y. Anikster[3,7], W. A. Gahl[1,2,10], M. C. Malicdan[1,2,10].* 1) Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA; 2) NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, Maryland, USA; 3) Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 4) Inova Translational Medical Institute, Fairfax, Virginia, USA; 5) Division of Pediatric Gastroenterology, Hepatology and Nutrition, Edmond and Lily Safra Children's Hospital; 6) Cincinnati Children's Hospital Medical Center, OH, USA; 7) Metabolic Disease Unit, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel-Hashomer, Israel; 8) The Dr Pinchas Borenstein Talpiot Medical Leadership Program, Sheba Medical Center, Tel-Hashomer, Israel; 9) Sheba Cancer Research Center, Sheba Medical Center, Tel-Hashomer, Israel; 10) Office of the Clinical Director, NHGRI, National Institutes of Health, Bethesda, Maryland, USA.

Statement of Purpose: Protein Losing Enteropathy (PLE) is a clinical disorder of protein loss from the gastrointestinal system that results in hypoproteinemia and malnutrition. PLE is associated with a wide range of gastrointestinal disorders, including a unique syndrome associated with biallelic mutations in the *DGAT1* gene. Although previously identified in a single family, *DGAT1* mutations could be responsible for many undiagnosed cases of PLE in infancy. **Methods:** We investigated three children in two families presenting with severe diarrhea, hypoalbuminemia and PLE, using clinical studies, homozygosity mapping, exome sequencing, and expression studies at the transcript and protein levels. **Results:** In one family, homozygosity mapping using SNP arrays marked the *DGAT1* gene as the best candidate gene for causing PLE in the proband. Sequencing of the *DGAT1* coding region, including exon-intron boundaries and the promoter, identified a novel homozygous missense mutation, NM_012079. 5: c. 884T>C, Leu295Pro in the highly conserved Membrane Bound O-acyl Transferase (MBOAT) domain. In the second family, exome sequencing identified a previously reported splice site mutation in intron 8. Expression studies and western blotting verified reduced amounts of DGAT1. **Conclusions:** These cases of DGAT1 deficiency extend the molecular and phenotypic spectrum of this rare condition, which may be a more common cause of PLE than previously recognized.

2268F

Progression of behavioral and CNS deficits in a murine model of chronic neuronopathic Gaucher Disease. *Y. Sun[1,3], M. Dai[2], B. Liou[1], X. Wang[2], B. Swope[2,3], V. Inskeep[1], W. Zhang[4], G. A. Grabowski[5], D. Pan[2,3].* 1) Division of Human Genetics, Division of Human Genetics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 2) Experimental Hematology and Cancer Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 3) Department of Pediatrics, University of Cincinnati School of Medicine, Cincinnati, OH; 4) Division of Pathology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 5) Synageva BioPharma Corp, Lexington, MA.

The neuronopathic Gaucher Disease (nGD, Types 2 and 3) is the most frequent variant of Gaucher disease world-wide and has no effective specific treatment. Mutations of *GBA1* lead to defective acid β-glucosidase function and resultant accumulation of its substrates, glucosylceramide (GC) and glucosylsphingosine (GS). The CNS deficits in nGD were characterized by chronological behavioral profiles and the age of onset in a chronic nGD mouse model (9V/null). Abnormal weight gain emerged in 9V/null mice at 9- and 12-months for female and male mice, respectively. Compared to WT mice, significantly increased ratios of tissue to body weight were present for livers starting at 9-months and for spleen at 24-month, indicating the ages of onset for hepato- and spleno-megaly in 9V/null. Progressive brain accumulations of substrates were found in 9V/null mice as early as 6-months for GC and 3-months for GS. α-Synuclein, which is central to pathology in Parkinson's diseases, was progressively increased in 9V/null cortex from 6 to 12 month by immunoblotting analysis. α-Synuclein aggregates were distributed in neurons and astrocytes in hippocampus and cerebral cortex and P-Tau signals, a marker for neurodegeneration, were strong in cerebral cortex, hippocampus and cerebellum in 12-month 9V/null mice. Repeated open-field tests of the 9V/null mice compared to WT mice showed less habituation to the environment by spending less time grooming and more time exploring the open field by 9-months and older, indicating the onset of short-term memory deficits. In marble burying test, the 9V/null mice had a shorter latency to initiate burying activity at 6-months, but the latency increased significantly starting at 12-months and buried significantly more marbles to completion than the WT group, suggesting an abnormal response to the instinctive behavior and an abnormal activity in non-associative anxiety-like behavior. The conditional fear tests revealed that the 9V/null mice froze less than age- and gender-matched WT by12-months, indicating hippocampus-related memory defects. 9V/null mice showed abnormal gait that reached significance after 9-months. Unlike several acute nGD mouse models with death in 14-45 days, the chronic nature of the 9V/null neurobehavioral dysfunction, biochemical, and histopathological abnormalities provide a novel foundation for experimental designs to gain understanding of the progressive pathogenesis and evaluation of potential CNS therapies for nGD.

2269T

Behavioral and cholinergic pathway abnormalities in male MPS IIIB mice. *S. Le[1], S. Kan[1], B. Benedict[1], Q. Bui[1], J. Cushman[2], M. Sands[3], P. Dickson[1].* 1) Department of Pediatrics, LA Biomed at Harbor-UCLA, Torrance, CA; 2) Behavioral Testing Core, UCLA, Los Angeles, CA; 3) Department of Medicine, Washington University in Saint Louis School of Medicine, St. Louis, MO.

Mucopolysaccharidosis type IIIB (MPS IIIB or Sanfilippo B) causes progressive neurological deterioration. Patients with MPS IIIB have reduced fear associated with volume loss in the amygdala. In this study, we used 12 MPS IIIB and 12 littermate controls male mice age 16-20 weeks to determine whether this model would be similarly affected as human patients. Affected mice showed reduced anxiety, with a decrease in the number of stretch-attend postures during the elevated plus maze (p=0. 01, F=13. 7), an increased tendency to approach a novel mouse (p=0. 067, F=3. 66), and an increased tendency to linger in the center of an open field (p=0. 032, F=5. 28). Radial arm maze testing showed a longer latency to find reward during training (p=0. 025, F=5. 96). Acetylcholinesterase (AChE) activity staining and stereology volume measurement of amygdale were performed on MPS IIIB mice aged 41-42 weeks. Upon staining, we noticed an overall reduction in AChE activity staining in affected animals. To verify our observation, we measured AChE and choline acetyltransferase (ChAT) activity in whole brain homogenates of the mice that were used in the behavioral experiment. We found a 12. 4% reduction in mean AChE activity (p<0. 001) and no difference in ChAT activity in affected animals. However, as the amygdala may be specifically affected in MPS IIIB, we further evaluated AChE and ChAT in MPS IIIB affected mice and carrier controls using a hemicoronal section (thickness 2 mm) that contained amygdala. AChE activity was 25% lower (p=0. 006) and ChAT activity 14% lower (p=0. 002) in affected animals in that section of the brain. Cholinergic pathways are affected in adult-onset dementias, including Alzheimer disease. Our results suggest that male MPS IIIB mice display neurobehavioral deficits at a relatively early age, and that as in adult dementias, they may display deficits in cholinergic pathways.

2270F

Rare non-synonymous variations in the human ferroportin iron transporter gene (haemochromatosis type 4): the quest for causal mutations. *G. Le Gac[1,3], I. Callebaut[2], S. Pissard[4], C. Kannengiesser[5], V. Gérolami[6], C. Ged[7], F. Cartault[9], J. Rochette[8], C. Ka[3], C. Férec[3].* 1) Human Molecular Genetics , Inserm U1078, BREST, France; 2) IMPMC, Sorbonne Universités - UMR CNRS 7590, UPMC Univ Paris 06, Muséum d'Histoire Naturelle, IRD UMR 206, Paris, France; 3) Inserm U1078, Université de Brest, SFR SnInBioS, CHRU de Brest, Laboratoire de Génétique Moléculaire et d'Histocompatibilité, Etablissement Français du Sang - Bretagne, Brest, France; 4) UPEC (Université Paris Est Creteil), GHU Henri Mondor, Laboratoire de Génétique, Créteil, France; 5) Inserm U1149 - Center for Research on Inflammation, Université Paris Diderot, AP-HP, Hôpital Bichat, Département de Génétique, Paris, France; 6) CHU de Marseille, Hôpital Conception, Biologie Moléculaire; 7) Inserm U1035, Biothérapies des Maladies Génétiques et Cancers, Université de Bordeaux, CHU de Bordeaux, Pôle de Biologie et Pathologie, Bordeaux, France; 8) UPJV EA4666, CHU d'Amiens, Laboratoire de Génétique Moléculaire, Amiens, France; 9) CHU de la Réunion, Service de Génétique, Saint-Denis, France;9Inserm CIC0502, CHRU de Brest, Brest, France.

Haemochromatosis type 4 is a rare form of primary iron overload transmitted as an autosomal dominant trait caused by mutations in the gene encoding the iron transport protein ferroportin 1 (*SLC40A1*). *SLC40A1* mutations fall into two functional categories (loss- versus gain-of-function) underlying two distinct clinical entities (haemochromatosis type 4A-B). However, the vast majority of *SLC40A1* mutations are rare missense variations, with only a few showing strong evidence of causality. The present study reports the results of an integrated approach collecting genetic and phenotypic data from 44 suspected haemochromatosis type 4 patients, with comprehensive structural (3D model) and functional annotations. 1,2 Causality was demonstrated for 10 missense variants, showing a clear dichotomy between the two haemochromatosis type 4 subtypes. Two subgroups of loss-of-function mutations were distinguished: one impairing cell surface expression and one altering only iron egress. A new gain-of-function mutation was identified, and the degradation of ferroportin on hepcidin binding was shown to probably depend on the integrity of a large extracellular loop outside of the hepcidin-binding domain. Eight further missense variations, on the other hand, were shown to have no discernible effects at either protein or RNA level; these were found in apparently isolated patients and were associated with a less severe phenotype. The present findings illustrate the importance of combining in-silico and biochemical approaches to fully distinguish pathogenic *SLC40A1* mutations from benign variants. This has profound implications for patient management. 1 Le Gac G *et al.* Hum Mut. 2013;34:1371-802 Callebaut I *et al.* Hum Mol Genet. 2014;23:4479-90.

2271T

A Zebrafish Knockout of Argininosuccinate Lyase Separates Ureagenic and Biosynthetic Functions. *O. A. Shchelochkov[1], R. Bowman[1], L. Teesch[2], L. Yu[3], C. M. Brenner[4], R. A. Cornell[5].* 1) Stead Department of Pediatrics, University of Iowa Hospitals & Clinics, Iowa City, IA; 2) Mass Spectrometry Facility, University of Iowa, Iowa City, IA; 3) NMR Core Facility, University of Iowa, Iowa City, IA; 4) Carver College of Medicine, Iowa City, IA; 5) Anatomy & Cell Biology, University of Iowa, Iowa City, IA.

Argininosuccinate lyase (ASL) deficiency is a urea cycle disorder characterized by hyperammonemia and complications unrelated to hyperammonemic injury. Notably, liver dysfunction, hypertension and intellectual disability can occur in the absence of discernable hyperammonemia. In addition, urea cycle enzymes are expressed in a variety of cells that do not produce urea. We hypothesize that impaired metabolism of arginine and its downstream metabolites is responsible for non-hyperammonemic complications. However, a barrier towards characterizing the role of ASL in non-hyperammonemic complications is the difficulty in separating the ureagenic and biosynthetic functions of this enzyme. Though zebrafish excrete ammonia through their gills, the zebrafish genome encodes orthologues of all genes that would be termed urea cycle components. We therefore chose the zebrafish as a model to determine the non-ureagenic functions of ASL. Using TALEN technology, we generated a homozygous ASL knockout. Tissue amino acid analysis of ASL knockout larva showed the presence of argininosuccinic acid, a specific biochemical marker of ASL deficiency. Development was normal and the fish were viable and fertile for the first 4 months of life. However, ASL knockouts exhibited reduced reproduction after 4 months and failure to thrive by age 8. 5 months at which time mutants averaged less than 60% of the body weight and less than 90% of the length of wild-types (WT), $p < 0.01$ for each growth deficit. To further characterize ASL function in zebrafish, we monitored urea synthesis in the first 2 days post fertilization. This analysis indicates a functional urea cycle at this time point as WT embryos produced 361 mM urea. However, exposure of WT embryos to [15]NH4Cl led to incorporation into glutamine and citrulline but not into urea. Moreover, survival of embryos exposed to 5 mM ammonia chloride was equivalent in WT and ASL-deficient embryos. To test the hypothesis that ASL is essential for arginine synthesis, we soaked WT and ASL-deficient embryos in 4-amino-5-methylamino-2',7'-difluorofluorescein diacetate, a fluorescent nitric oxide (NO) probe. Mutants showed ~30% decreased fluorescence in comparison to WT ($p < 0.05$) supporting the notion that ASL is necessary in production of NO derived from arginine. These findings suggest that despite a functional urea cycle in embryos, ASL is not required for resistance to ammonia toxicity but is involved in synthesis of endogenous arginine and NO.

2272F

Insights into the pathophysiology of propionic acidemia provided by metabolomic analysis. *M. J. Miller[1], A. D. Kennedy[2], T. R. Donti[1], P. S. Atwal[1], L. A. Miller[2], M. V. Milburn[2], J. A. Ryals[2], Q. Sun[1], V. R. Sutton[1], S. H. Elsea[1].* 1) Dept. of Molecular and Human Genetics, Medical Genetics Laboratory, Baylor College of Medicine, Houston, TX; 2) Metabolon Inc. , Durham, NC.

Propionic acidemia is an autosomal recessive multisystem disorder that often presents in the neonatal period with acidosis, vomiting, and lethargy. More variable features include hyperammonemia as well as cardiovascular, central nervous system, gastrointestinal, immune, and renal symptoms. Underlying these phenotypic outcomes are myriad primary and secondary metabolic pathway perturbations which are not completely understood. To further explore the complex biochemical basis of this disease, we used a state-of-the-art metabolomic platform to analyze 9 plasma specimens collected from 8 different patients with propionic acidemia, one of whom had received a liver transplant. Of the approximately 900 small molecule analytes detected by this analysis, ~200 showed a significant perturbation (false discovery rate <0. 05) in patients with propionic acidemia when compared to a cohort of 70 individuals without this disease. The list of significantly perturbed analytes includes multiple biomarkers within the branch chain amino acid (BCAA) metabolic pathway, diet/treatment related compounds, and spectrally unique but, as of yet, unknown biochemicals. Numerous significant analyte perturbations were also detected in metabolic pathways interconnected with the BCAA pathway including previously reported dysregulation of the TCA and urea cycles. Building on these earlier findings we identified a significant reduction of the plasma levels of the urea cycle compound N-acetylglutamate and significant perturbations of the TCA cycle intermediates fumarate and malate (accumulation) as well as succinylcarnitine, cis-aconitate, and citrate (reduction). While marked elevations of many classic propionic acidemia biomarkers persisted in the patient with a donor liver, N-acetylglutamate as well as succinylcarnitine, cis-aconitate, and citrate appeared to normalize thus providing insights into the metabolic changes that occur after transplant. Taken together, these data lend additional support to a growing body of evidence advocating for the expanded use of the N-acetylglutamate analog, carglumic acid, as well as citrate in the critical care of patients with propionic acidemia. More generally, our results shed light on the pathophysiology of propionic acidemia and highlight the potential of metabolomic analyses to improve surveillance and treatment of patients with inborn errors of metabolism.

2273T

Musculardystrophy and Mitochondrial ca^{+2} up take disorder. *W. M. EYAID[1], W. shehhi[1], K. Dress[1], M. Balwi[3], w. Twaijri[1], A. Alassiri[4], V. K. Mootha[2].* 1) PEDIATRICS, KING ABDULAZIZ MEDICAL CITY, RIYADH, CENTRAL, Saudi Arabia; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA; 3) Dep. of pathology-molecular lab. KING ABDULAZIZ MEDICAL CITY, RIYADH, CENTRAL, Saudi Arabia; 4) Dep. of pathology-anatomical pathology,KING ABDULAZIZ MEDICAL CITY, RIYADH, CENTRAL, Saudi Arabia.

Mitochondrial calcium uptake has a central role in cell physiology by stimulating ATP production, shaping cytosolic calcium transients and regulating cell death. Mitochondrial Ca+2 uptake is mediated by the Ca+2 uniporter complex in the inner mitochondrial membrane, MCU. The MICU1 is a subunit of the MCU. It regulates channel opening in response to intracellular calcium content and calcium transients. Here we report a mutation in the MICU1 gene in two brothers who presented with incidental finding of elevated creatinine kinase, CK level. Both of them are having mild proximal weakness, and have no abnormal movement or learning difficulty. Muscle biopsy showed minimal myopathic changes. The Whole Exome Sequencing, WES showed that they are having homozygous mutation c. 547C>T (p. Q183X) in the MUC1 gene, and both parents were heterozygous. This is the second report in the literature of human mutation in the MUC1 gene. Our cases' clinical presentation is different from the cases reported before. Although Mutations in the MICU1 gene are directly linked to a form of neuromuscular disease, however, it is still unknown how these mutations lead to myopathy precisely.

2274F

1-^{13}C-Propionate Oxidation as a Measure of Hepatic Methylmalonyl-CoA Mutase (MUT) Function in Methylmalonic Acidemia (MMA). *I. Manoli[1], E. Harrington[1], S. Smith[2], J. Hattenbach[2], J. L. Sloan[1], K. Y. Chen[2], C. P. Venditti[1].* 1) Genetics and Molecular Biology Branch, NHGRI, NIH, Bethesda, MD; 2) Clinical Endocrinology Branch, NIDDK, NIH, Bethesda, MD.

Isolated methylmalonic acidemia (MMA), caused by a defect in methylmalonyl-CoA mutase (MUT) or its cofactor adenosylcobalamin (*cblA*, *cblB*, *cblD*-MMA), is characterized by increased morbidity and mortality. Elective liver (LT) or combined liver and kidney transplantation (LKT) are required to stabilize the most severely affected patients but do not correct the circulating metabolite perturbations. Studies in MMA mice have proven that 1-13C-propionate oxidation can be used to track whole body MUT activity conferred by viral or germ line transgenes, even in the presence of an expanded tracee pool and have encouraged translation to patients, with the aim of using this approach as one of the outcome measures in gene therapy trials. We measured sodium-1-13C-propionate oxidation in 27 patients with isolated MMA (*mut0* N=16, *mut-* N=6, *cblA* N=5, *cblB* N=1, age range 4-41y), 6 healthy volunteers (age range 21-33y) and 13 heterozygote controls (age range 35-62y). Resting energy expenditure was measured and serial breath samples were collected over 2 hours following an oral single bolus of 0. 5mg/kg of the tracer. Conversion of the majority of the labeled propionate to 13CO2 occured within 30 min in controls. Four healthy volunteers repeated the study three times each in order to evaluate inter- and intra-individual variability. The results demonstrated excellent reproducibility. Similarly, reproducibility was established with repeat testing in 8 MMA patients at separate times, despite a different plasma methylmalonic acid pool in two patients (1741 *vs* 2246 and 2260 vs 719 µmol/L). Decreased propionate oxidation was observed in all MMA patients vs. controls (P<0. 0001). The most severe patients, who carried two stop mutations (N=3), showed almost no oxidation of label, despite an isolated KT in one patient. In contrast, LKT recipients (N=2) showed a complete restoration of oxidation rates to control levels (P=0. 004 compared to non-transplanted patients), while only minimal oxidation was evident in a recipient after an auxiliary liver allograft post KT. B12-responsive patients with KT (N=2) showed near normal recovery, while a significant difference in measured activity was observed in one *cblA* patient pre and post B12 injection. 13C breath testing represents a non-invasive method to assess whole body *in vivo* MUT activity and could be used as prognostic or outcome measure for interventions, such as B12 supplementation, transplantation and gene or cell-therapy in MMA.

2275T

Gaucher Type 2 Neural Models from Human Induced Pluripotent Stem Cells Cindy E. McKinney, Ph. D. , Megan Corrigan-Cummins, OMS, Katherine Baumgarner, M. S. Stem Cell Lab, Edward Via College of Osteopathic Medicine (VCOM-CC) and Gibbs Research Institute, Spartanburg, SC. *C. McKinney, M. Corrigan-Cummins, K. Baumgarner.* Genetics and Biochemistry and Stem Cell Lab, Edward Via College of Osteopathic Medicine and Gibbs Research Institute, Spartanburg, SC.

The ability to reprogram a variety of human tissue sources (for example; blood PBMCs and fibroblasts) to create human induced pluripotent stem cells (hiPSC) has profound implications for studies of neurodevelopmental disease and for applications in regenerative medicine. Derivation of neuronal cells *in vitro,* by differentiating Gaucher type 2 (GD2) hiPSCs, allows for characterization and manipulation of these cells in the hope of providing insights into this fatal neurodegenerative genetic disorder. Previous studies were limited by constraints in obtaining GD2 patient and control brain tissues. The access provided by GD2 hiPSC derived neural models allows for (1) the opportunity to study cellular function in the GD2 cells and (2) the opportunity to develop drug therapy modalities that can be tested in these models and potentially moved to pharmacological development. We have obtained GD2 fibroblasts from the NIGMS Biorepsoitory (Coriell Institute) where some lines are L444P homozygotes or L444P compound heterozygotes. Control neonatal fibroblasts (BJ from ATCC) and control PBMCs from VCOM student donors were also obtained as controls. We have created hiPSC lines from all of these sources by reprogramming with the four Yamanaka factors using a non-integrating Sendai virus carrier. Characterization of these hiPSC includes screens for mycoplasma and STRs, karyotyping and immunocytochemistry for hiPSC pluripotent markers (Oct4, Sox2, Nanog, TRA1-60 and SSEA4). *In vitro* differentiation protocols show these lines are capable of generating endoderm, mesoderm and ectoderm tissues. Using Western blots, we have compared glucocerebrosidase levels in the fibroblasts and hiPSC lines with a control lysosomal protein (LIMP-2) and a housekeeping enzyme (GAPDH). Digital droplet PCR (ddPCR) using cDNA made from the cell lines shows the original GD2 fibroblasts lines have decreased GBA1 (lysosomal glucocerebrosidase) gene expression and a moderate increase in GBA2 (non-lysosomal glucocerebrosidase) gene expression compared to controls. Using the GD2 and control hiPSC lines, we derived neural progenitor cells (NPCs) expressing the neuromarkers, nestin and/or hSynapsin-EGFP. The expression of EGFP from the hSyapsin promoter in the NPCs allows for cell sorting of a homogeneous population of NPCs. The GD2 derived NPC are being further differentiated to neurons and astrocytes.

2276F

Clarifying the Phenotype of *NGLY1* Deficiency, the First Congenital Disorder of Deglycosylation. *C. Lam[1], C. Ferreira[1], D. Krasnewich[2], C. Toro[3], L. Latham[1], W. Zein[4], T. Lehky[5], C. Brewer[6], K. King[6], C. Wassif[6], S. Rosenzweig[7], J. Lyons[8], W. Gahl[1, 3], L. Wolfe[1, 3].* 1) NHGRI, NIH, Bethesda, MD; 2) NIGMS, NIH, Bethesda, MD; 3) UDP, NIH, Bethesda, MD; 4) NEI, NIH, Bethesda, MD; 5) NINDS, NIH, Bethesda, MD; 6) NIDCD, NIH, Bethesda, MD; 7) CC, NIH, Bethesda, MD; 8) NIAID, NIH, Bethesda, MD.

Background: *NGLY1* deficiency was first described to cause human disease in 2012. Encoded by *NGLY1*, N-glycanase 1 is a cytosolic enzyme that catalyzes the cleavage of the beta-aspartyl glycosylamine bond of N-linked glycoproteins releasing intact N-glycan species from proteins bound for degradation. Thus far, the human phenotype for *NGLY1* deficiency has been described in the literature in case reports and chart reviews. In this study, we clarify the clinical spectrum of *NGLY1* deficiency through a prospective natural history protocol. *Methods*: Subjects with *NGLY1* deficiency were enrolled in the NHGRI study 14-HG-0071 (NCT02089789), and 76-HG-0238 (NCT00369421). Under these protocols, subjects were admitted to the NIH Clinical Center and underwent evaluations using established clinical, radiographic, biochemical, and molecular methods. *Results*: We have evaluated 12 patients ages 2 to 21 with confirmed, biallelic, pathogenic *NGLY1* mutations. In these subjects, we explored previously reported clinical characteristics. We found decreased tear production (10/11 subjects), peripheral retinal pigmentary changes (6/11), optic nerve pallor (7/11), small cerebrum (7/11) and cerebellum (3/11), deficit of N-acetylaspartate (10/11) and elevated cerebral choline (4/11), history of seizures (6/12), axonal and/or demyelinating sensorimotor polyneuropathy (11/11), decreased sweat response (8/11), normal or near normal peripheral hearing sensitivity (9/11) with delayed and dyssynchronous transmission through the brainstem (9/11), mostly resolved transaminitis (8/9), abnormal liver texture on ultrasound (4/12), delayed bone age (8/10), complex hyperkinetic movement disorder (12/12), and global developmental delay with socialization as a relative strength (11/11). Additionally, we identified previously unreported features including consistently low CSF protein and CSF albumin, hyper-immune response to the rubella or rubeola vaccination (9/11), and a lower than predicted resting energy expenditure (9/12). *Discussion*: This study constitutes the largest case series of patients with *NGLY1* deficiency reported so far. Our prospective phenotyping has allowed us to expand the clinical spectrum of this disorder, and has led to the hypothesis that hypo-glycosylation contributes to the pathogenesis, given the phenotypic overlap between *NGLY1* deficiency and congenital disorders of glycosylation. We hope that these data will aid in development and evaluation of therapies for this disorder.

2277T

Hyperammonemia due to carbonic anhydrase VA deficiency: a case report. *H. H. C. Lee[1,2], E. Gorman[1], K. Bagley[1], V. W. Zhang[1], L. J. Wong[1], J. Zhang[1].* 1) Mitochondrial Laboratory, Baylor Miraca Genetics Laboratories, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Kowloon West Cluster Laboratory Genetic Service, Department of Pathology, Princess Margaret Hospital, Hong Kong.

BACKGROUND. *CA5A* encodes the intramitochondrial carbonic anhydrase type VA [MIM 114761], which provides bicarbonate for various mitochondrial enzymes including those involved in the urea cycle. A deficiency of this carbonic anhydrase caused by homozygous mutations was recently reported to cause hyperammonemia similar to that in proximal urea cycle defects, with inter- and intra-familial variability [MIM 615751]. **MATERIALS AND METHODS.** Cases received for testing for proximal urea cycle defects from May 2006 to April 2015 were retrieved from the Laboratory database. Clinical details submitted on the requisition forms were reviewed: those with lactic acidosis, hypoglycemia, hypocitrullinemia or positive newborn screening but with negative molecular results, as well as those tested negative with panels of urea cycle defects testing, were selected for *CA5A* Sanger sequencing. All seven coding exons were amplified and sequenced with gene-specific primers. **RESULTS**. A total of 909 cases (from 884 families) were retrieved and reviewed. Two samples were insufficient on arrival; 10 further tests were cancelled and not analyzed after requisition. Among the remaining 897 cases from 876 families, 54 were selected for Sanger sequencing of *CA5A*. Apparently homozygous NM_001739. 1(*CA5A*):c. 555G>A, a splice site mutation previously reported in Russian patients, was detected in one case born to non-consanguineous parents. The test was requested in 2011 when the child was 21 months old, who presented with hyperammonemia, hypoglycemia, acute mental status change and lactic acidosis; previous analyses for *CPS1*, *NAGS* and *OTC* were negative. **CONCLUSION.** Carbonic anhydrase VA deficiency is a rare but important cause of hyperammonemia that should be considered together with other urea cycle defects. Regarding the presence of pseudogenes, gene-specific primers could facilitate amplification and sequencing of the true gene. Other clinical and biochemical information would aid case selection for *CA5A* analysis.

2278F

Efficacy of an Electronic Medical Record Tool for Improving Detection of Hyperammonemia in Neonates - A Two-Year Data Review. *S. A. S. Vergano[1,2], F. Cusick[3], N. Sondheimer[4,5].* 1) Medical Genetics and Metabolism, Children's Hospital of The King's Daughters, Norfolk, VA; 2) Department of Pediatrics, Eastern Virginia Medical School, Norfolk, VA; 3) Information Services, The Children's Hospital of Philadelphia, Philadelphia, PA; 4) Department of Metabolism, The Children's Hospital of Philadelphia, PA; 5) Department of Pediatrics, Perelman School of Medicine, Philadelphia, PA.

Hyperammonemia is a rare but life-threatening condition in neonates that requires prompt recognition and treatment. The causes of hyperammonemia can be many and include liver failure, medications, infection, and inborn errors of metabolism (IEM). The latter is often the most difficult to recognize, as the symptoms of hyperammonemia often mimic other more common situations, including sepsis. Due to the rarity of IEMs, hyperammonemia is often not diagnosed until much later in an infant's course, after irreversible brain damage has occurred. In an attempt to improve detection rates of hyperammonemia in neonates, a Best Practice Alert (BPA) was created at The Children's Hospital of Philadelphia (CHOP) through the EPIC(c) electronic medical record program. The alert was created to fire on infants 2-7 days of age on whom a first-time blood gas was obtained. The details and methods of the creation of the BPA were then published in Molecular Genetics and Metabolism (Vergano et al, 2013). Since then, the BPA has become a software package owned by EPIC(c) which can be purchased by other hospitals wishing to utilize the alert. In April of 2014, it was installed within the Sentara Health System in Eastern Virginia. We reviewed the data generated from the BPA since its implementation in both institutions to determine the efficacy of the tool in identifying infants with hyperammonemia secondary to IEMs. At CHOP, the alert has fired almost 1000 times, resulting in about 160 orders of ammonia levels, and identifying one infant with hyperammonemia with MMA. This is consistent with the low positive-predictive value anticipated with this tool (<1%) and the overall prevalence of IEMs, estimated at about 10-15/100,000 births. At Sentara, the alert has fired 114 times and correctly identified one hyperammonemic infant with an FAOD. There do not appear to have been any neonates with hyperammonemia who met inclusion criteria for the alert and were subsequently missed. We report here the statistics from this BPA in two institutions to evaluate its the utility and efficacy. Thus far, it has identified those infants for whom it was targeted, with no apparent false negative results. We hope this potentially life-saving electronic alert will continue be implemented in other hospitals and health systems across the country, enabling faster recognition of hyperammonemia, and ultimately improve the outcomes of children with these condition.

2279T

Development of a new MRM-MS assay for the diagnosis of Morquio Syndrome in Dried Blood Spots. *C. Cozma[1], S. Eichler[1], S. Oppermann[1], M. I. Iurascu[1], A. K. Giese[2], A. Rolfs[1,2].* 1) Centogene AG, Rostock, Germany; 2) Albrecht-Kossel-Institute, University of Rostock, Rostock, Germany.

Background and aims: Mucopolysaccharidosis IVA (MPS IVA; Morquio A disease) is an autosomal recessive disease caused and characterized by a decreased activity of N-acetylgalactosamine-6-sulfate sulfatase (*GALNS*, EC 3. 1. 6. 4), resulting in accumulation of glycosaminoglycans in tissues that lead to secondary organ damage. Recently approved enzyme replacement therapy renders the easy and early identification of MPS IVA of out-most importance. We propose a completely new assay for the stable and reproducible detection of *GALNS* deficiency in dry blood spots (DBS). **Methods:** For the validation blood samples were taken from 59 healthy individuals and 24 randomly selected genetically confirmed MPS IVA patients. The MPS IVA patients presented 18 different *GALNS* mutations, 11 of which never before described in literature. All patients presented MPS IVA phenotype and a low *GALNS* activity in leucocytes. The material extracted from DBS was incubated with a 4-methylumbelliferyl-β-D-galactopyranoside-6-sulfate as a specific substrate for *GALNS*. In a second step the intermediary enzymatic product is incubated with beta-galactosidase. Final product, 4-methylumbelliferone, obtained after adding exogenous beta-galactosidase, was quantified by LC/MRM-MS (liquid-chromatography/multiple-reaction-monitoring mass-spectrometry). 4-propyl-5-hydroxy-7-methyl-2h-chromen-2-one was used as internal standard, a compound with a similar molecular structure and fragmentation pattern in negative ion mode as 4-methylumbelliferone. **Results:** The enzymatic assay yielded a positive and negative predictive value of 1. 0 for genetically confirmed MPS IVA patients (*GALNS* activity of 0. 35 ± 0. 21 µmol/L/h) and for controls with normal *GALNS* activity (23. 1 ± 5. 3 µmol/L/h). With present enzymatic conditions, the reaction yield in dried blood spots is at least 20 fold higher than any previously reported data with other assays. **Conclusion:** The present LC/MRM-MS based assay for MPS IVA diagnosis provides an easy, highly-standardized, accurate and innovative quantification of the enzymatic product in vitro and distinguishes perfectly between MPS IVA affected patients and normal controls. This technique will significantly simplify the early detection of MPS IVA patients.

2280F

Cerebrotendinous xanthomatosis in an adult with soft tissue masses andprogressive neurological decline. *A. Alhariri[1], V. OZA[2], K. HAMILTON[4], K. CORDORO[2], M. MALLOY[3], A. SLAVOTINEK[1].* 1) Medical Genetics Division, UCSF, SAN FRANCISCO, CA; 2) Department of Dermatology, UCSF, SAN FRANCISCO, CA; 3) Department of Medicine, UCSF, SAN FRANCISCO, CA; 4) Department of Neurology, UCSF, SAN FRANCISCO, CA.

Cerebrotendinous xanthomatosis (CTX) is a treatable autosomal recessivelipid storage inborn error of metabolism. It is caused by genetic deficiency of the 27-hydroxylase enzyme (encoded by CYP27A1 gene on 2q35). It is important for cholesterol metabolism, especially in bile acid synthesis and in the 25-hydroxylation of vitamin D3 in the liver. Reduced bile acid synthesis and cholestanol accumulation in the tissues cause the disease. Clinical manifestations can vary in the onset and affected organs. Major manifestations include infantile onset chronic diarrhea, childhood cataracts, adult onset tendon-distributed xanthomas, adult onset progressive neurologic impairment and less frequently, coronary heart disease, osteoporosis and optic disc and retina abnormalities. Early diagnosis and treatment with chenodeoxycholic acid may prevent further complications, mainly neurological manifestations. Here we present a 36 year old Cantonese male who was referred to medical genetics for evaluation of progressive spastic paraplegia and large, soft tissue masses in ankles. He started to have symptoms around age 20 years. He presented with multiple soft tissue lesions at bilateral Achilles tendons that were 5-10 cm in size, a soft tissue lesion at the right posterior elbow and progressive ataxia requiring wheelchair for ambulation. No history of diarrhea. Family history is unremarkable. Brain and cervical spine MRI showed global cerebral and cerebellar volume loss and extensive non-enhancing white matter symmetric hyperintensities. His cholestanol levels were 10 fold above upper normal limits 34. 76 ug/ml (0. 86 to 3. 71). Eye exam showed mild cataract. His fasting lipid profile showed total cholesterol 198 (<200 mg/dL), triglycerides 184 (<150 mg/dL), HDL 66 (>39 mg/dL), LDL 95 (<130 mg/dL). Skin biopsy confirmed xanthomas. Chenodeoxycholic acid (CDCA) 250 mg twice daily will be started. Testing for CYP27A1 mutations is pending. We conclude that the diagnosis of CTX might be suggested by soft tissue tumors at the tendons distribution along with progressive neurologic decline, cataract and other manifestations of cholestanol deposits in multiple tissues. Diagnosis can be made by clinical presentation, elevated serum cholestanol levels, evidence of xanthomas in soft tissue growths and molecular testing for CYP27A1 gene mutations. Establishing an early diagnosis of CTX may prevent the neurological decline and improve the morbidity and mortality of this condition.

2281T

Real-world experience in the diagnosis of neuronal ceroid lipofuscinosis type 2 (CLN2): Report from an international collaboration of experts. *N. Miller[1], S. A. Mole[2], J. L. Cohen-Pfeffer[1], R. Crystal[3], E. de los Reyes[4], Y. Eto[5], M. Fietz[6], B. Héron[7], E. Izzo[1], A. Kohlschütter[8], C. M. Lourenço[9], I. Noher de Halac[10], D. A. Pearce[11], M. S. Pérez-Poyato[12], A. Simonati[13], A. Schulz[8].* 1) BioMarin Pharmaceutical Inc. , Novato, CA, USA; 2) MRC Laboratory for Molecular Cell Biology, University College London, London, UK; 3) Department of Genetic Medicine, Weill Cornell Medical College, New York, NY, USA; 4) Department of Pediatric Neurology, Nationwide Children's Hospital, The Ohio State University, Columbus, OH, USA; 5) Advanced Clinical Research Center, Southern Tohoku Brain Research Center, Kawasaki, Japan; 6) SA Pathology, North Adelaide, South Australia, Australia; 7) Department of Pediatric Neurology, Trousseau Hospital, CHU Paris Est, Paris, France; 8) Department of Pediatrics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; 9) Department of Medical Genetics, School of Medicine of Ribeirao Preto, University of São Paulo, São Paulo, Brazil; 10) Universidad Nacional de Córdoba, Facultad de Ciencias Médicas, Cordoba, Argentina; 11) Sanford Children's Health Research Center, Sioux Falls, South Dakota, USA; 12) Pediatric Neurology, Hospital Universitario Marqués de Valdecilla, Cantabria, Spain; 13) Department of Neurological and Movement Sciences-Neurology, University of Verona, Verona, Italy.

Background: CLN2 disorder is a lysosomal storage disorder resulting from TPP1 enzyme deficiency that causes progressive neurological degeneration and early mortality. CLN2 disorder is rare and often unsuspected, leading to delays in diagnosis. **Methods:** In late 2014, 18 international CLN2 experts (clinicians, academic researchers, and laboratory directors) answered a comprehensive survey on CLN2 disorder and a subset met to discuss experiences, current practices and shortcomings in diagnosis of CLN2. **Results:** 70% of laboratory experts considered the standard for CLN2 diagnosis to be a demonstrated decrease in TPP1 enzyme activity, with the remaining experts favoring molecular detection of pathogenic *CLN2*/*TPP1* mutations. Delays in the diagnosis of CLN2 were identified as a crucial concern: 82% of the group responded that patient referral to a specialist can typically take longer than one year. Laboratory experts identified the challenge in reaching a suspicion of CLN2 (50%) and lack of awareness of available tests (83%) as common reasons for delays. **Discussion:** Experts agreed that reliable techniques exist for CLN2 diagnosis and identified timely referral as a key challenge. An upcoming CLN2 expert meeting will define laboratory-based screening and diagnostic guidelines in order to establish best practices for use of biochemical genetics testing in CLN2 diagnosis.

2282F

Inborn errors of metabolism in the molecular era: will biochemical analyses became a second-tier test? *L. S. A. Costa[1,2], F. B PIAZZON[1,2], F. P MONTEIRO[2], J. P. KITAJIMA[2], I. CORREIA[2], M. DELLAMANO[2], E. L. FREITAS[2], D. SCHLESINGER[2], C. BUENO[1], F. KOK[1,2].* 1) UNIVERSIDADE DE São PAULO - USP-SP, SAO PAULO, SAO PAULO, Brazil; 2) MENDELICS GENOMIC ANALYSIS, SAO PAULO - SP, BRAZIL.

Molecular analysis using next-generation sequence (NGS) technology already has a huge impact in clinical practice and is a fast and reliable way to provide diagnosis for rare, atypical or genetically heterogeneous conditions. In the past 2 years, we performed 720 clinical whole exome sequencing (WES) and identified 63 different IEM in 76 individuals, 35 of them without a previous suspicion of IEM. Diagnosed conditions belonged to the following group of diseases: LSD (13), peroxisomal (2), mitochondrial (13), amino acids (7), CDGs (6), GSD (3), neurotransmitter (3), nucleotides (3), metal (2), organic & FAO (7) and other (18). Among the IEM diagnosed, many conditions had only a couple of reports, as the X-linked SSR4-CDG, the purine synthesis ATIC deficiency, the inositol glycan anchor synthesis defect PGAP1, PGAP2 and PIGO and the isoleucine metabolism EHCS1 deficiency. A biochemical confirmatory test was performed when appropriate. In countries with limited access to biochemical tests, the use of WES might overcome most of diagnostic hurdles related to turnaround time, quality, cost, availability and dependence on overseas service provider. With a single test, we may answer several different questions.

2283T

Screening for mucopolysaccharidosis disorders in children and young adults presenting with bilateral hip disease. *G. Scharer[1], N. Mendelsohn[1], P. Stasikelis[4], S. Gutknecht[2], B. Smith[3], T. Wood[3], D. Pond[1], A. Bellomo[3], R. Olson[1], L. Read[1], K. Mason[4], T. Ramsey[3], L. Pollard[3], M. Barrett[1], C. Rogers[3,4].* 1) Children's Hospitals and Clinics of Minnesota, Minneapolis, MN, United States; 2) Gillette Children's Specialty HealthCare, St. Paul, MN, United States; 3) Greenwood Genetic Center, Greenwood, SC, United States; 4) Greenville Shriners Hospital, Greenville, SC, United States.

Introduction: Mucopolysaccharidosis (MPS) IVA and MPS VI are usually diagnosed by testing urine for elevated mucopolysaccharides followed by confirmatory enzyme testing for deficient N-acetylgalactosamine-6-sulfatase (GALNS) or arylsulfatase B (ARSB), respectively. Urine testing can be inconclusive. A 10-year-old boy with bilateral hip pain beginning at 9 years old was diagnosed with Legg-Calve-Perthes (LCP) and subtle spine abnormalities. Quantitative urine testing was normal, but thin layer chromatography revealed dermatan sulfate. Enzyme testing confirmed MPS VI. Two MPS IVA patients presented with hip dysplasia, joint pain, and normal or mildly elevated urine testing for glycosaminoglycans. Enzyme testing confirmed MPS IVA. Bilateral hip disease is documented as an early finding in MPS. We hypothesized enzyme testing in patients with bilateral hip disease may detect attenuated MPS IVA or VI. **Methods:** Participants were selected from 2 healthcare systems in Minnesota (ages ≤21 years) and 1 in South Carolina (SC) (ages <19 years) with diagnoses of bilateral spondyloepiphyseal dysplasia, multiple epiphyseal dysplasia (MED), LCP, unclassified skeletal hip dysplasia, or multi-joint pain (SC only). Blood samples collected for leukocyte enzyme testing (GALNS, ARSB) were analyzed by Greenwood Genetic Center using standard procedures. Abnormal results led to sequencing the associated gene. **Results:** In total, 3,652 charts were reviewed, 139 were eligible, and 39 patients enrolled. One was diagnosed with MPS IVA—a 10-year-old girl with a misdiagnosis of MED, presented with hip and shoulder pain, short stature, and mildly flat facies. Radiographs suggested attenuated MPS IVA. GALNS enzyme activity was deficient, and 2 pathogenic changes were found in the *GALNS* gene (p. D233N, p. P498S). Another patient had low arylsulfatase B activity (28. 2, normal 57-493), which was normal (85. 7) upon repeat analysis. Sequencing of *ARSB* revealed 2 polymorphisms (p. V358M, p. S384N), each of which causes moderately reduced enzyme activity *in vitro*, likely explaining the enzyme results. **Conclusion:** MPS VI and MPS IVA are rare disorders with hip disease as a common symptom. Of the 39 patients enrolled in the study, one was diagnosed with MPS IVA. The authors suggest consideration of MPS disease in children with bilateral hip disease to include enzyme screening.

2284F

Exome sequencing to identify the genetic bases for lysosomal storage diseases of unknown etiology. *J. Xing[1, 2], N. Wang[1, 2], E. Gedvilaite[1], Y. Zhang[1, 2], D. Kumar[3], R. Donnelly[4], D. Sleat[5, 6], P. Lobel[5, 6].* 1) Dept of Genetics, Rutgers University, Piscataway, NJ; 2) Human Genetics Institute of New Jersey, Rutgers University, Piscataway, NJ; 3) Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ; 4) Molecular Resource Facility, Rutgers - New Jersey Medical School, Newark, NJ; 5) Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, NJ; 6) Dept of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers University, Piscataway, NJ.

Lysosomes are membrane-bound, acidic eukaryotic cellular organelles. As an enzyme container, they play important roles in the degradation of macromolecules. Monogenic mutations resulting in the loss of enzyme activities in the lysosome may lead to severe health problem, such as neurodegeneration, early death, etc. These conditions are categorized as lysosome storage diseases (LSDs). The diagnosis of LSDs is typically straightforward, but in some cases, mutations that result in atypical clinical presentation or defects in previously undescribed lysosomal disease genes may complicate the identification of the underlying genetic defect. Here, we performed whole exome sequencing on 14 suspected LSD cases, with the goal of finding the causal mutations in each case. From the raw sequence data, we first identified DNA variants in each individual using three variant discovery pipelines: the Genome Analysis Toolkit, LifeScope and CLC Genomics Workbench. We then used the Variant Annotation Analysis Selection Tool (VAAST) to prioritize disease-causing mutations in 848 candidate LSD genes. As a probabilistic disease gene finder, VAAST integrates allele frequency, amino acid substitution severity and conservation information into a composite likelihood framework. Different from hard filtering methods, VAAST preserves all the candidates by listing them according to their disease-causing potential. So far a number of candidate variants have been identified, and we are performing downstream mutation validation and proteomic analyses to investigate the potential connection between our candidate variants and LSDs. Our project makes use of bioinformatics analyses to decode enormous exome sequencing data, narrowing down candidate lists and largely increasing the efficiency of downstream proteomic studies. Our results will shed light on the genetic basis of LSDs.

2285T

Niemann-Pick disease type B: a not rare diagnosis in patients referred for investigation of Gaucher disease type 1. *R. Giugliani, F. M. Sebastiao, K. Michelin-Tirelli, F. Bitencourt, F. B. Trapp, M. G. Burin.* Medical Genetics Service, HCPA/UFRGS, Porto Alegre, RS, Brazil.

The Laboratory of Inborn Errors of Metabolism/Medical Genetics Service/Hospital de Clínicas de Porto Alegre is a reference center for the diagnosis of lysosomal storage diseases (LSDs) in Brazil, including Gaucher disease (GD) and Niemann-Pick disease (NPD). GD is caused by β-glucosidase deficiency and classified as GD type 1 (GD-1), more attenuated and more common, known as non neurophatic, and GD type 2 and GD 3, more severe forms with neurological involvement. NPD is caused by acid sphingomyelinase deficiency and can be classified as NPD type A, a severe form with massive visceromegaly and early and pronounced neurological deterioration, with low life expectance, or as NPD type B (NPD-B), a more attenuated type, with little or none neurological involvement, with patients surviving into adulthood. GD-1 and NPD-B have several clinical manifestations in common, as marked hepatosplenomegaly and hematological problems, and in both conditions plasma chitotriosidase is usually increased. Therefore, when there is a clinical suspicion of GD-1, it is important to think in NPD-B as diagnostic possibility as well. Final diagnosis is usually obtained by the identification of the specific enzyme deficiency in leukocytes or fibroblasts. The purpose of this work was to identify the proportion of cases diagnosed as NPD-B in our lab in whom the initial suspicion was GD-1. We performed a retrospective analysis of our records (from 1988 to 2015) of all patients in whom acid sphingomyelinase deficiency was identified. We retrieved 202 NPD cases, being 89 of them with NPD-B, 22 with NPD-A, and 91 who could not be classified as no phenotypical information was available. In the 89 known cases of NPD-B, 18 (20 %) where referred with the initial suspicion of GD-1 (surprisingly, in 3 of them there was a biochemical diagnosis of GD-1 established in another lab). The main signs and symptoms in the 18 cases of NPD-B in whom the first suspicion was GD-1 were hepatosplenomegaly, anemia, thrombocytopenia, pancytopenia, and growth retardation, quite common in GD-1 as well. These results highlight the need to investigate NPD-B (through the measurement of acid sphingomyelinase activity) in all cases with GD-1 suspicion in whom this diagnosis was not confirmed, once, in our experience, 20% of NPD-B diagnoses were obtained incidentally in patients referred for the investigation of GD-1. The correct identification of NPD-B becomes more important as ERT for this disease is being developed.

2286F

Comprehensive screening of Urine sample for Metabolic Disorder using Gas Chromatography /Mass spectrometry. *B. Varughese, C. KUMARI, P. KADAM, S. K. POLIPALLI, S. KAPOOR.* PEDIATRICS RESEARCH AND GENETIC LAB, DEPARTMENT OF PEDIATRICS, MAULANA AZAD MEDICAL COLLEGE AND UNIVERSITY OF DELHI, DELHI, India.

Background & Aim:GC-MS has been widely used worldwide in the diagnosis of IEM because of its high sensitivity, accuracy and effectiveness in the analysis of multiple compounds simultaneously. We understand the need for a rapid and precise diagnosis in the clinically suspected metabolic disorders in New Born Screening in India. This study was designed to investigate the organic acid profile in clinically suspected metabolic disorder patients. Material and Method:Over a period of one year (2014-2015), a total of 1500 clinically suspected children were screened for metabolic abnormalities. Urine samples were analyzed for abnormal constituents like ammonia, lactate, reducing sugar, proteins, ketone bodies by routine laboratory chemical tests. Organic Acid extraction from urine was carried out using ethyl acetate followed by derivatization using N,O- bis- trimethylsilyltrifluroacetamide (BSTFA) and analysed for phenylketones, organic acids, ketoacids, succinylaccetone & glycines by GC/MS analysis supported by NIST v 2. 0g mass spectra library. Results:Out of 1500 screened only 37 (2. 46%) children were diagnosed to have IEM . In our study , the most common cases reported are, Lactic acidosis 13 (35. 1 %) ,Glutaric acidemia 6 (16. 2%), Methylmalonic aciduria 5 (13. 5%), followed by Tyrosinemia, Neuroblastoma, Maple syrup urine disorder , Gama- Amino butyric aciduria which collectively contributes 8 (5. 40%). Isovaleric acidemia 1 (2. 7 %) , Citrullinemia, Beta- Ketothiolase deficiency 1 (2. 7 %), propionic acidemia 1 (2. 7 %) and orotic aciduria1 (2. 7 %) . The common clinical manifestations observed among all participants were neurological features like convulsions (25. 7%), delayed milestones (17. 9%), and followed by metabolic acidosis (17. 2%) and hypoglycemia (10. 1%). Conclusion: High index of suspicion from clinicians supported by preliminary screening tests can aid in early presumptive diagnosis, which helps in initiating early treatment to prevent lethal neurological complications.

2287T

Clinical metabolomic profiling for the diagnosis of neurometabolic disorders associated with seizures and global developmental delay.

T. Donti[1], G. Cappuccio[2], M. Miller[1], P. Atwal[1], A. Kennedy[3], L. Miller[3], M. Milburn[3], J. Ryals[3], A. Cardon[4], C. Bacino[1], L. Emrick[1,4], J. Hertecant[5], A. Beaudet[1], M. Bainbridge[1], B. Graham[1], Q. Sun[1], V. R. Sutton[1], S. Elsea[1]. 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Department of Translational Medical Sciences, Section of Pediatrics, Federico II University, Naples, Italy; 3) Metabolon Inc. , Durham, NC; 4) Section of Pediatric Neurology and Neuroscience, Baylor College of Medicine, Houston, TX; 5) Tawam Hospital, UAE.

Neurometabolic disorders may present in infancy or early childhood with seizures, hypotonia and developmental delay or other focal neurologic signs. In individuals presenting with undifferentiated phenotypes such as developmental delay, hypotonia and seizures, the list of differential diagnoses is often very long and includes metabolic/neurometabolic, genomic, and other Mendelian disorders. We have previously documented the ability of untargeted metabolomic profiling to diagnose a variety of amino acid, organic acid, and fatty acid oxidation disorders. The untargeted nature of global metabolomic profiling allows screening for many metabolic disorders in one plasma sample. Here, we report 5 patients with various neurometabolic disorders identifiable by metabolomic analysis of plasma. The disorders include aromatic amino acid decarboxylase (AADC) deficiency, 4-aminobutyrate aminotransferase (ABAT) deficiency, adenylosuccinate lyase (ADSL) deficiency, citrate transporter deficiency, and urocanase deficiency. Four of the five (all except ADSL) had variants of uncertain significance (VUS) in related genes on whole exome sequencing (WES) which prompted the metabolomic profiling. We observed elevations of 3-methoxytyrosine, 2-pyrrolidinone, succinyladenosine, citrate and imidazole propionic acid in AADC, ABAT, ADSL, citrate transporter and urocanase deficiency, respectively. The perturbations in the metabolomic profile of plasma from these patients are unique and specific for these particular disorders. The standard diagnostic test for AADC, ABAT, and ADSL deficiency is CSF neurotransmitter analysis (although our patient with ABAT had this test and the panel did not include GABA or other diagnostic metabolites) and for urocanase deficiency, it is an enzyme activity assay from a liver biopsy. Citrate transporter deficiency is a rare disorder for which no clinically-available biochemical test is available. These cases demonstrate the ability of untargeted metabolomic profiling in confirming VUS found on WES, as well as the ability to use metabolomic profiling alone to screen for a wide range of metabolic/neurometabolic disorders.

2288F

Known and Novel ATP6 Mutations in the Mitochondrial Genome: Delineation of the phenotypic spectrum and functional effects.

R. D. Ganetzky[1], E. McCormick[1, 2], E. Polyak[1, 2], M. Peng[1, 2], M. Bennett[3], E. Nakamaru-Ogiso[4], R. Rodenburg[5,6], M. J. Falk[1,2]. 1) Section of Biochemical Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Clinical Genetics, Children's Hospital of Philadelphia; 3) Department of Pathology, Children's Hospital of Philadelphia; 4) ohnson Research Foundation, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania; 5) VU University Medical Center; 6) Department of Pediatrics, Radboud University Nijmegen Medical Center.

BACKGROUND: ATP synthase (complex V, CV) is an essential mitochondrial OXPHOS complex, wherein the proton gradient generated by the electron transport system is dissipated to generate energy as ATP. Ten pathogenic mutations have previously been reported in *ATP6*, one of two mtDNA-encoded complex V subunits. Clinical phenotypes that have been reported in association with *ATP6* mutations include primary lactic acidosis, Leigh syndrome and NARP. A growing number of novel *ATP6* variants of uncertain significance (VUS) have been found in a range of other phenotypic presentations. The lack of CLIA-approved functional assays to validate complex V functional effects prohibits accurate classification of novel *ATP6* variants. **METHODS:** A systematic review was performed of all reported *ATP6* mutations. Clinical and biochemical phenotypes were also evaluated in 8 patients from 6 families who have novel VUS in *ATP6* seen at The Children's Hospital of Philadelphia. Functional validation is underway of 4 methods to assess complex V activity/function: 1) spectrophotometric CV enzyme activity analysis, 2) blue native gel analysis with in-gel CV activity assay, 3) fluorometric assay of ATP synthetic rate, and 4) high-resolution respirometry analysis of integrated OXPHOS capacity with simultaneous fluorescence analysis of membrane potential (Oxygraph-2K, Oroboros). **RESULTS:** We have collated all reported *ATP6* phenotypes, functional effects of individual variants, functional assay results, inheritance, and heteroplasmy loads. A high correlation exists between heteroplasmy load and phenotype in reported pathogenic *ATP6* mutations. Proposed pathophysiology includes inefficient coupling between proton transport and ATP synthesis, loss of proton transport and accumulation of reactive oxygen species. Functional analyses are underway with to assess each assay's relative sensitivity, specificity, cost, and ease of use. Subsequently, these methods will be applied to assess CV activity in patient fibroblasts and cybrids from individuals with VUS in *ATP6*. **CONCLUSION:** While a handful of *ATP6* mutations have been validated to cause classical mitochondrial disease phenotypes, determining the functional relevance of novel *ATP6* variants being identified in additional phenotypes is a challenge. Assessing multiple diagnostic modalities will enable the development of robust clinical diagnostic assays to inform the clinical significance of novel *ATP6* mutations.

2289T

Extremely skewed X-inactivation leading to a manifestation of mucopolysaccharidosis type II in a female patient. *L. Dvorakova[1], M. Reboun[1], T. Veselkova[1], J. Vcelak[2], G. Storkanova[1], L. Stolnaja[1], M. Hrebicek[1], D. Musalkova[1], H. Poupetova[1], K. Peskova[1], J. Zeman[3].* 1) Inst Inherit Metabol Disorders, General University Hospital, Prague 2, Czech Republic; 2) Institute of Endocrinology, Prague, Czech Republic; 3) Department of Paediatrics and Adolescent Medicine, General University Hospital in Prague and First Faculty of Medicine, Charles University in Prague, Czech Republic.

Mucopolysaccharidosis type II (Hunter syndrome, MPS II, OMIM 309900) is an X-linked recessively inherited lysosomal storage disorder resulting from a deficiency of iduronate-2-sulfatase activity (IDS, EC 3. 1. 6. 13). The vast majority of patients are males, while only 16 female patients have been reported so far. We present a 3-year old girl with coarse face, mild hepatosplenomegaly, dysostosis multiplex and mild psychomotor retardation. Enzyme assay performed in both leukocytes and serum showed deficiency of IDS activity comparable to hemizygous MPS II patients. Diagnosis of MPS II was also established at the molecular level. A heterozygous substitution c. 1403G>A (p. Arg468Gln), a common MPS II mutation, was found as an apparent *de novo* change in her *IDS* gene. No other mutation, deletion, splicing mutation or recombination *IDS-IDS2*, which could explain the manifestation of MPS II in the girl was found. The analyses of the patient's cDNA using Sanger sequencing as well as NGS showed that only the mutated allele was transcribed. This finding led to a suspicion of X-chromosome skewing. Both approaches used for determination of X-inactivation status, DNA-methylation analysis and transcription-based assay, identically showed preferential inactivation of maternal allele (96/4 to 99/1) in patient's blood leukocytes and buccal swabs. Our results confirm that while overwhelming majority of MPS II carriers are asymptomatic, almost exclusive expression of the mutated allele due to extremely skewed X-inactivation leads to manifestation MPS II in females. Support: IGA MZ CR NT 14015, RVO-VFN64165/2012, PRVOUK-P24/LF1/3, GAUK No. 42314, MZ CR - RVO (EU, 00023761).

2290F

Genetic investigation of Maple Syrup Urine Disease in Iranian population. *M. Abiri[1], M. R Alaei[2], S. Dabagh Bagheri[3], H. Bagherian[4], S. Zeinali[4,5].* 1) Medical Genetics, Tehran university of medical sciences, Tehran, Iran; 2) Mohammad Reza Alaei, Pediatric Endocrinology and Metabolism, Mofid Children's Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran,MD; 3) Samira Dabagh Bagheri,Medical Genetics Laboratory, Kawsar Human Genetics Research Center, Tehran, Iran,BSc; 4) Hamideh Bagherian,Medical Genetics Laboratory, Kawsar Human Genetics Research Center, Tehran, Iran, MD; 5) Sirous Zeinali, Department of Molecular Medicine, Biotechnology Research Center, Pasteur Institute of Iran, Tehran, Iran, PHD.

Maple syrup urine disease (MSUD; OMIM 248600) is a rare inborn error of metabolism with the population incidence of 1:185. 000. The disease frequency is higher in population with higher rate of consanguineous marriage. Also, Iran is an immense country with 38. 6% rate of consanguineous marriage. So it is expected to have a higher rate of patients. So in our country early detection of the disease and disease management using special diets or the other forms of therapy is very important. Untreated patients die soon after the birth and if not they may show different physical and mental abnormalities. The disease mainly caused by deficient activity of a mitochondrial enzyme called BCKD (BCKD; EC 1. 2. 4. 4.) complex. It is a branched-chain α-keto acid dehydrogenase complex which catalyze oxidative decarboxylation of branched-chainα-keto acids that derived from branched-chain amino acids (BCAA). Improper activity of the enzyme complex leads to accumulation of leucine, isoleucine and valine and their respective metabolites in cells and body fluid. Phenotypic expression of the disease depends on the amount of BCKAD activity. Different types of mutation have been previously disclosed in *BCKDHA, BCKDHB, DBT* and *DLD* gene in other populations. Since the MSUD mutational spectrum has not previously investigated in Iran, this study aims to investigate the molecular basis and biochemical profile of patients. Here in, we surveyed clinical and biochemical profile of 26 MSUD families. Now, molecular genetic testing of 15 cases is available and the others are still ongoing. In this study, we used 2 sets of polymorphic STR markers linked to the mentioned genes to find the probable pathogenic gene in families quickly and indirectly. STR markers were selected based on their heterozygosity in Iranian population. Then to find the exact pathogenic change, mutation analysis of probable gene/ genes was done by amplification of all exons and flanking intronic sequences, followed by bidirectional sequencing. All identified mutations were novel except one that reported previously in Turkish population. Also our findings show that there is no hot spot mutation in any of the investigated gene and mutations are scattered thought the mentioned genes. *BCKDHA* gene was the most common gene involved in genetic status of Iran. Structural modellings of the novel mutations were done by different available soft wares to predict the severity and probable clinical consequence.

2291T

Unique presentation of LHON/MELAS overlap syndrome caused by very rare mutation in *MTND5*. *T. Honzik[1], P. Liskova[2, 3], M. Tesarova[1], V. Kucerova Vidrova[1], H. Hansikova[1], H. Kolarova[1].* 1) Dpt of Pediatrics, First Fac. Med. , Charles Univ, Gen Univ Hosp, Prague, Czech Republic; 2) Lab of Biol Pat of the Eye, Inst of Inherit Metab Dis, First Fac. Med. , Charles Univ, Gen Univ Hosp, Prague, Czech Republic; 3) Dpt of Ophthalmology, First Fac. Med. , Charles Univ, Gen Univ Hosp, Prague, Czech Republic.

Background: Leber hereditary optic neuropathy (LHON) and mitochondrial encephalopathy, myopathy, lactic acidosis and stroke-like episodes (MELAS) syndromes are mitochondrially inherited disorders characterized by acute visual failure and variable multiorgan system presentation, respectively. *Materials and Methods:* A 12-year-old girl with otherwise unremarkable medical history presented with abrupt, painless loss of vision. Over the next months, she developed moderate sensorineural hearing loss, vertigo, migraines, anhedonia and thyroiditis. Ocular fundoscopic examination confirmed bilateral optic nerve atrophy. Metabolic workup documented elevated cerebrospinal fluid lactate. Initial genetic analyses excluded three of the most common LHON mutations in our patient. Subsequently, Sanger sequencing of the entire mitochondrial DNA (mtDNA) genome was performed. *Results:* Whole mtDNA sequencing revealed a pathogenic heteroplasmic mutation m. 13046T>C in *MTND5* encoding the ND5 subunit of complex I. This particular variant has been previously described in a single case report of MELAS/Leigh syndrome. Based on the constellation of clinical symptoms in our patient, we diagnose the condition as LHON/MELAS overlap syndrome. *Conclusions:* We describe a unique presentation of LHON/MELAS overlap syndrome resulting from m. 13046T>C mutation in a 12-year-old girl. In patients with sudden vision loss in which three of the most common LHON mitochondrial mutations have been ruled out, molecular genetic examination should be extended to other mtDNA-encoded subunits of *MTND5* complex I. Furthermore, atypical clinical presentations must be considered, even in well-described phenotypes. *This work was supported by grants IGA NT 14156/3, IGA NT 13114/4, GAUK 38515/2015 and by institutional research support by Charles University in Prague (PRVOUK P24/LF1/3, RVO-VFN 64165, SVV UK 260148/2015).* .

2292F

Identification of a novel deletion in the *PEX1* gene resulting in partial PEX1 protein function in patients with Zellweger Spectrum Disorder. *E. Di Pietro[1], C. Argyriou[1], P. Saberian[1], S. Birjandian[1], S. J. Steinberg[2], N. E. Braverman[1].* 1) Department of Human Genetics, Research Institute of the MUHC, Montreal, PQ, Canada; 2) 2McKusick Nathans Institute of Genetic Medicine , Johns Hopkins School of Medicine, Baltimore, Maryland.

Zellweger spectrum disorder is a group of autosomal recessive multi-systemic disorders that are caused by defects in any one of 13 PEX genes that encode PEX proteins, or peroxins, required for peroxisome assembly. Clinical severity ranges from severe to milder and is largely correlated to the effect of the PEX gene mutations on the function of the encoded protein. The majority of reported mutations are in the PEX1 gene, and a common mutation, PEX1-G843D, encodes a protein with residual function and an intermediate- milder phenotype. The PEX1 protein is a AAA protein (ATPases Associated with diverse cellular Activities) that forms a hetero-hexamer with the PEX6 AAA protein to generate a peroxisome associated complex that recycles the PEX5 receptor from the peroxisome membrane to the cytosol in order to ensure additional rounds of matrix protein import. Here we report the identification of novel, overlapping terminal deletions in the PEX1 gene in 3 unrelated patients with intermediate-milder ZSD phenotypes. We identified the deletions, c. 3767+334_c. 3852*3356 in two patients and c. 3767+451_c. 3852*5312 in one patient, each in compound heterozygosity with a PEX1 null allele. We characterized the cDNA and the resulting protein, which is truncated at the end of exon 23. The level of PEX1 protein is reduced in patient cell lines. Remarkably, the truncated protein retains residual function. There is residual matrix protein import in patient cell lines and expression of the truncated protein in a PEX1 null cell line resulted in approximately 50% recovery of peroxisomal import as compared with expression of normal PEX1 protein. We show that treatment with Betaine, a nonspecific chemical chaperone, also improved peroxisome assembly in the PEX1 exon 24-deletion cell lines. We conclude that the truncated PEX1 protein is responsible for the milder patient phenotype and that betaine treatment could be considered in patients with this allele.

2293T

International Fabry Disease Genotype/Phenotype Database: Phenotypic Classification of 46 novel pathogenic *GLA* mutations for Type 1 "classic" or Type 2 "Later-Onset" Phenotype by clinical, *in vitro* expression and *in silico* analyses. *D. O. Doheny, C. Sollis-Villa, S. Kadirvel, B. H. Lee, B. Chen, I. Nazarenko, R. Montel, R. Srinivasan, M. Yasuda, C. Yu, R. J. Desnick.* Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

X-linked Fabry disease results from mutations in the α-galactosidase A gene (*GLA*) that lead to the absent or deficient activity of α-galactosidase A (α-Gal A), and progressive accumulation of glycosphingolipids with terminal a-galactosyl moieties. Mutations causing the severe, early-onset Type 1 "classic" phenotype result in ecssentially no α-Gal A activity in affected males and include nonsense, frameshift, consensus splice site, large deletions/insertions and specific missense mutations. Mutations causing the Type 2 "later-onset" phenotype encode residual α-Gal A activity and include alternative splicing and specific missense mutations. Notably, certain common missense mutations are benign or likely benign, including encoding E66Q, SI26G, c. 639+6A>C and D313Y as well as R118C and A143T, previously thought to cause the Type 2 phenotype. From 2005 to 2015, 2912 individuals were tested, including 1056 (36. 3%) probands of which 351 were mutation positive (217 males, 134 females). The 5 most common reported mutations were encoding N215S, R363H, R112C, R227X, and R112H. Noteably, 46 novel *GLA* mutations were identified: 14 (30%) were clearly novel Type 1 mutations including 1 nonsense (Q57X), 6 deletions (c. 59_84del26, c. 234delA, c. 966delC, c. 1193_1196delAATG, c. 1226_1231delCCACAG, and large deletion of exons 2-7,), 3 insertions (c. 147_148insCGC, c. 966insC, c. 1094dupA), 1 double missense mutation (novel R4M and previously reported Y207S), and 3 consensus splice site mutations (IVS3+2T>C, IVS5-1G>A, IVS6+1delG). Mutations causing the Type 2 phenotype included two alternative splicing defects (IVS4+876T>G, IVS4+878T>C). In addition, there were 30 (65%) novel missense mutations encoding (L19Q, L21P, L21R, A37P, E48Q, G80D, Y86H, C90R, C90Y, Y134D, T141N, L167V, R196G, I198T, V199A, S201P, E203D, Y207H, I232S, W245G, A257D, V269E, G274R, M290T, G360R, Q250H, C378S, C382R, P389L, L417R) which could cause the Type 1 or 2 phenotype. To determine the pathogenicity and phenotype of the missense lesions, clinical data, *in vitro* expression assays, and *in silico* structural studies are underway. These studies are essential to predict the clinical phenotype of patients with pathogenic mutations and to identify likely benign lesions, particularly for cascade family testing and newborn screening, and for appropriate therapeutic intervention, including early enzyme therapy in Type 1 classically affected males.

2294F

Molecular and functional characterization of mutations in Congenital Adrenal Hyperplasia patients (CYP21A2 gene): functional genomics. *R. Prasad[1], R. Khajuria[1], R. walia[2], A. Bhansali[2].* 1) Biochemistry, PGIMER, Chandigarh, Chandigarh, India; 2) Endocrinology,PGIMERchandigarh,chandigarh,india.

Congenital Adrenal Hyperplasia (CAH) is autosomal recessive disease, with a wide range of clinical manifestations from severe classical form to late onset form. Our aim was to determine the frequency of common CYP21A2 gene mutations, identify novel mutations and their functional analysis. Clinical and hormonal evaluations were used to categorize the patients in Salt Wasting (SW), Simple Virilizing (SV) and Non Classical (NC) forms. Molecular analysis of CYP21A2 was performed in 55 patients for detection of common mutations viz gene deletion,Q318X, R356W, V281L, i2g, F306 + T or L307 frame shift mutation and P267L. Polymorphisms viz D183E & S268T were identified in 55 patients and 55 controls. Novel mutations were identified by SSCP technique and subsequently sequencing of amplified product. Functional implication of novel mutations was analyzed by construction of mutant plasmids by site-directed mutagenesis and subsequently by their in vitro expression in COS-7 cells. Mutation severity was assessed by calculating enzyme activity and their kinetic constant of respective mutants. Mutation severity prediction softwares namely PROVEAN, SIFT and polyphen were used to predict the effect of novel mutations on 21-hydroxylase enzyme. Disease causing mutations were identified in patients comprising SW (n=14), SV (n=26) and NC(n=15). 8 bp gene deletion was found in 10 patients (18. 1 %) in homozygous condition whereas heterozygous condition prevailed among 5 patients (9%). Frequency of other known mutations were R356W mutation (12. 7 %), Q318x (14. 5 %) ,V281L (12. 2 %), i2g (13. 6 %), F306 + T (12. 7 %) and P267L (5. 4 %). H365N, F306V, P357P, D234D are novel mutations in CYP21A2 gene. Each novel mutation was present at frequency of 1. 8 %. H365N mutant retained 63. 23 % enzyme activity while F306V mutant had 46. 13% activity as compared to wild type enzyme for 17OHP as substrate. H365N mutant had 64. 17 % activity while F306V mutant had 57. 77 % activity as compared to wild type enzyme for progesterone as substrate. Kinetic analysis reveals significant alteration in kinetic constant of mutant to that of wild type. Further, there was significant change in activation energy of mutant to that of wild type. Polyphen-2 predicts that F306V and H365N are probably damaging mutations. PROVEAN assigns H365N and F306V as deleterious mutations. SIFT predicts that both the novel mutations affect protein function as a result of substitution of amino acid at the respective positions.

2295T

Severe Fetal Onset of Glycogenosis type IV (Andersen disease) Presenting as Hydrops Fetalis – Report of a case. *H. Alrukban[1], P. Shannon[2], R. Tietelbaum[3], D. Chitayat[1,3].* 1) Genetics and Metabolics, The Hospital For Sick Childrens, toronto, Canada; 2) Department of Laboratory Medicine and Pathobiology and Department of Obstetrics and Gynecology; 3) The Prenatal Diagnosis and Medical Genetics Program, Mount Sinai Hospital, University of Toronto, Toronto, Ontario, Canada.

Glycogen storage disease type 4 is an autosomal recessive disease caused by mutation in Glycogen branching enzyme (GBE) gene. This enzyme catalyzes the attachment of short glucosyl chains to a naked peripheral chain of nascent glycogen. Deficiency results in abnormal structure, known as polyglucosan that has longer chain lengths and fewer branch points. It accumulates in various body tissues including liver, muscles, nervous system and heart. The clinical presentation involves the classic liver form leading to liver cirrhosis and failure and the rare neuromuscular form. The neuromuscular presentation is variable in severity and onset and includes congenital, juvenile and adult presentations. The congenital type is characterized by antenatal polyhydrominos, hydrops fetalis and decreased fetal movement in the third trimester caused by accumulation of the polyglucosan in the skeletal and neural tissues. Affected babies present with hypotonia, hyporeflexia, cardiac dysfunction, and arthrogryposis. Majority of the cases required immediate resuscitation and respiratory support and died in the neonatal period. We report a male neonate who was born to a consanguineous couple of Indian descent, who had lost a previous baby 30 minutes after his birth with limited information on the cause of death. Their current pregnancy was complicated with hydrops fetalis of an unknown etiology. The baby died shortly after birth despite vigorous resuscitation. On autopsy examination he was found to have a myopathic face with small mouth and high arched palate, subcutaneous edema and bilateral mild talipes. The lungs were hypoplastic and the myocardium of the right ventricle appeared thickened. Microscopic examination demonstrated innumerable polyglucosan bodies in the skeletal muscle, myocardium, liver and central nervous system consistent with GSD4. Sequencing the GBE1 gene showed a homozygous novel missense mutation, c. 321G_A, predicted to result in premature protein termination (p. Trp107*). This finding expands the spectrum of mutations associated with the congenital form. To date total of 38 mutations in the GBE1 gene have been reported with different phenotypes and include missense, nonsense, insertion, deletion and splice site mutations. Majority of the mutations reported with the congenital type cause truncation of the protein, explaining the severity of the phenotype. Our case shows the importance of a thorough autopsy in delineating the etiology of hydrops fetalis.

2296F

Evidence of inflammation in organs of male reproductive tract of the MPS I mouse model. *V. D Almeida[1], C. C. Nascimento[1], O. Aguiar-Jr[2].* 1) Psychobiology, Univ Federal de Sao Paulo, Sao Paulo, Sao Paulo, Brazil; 2) Biosciences, Univ Federal de Sao Paulo, Sao Paulo, Brazil.

Several pathophysiologic mechanisms have been described in mucopolysaccharidoses (MPS) resulting in cellular, tissue and organ damage, demonstrated in patients and animal models. We have previously investigated some male reproductive parameters in MPS I murine model and detected lower sperm production, lower weight of seminal vesicles and testicular histopathological alterations, especially in the interstitial compartment, including high number of vacuolated cells and a higher interstitial area when compared to controls. In order to investigate these alterations in more detail, we evaluated the testicular ultrastructure of 6-month-old *Idua*-/- C56BL mice by transmission electron microscopy and examined the histology of epididymis, seminal vesicles and prostate. In testicular analysis, we detected some cell types that appear to be inflammatory infiltrates completely full of vacuoles, with lysosomal characteristics. Testicular macrophages interact in a paracrine way with Leydig cells and interfere in testosterone production, a potent androgen important for signaling in many tissues, besides reproductive tract. Considering that the observed cell types are indeed infiltrated macrophages, it is possible that testosterone production is unbalanced, which may also explain part of the symptomatology of the disease. Interestingly, these inflammatory infiltrates have also been identified in histological sections of epididymis, seminal vesicles and prostate. Inflammation is the cause of a significant portion of the morbidities and progressive decline suffered by MPS patients. However, the alterations observed in these structures have not been previously described. Cells of the immune system seem to be present in these infiltrates and we hypothesize that macrophages could be responsible for an inflammatory cascade related to damage or dysfunction caused by lysosomal storage in the male reproductive tract.

2297T

A new disorder of O-glycosylation due to a mutation in the X-linked OGT gene is associated with microcephaly, epilepsy, abnormal sleep architecture, intellectual disability and hypothyroidism. *A. Huq, F. Serajee.* Dept Pediatric Neurology, Children's Hosp Michigan, Detroit, MI.

Background: OGT gene on Xq13 encodes a glycosyltransferase that transfers a single N-acetylglucosamine from UDP-GlcNAc to a serine or threonine residue in many cytoplasmic and nuclear proteins. It is involved in cellular processes including chromatin structure, glucose homeostasis, and circadian oscillation of clock genes. Currently mutation in OGT is not known to cause any Mendelian disorder in human. Case Report: We studied a 6-year-old boy with microcephaly, intellectual disability, hypothyroidism, abnormal sleep pattern, nystagmus and epilepsy. His EEG revealed disorganized background and absent sleep spindle in addition to epileptiform waves. Investigations revealed an abnormal oligosaccharide pattern in urine. Exome sequencing was performed using genomic DNA isolated from the patient, parents and a healthy brother. Results: Exome sequencing revealed a hemizygous c. 775G>A (pA259T) mutation in exon 7 of the X-linked gene OGT causing the alanine at amino acid position 259 to be replaced by threonine. Mother was heterozygous for this alteration, while father and healthy brother did not show this change. The alteration is not observed in healthy cohorts, conserved throughout evolution and is predicted to be deleterious by in silico models. The p259 amino acid is located in the sixth tetratricopeptide repeat, important for binding to protein substrates and transfer of N-acetylglucosamine. Conclusion: Our patient's phenotype is consistent with OGT gene function and observed phenotype in mice with altered OGT. That mutations in an interacting protein of OGT (HCFC1) cause microcephaly, developmental delay and intractable epilepsy is also notable. Our patient likely represents a new disorder of O-glycosylation resulting from mutation in OGT gene.

2298F

The common MERRF mutation in a three generation pedigree with multiple symmetric lipomatosis. *S. Seneca[1], N. Revencu[2], J. Smet[3], A. Vanlander[3], K. Stouffs[1], A. Gheldof[1], R. Van Coster[3], L. De Meirleir[4], M. Bonduelle[1].* 1) Center for Medical Genetics, UZ Brussel & REGE, Vrije Universiteit brussel, Brussels, Belgium; 2) Center for Human Genetics, Catholic University of Louvain, Brussels, Belgium; 3) Department of Paediatric Neurology and Metabolism, Ghent University Hospital, Ghent, Belgium; 4) Department of Paediatric Neurology and REGE, UZ Brussel & Vrije Universiteit Brussel, Brussels, Belgium.

Background : Disorders of OXPHOS are among the most common conditions of metabolic disorders with mutations in both the mitochondrial and nuclear genomes underlying these diseases. Mutational hotspot positions of the mtDNA include the recurrent m. 8344A>G MERRF mutation in the *MT-TK* gene. Mitochondrial disorders often present as a multisystem disease with a very heterogeneous phenotype for an identical mutation, even within the same family. **Objective** : We investigated a possible mitochondrial dysfunction in a patient with multiple symmetric lipomatosis (MSL) by assessing the integrity of the mtDNA and investigating OXPHOS complex activities with, respectively, molecular and biochemical assays in several tissues of the index case. **Results** : Molecular investigation of the mitochondrial genome identified the presence of the common m. 8344A>G MERRF mutation in the *MT-TK* (tRNA Lys) gene of leukocytes, skeletal muscle and adipose tissue. Spectrophotometric analysis of muscle tissue revealed a decreased complexIV enzymatic activity. In addition, BN-PAGE analysis documented overall low activity of complexesI, III, and IV, and subcomplexes of complexV. The presence of the m. 8344A>G mutation was also seen in several other members of this pedigree, at least inthree generations. **Conclusions** : Multiple symmetric lipomatosis (MSL), also known as Launois–Bensaude syndrome or Madelung's disease, is a rare disorder predominantly seen in middle-aged male patients. In the past, publications have suggested, among others, an association with mitochondrial dysfunction. Although the typical manifestations of myoclonic epilepsy and ragged-red fibers disease were not present in our index patient, his mtDNA tested positive for the presence of the classic MERRF mutation. Therefore, we suggest that investigation of the mitochondrial genome in patients presenting with a familial history of MSL might be indicated. The common MERRF mutation can cause severe cellular dysfunction with a fatal outcome.

2299T

A distinct type of 3-methylglutaconic aciduria due to a mutation in the Translocase of Inner Mitochondrial Membrane 50 (*TIMM50*) gene. *F. Serajee, A. Huq.* Pediatrics, Wayne State University, Detroit, MI.

BACKGROUND: 3-methylglutaconic aciduria is a heterogeneous group of disorders and include several inborn errors of metabolism biochemically characterized by increased urinary excretion of 3-methylglutaconic acid. Defective leucine metabolism and various disorders affecting mitochondrial function have been identified as causes of conditions in which 3-methylglutaconic aciduria is a distinctive feature. We report a homozygous mutation in mitochondrial TIMM50 gene in a family with 3 affected siblings with a distinctive type of 3-methylglutaconic aciduria. METHODS: We investigated a patient of South Asian ancestry with intractable epilepsy, microcephaly, developmental delay, visual deficit spastic quadriplegia. Metabolic testing revealed large amount of 3-methylglutaconic acid in urine. Parents were consanguineous and two earlier-born sisters of the patient had a similar phenotype, 3-methylglutaconic aciduria and unexpected death after an infection during a visit to their country of origin. Detailed clinical history, imaging, EEG and metabolic testing were obtained for all affected persons. Full exome sequencing was performed using genomic DNA isolated from one surviving patient, two healthy siblings and both parents. RESULTS: Exome sequencing identified a homozygous c. 1114G>A (p. G372S) mutation in the gene TIMM50. There were no other candidate alterations in exome that could explain the phenotype in the proband. The G372 amino acid position in TIMM50 gene is completely conserved in eukaryotes all the way from the yeast Saccharomyces cerevisiae to humans. The p. G372S alteration is predicted to be probably damaging and deleterious by PolyPhen and SIFT in silico analyses, respectively. The p. G372S alteration is located in the conserved C-terminal domain of the Tim50 protein that interacts with the N-terminal domain of the Tim23 protein in the intermembrane space and regulates mitochondrial protein import of presequence-containing polypeptides Both parents are heterozygous. The patient's healthy sister is homozygous normal and her healthy brother is heterozygous. The segregation pattern is consistent with the TIMM50 gene mutation being the etiology of patient's 3-methylglutaconic aciduria. CONCLUSION: TIMM50 gene mutation results in a novel mitochondrial membrane associated disorders with 3-methyl glutaconic aciduria.

2300F

Molecular genetic study and urine analysis of Japanese patients with cerebral creatine deficiency syndromes. *H. Shimbo[1], H. Osaka[2], M. Tachikawa[3], S. Otsuki[4], S. Ito[4], T. Goto[5], Y. Tsuyusaki[5], N. Aida[6], K. Kurosawa[7], Y. Kurosawa[8], H. Kato[9], K. Takano[10], T. Wada[11]*. 1) Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Kanagawa, Japan; 2) Department of Pediatrics, Jichi Medical University, Tochigi. Japan; 3) Graduate School of Pharmaceutical Sciences, Tohoku University, Sendai, Japan; 4) Department of Pharmaceutical Microbiology, Kumamoto University, Kumamoto, Japan; 5) Department of Neurology, Kanagawa Children's Medical Center, Yokohama, Japan; 6) Department of Radiology, Kanagawa Children's Medical Center, Yokohama, Japan; 7) Department of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan; 8) Graduate School of Sport and Health Science, Ritsumeikan University, Kyoto, Japan; 9) Kanagawa Psychiatric Center, Yokohama, Japan; 10) Department of Medical Genetics, Shinshu University School of Medicine, Nagano, Japan; 11) Department of Medical Ethics and Medical Genetics, Kyoto University Graduate School of Medicine, Kyoto, Japan.

Introduction: Cerebral creatine deficiency syndromes (CCDS) are caused by genetic defects in L-arginine:glycine amidinotransferase (AGAT), guanidinoacetate methyltransferase (GAMT) , and creatine transporter 1 (encoded by *SLC6A8*). SLC6A8 deficiency leading cause of CCDS, is estimated to account for 1-2percentofmaleswith ID. The common clinical features of CCDS are intellectual disability (ID), speech delay, autistic behavior, and seizure. CCDS can be diagnosed bymeasurement of urinary creatine (CR), guanidinoacetic acid (GA), and creatinine (CN) in urine, decreased peak of creatine in magnetic resonance spectroscopy (1H-MRS) of thebrain, and molecular genetic analysis. Three types of CCDS can be differentiated from the abnormal ratio of CR to CN, or GA to CN in urine. Recently we have established a simple high performance liquid chromatography (HPLC) method to quantify CR, GA, and CN in urine and diagnosed 6 families with CCDS in Japan. We reviewed identified mutations in *SLC6A8*of 10 patients from 5 families and*GAMT*of one patient, and the results of their urinary analysis. **Methods:**The separation of CR, GA, and CN in urine was performed on a ion chromatography column with UV detection. Genomic DNA and RNA extracted from the leucocytes of patients, were subjected to PCR and Sanger sequencing. **Results:**CR, GA, and CN were separated clearly using ion chromatography column. We succeeded to detect 5 families with SLC6A8 deficiency and one patient with GAMT deficiency. **Conclusion:**HPLC method would contribute to early diagnosis and intervention for patients with CCDS. CCDS, especially AGAT or GAMT deficiency, can be treatable with creatine, andSLC6A8 deficiencymay be treatable with cyclocreatine in the near future. Most patients withSLC6A8 deficiency appear to remain undiagnosed in Japan, and the systematic screening system should be established using this method.

2301T

Characterizing the genetic basis of Leigh Syndrome, the most common mitochondrial disorder affecting children. *N. J. Lake[1,2], A. G. Compton[1,2], S. E. Calvo[3,4,5], V. K. Mootha[3,4,5], D. R. Thorburn[1,2,6]*. 1) Murdoch Childrens Research Institute, Melbourne, Victoria, Australia; 2) Department of Paediatrics, University of Melbourne, Australia; 3) Howard Hughes Medical Institute and Department of Molecular Biology, Massachusetts General Hospital, Boston, USA; 4) Department of Systems Biology, Harvard Medical School, Boston, USA; 5) Broad Institute of Harvard and MIT, Boston, USA; 6) Victorian Clinical Genetic Services, Royal Children's Hospital, Melbourne, Australia.

Leigh syndrome (LS) is the most common pediatric clinical presentation of inherited mitochondrial energy generation disorders. LS is genetically heterogeneous, and can be caused by mutations in over 70 different genes. Here we present an update on 67 patients from a previously ascertained cohort, comprised of 35 patients with definite LS and 32 Leigh-like patients, diagnosed according to stringent diagnostic criteria. At the time of publication in 1996[1], the genetic basis of ~25% (18 patients) was known. Using contemporary technologies, including massively parallel sequencing (MPS), we have now established the genetic basis of disease in 75% of the cohort (50 patients), including in 34 of the 35 LS patients and 16 of 32 patients with Leigh-like syndrome. Pathogenic mutations in seventeen different disease genes and a mitochondrial DNA deletion were identified. Our results reveal multiple novel pathogenic mutations in known LS disease genes including 3 in *SLC19A3*, and expand the phenotypic spectrum of known LS disease genes such as *SERAC1*. Furthermore we expand the phenotypic spectrum of known mitochondrial disease genes to include LS such as *C10orf2*. The mutations were inherited in a maternal (46%), autosomal recessive (42%) or X-linked (12%) manner. The identified disease genes encoded proteins associated with the activity of oxidative phosphorylation (OXPHOS) complexes I (26%), IV (18%), V (22%) or pyruvate dehydrogenase (22%). Molecular defects in genes associated with a combined OXPHOS deficiency were observed in 12% of patients. We also show examples where routine approaches would have missed the genetic diagnosis or resulted in miscalling of the causative mutations, and illustrate how careful analysis of MPS data can achieve a genetic diagnosis where routine filtering failed to identify any candidates. Overall, our results highlight the significant genetic heterogeneity underlying LS, and provide key insight into the etiological basis of disease in a large well characterized cohort. Evaluation of diagnostic outcomes from the application of MPS will facilitate the development of diagnostic clinical exome testing in LS patients. 1. Rahman S, Blok RB, Dahl HH, et al. Leigh syndrome: clinical features and biochemical and DNA abnormalities. Ann Neurol. 1996 Mar;39(3):343-51.

2302F

***PUS1* and *COX10* mutations in three Czech patients with cytochrome c oxidase deficiency and hematological symptoms.** *M. Tesarova[1], A. Vondrackova[1], H. Kratochvilova[1], M. Rodinova[1], V. Dvorakova[1], V. Stranecky[2], J. Spacilova[1], H. Hansikova[1], J. Zeman[1], T. Honzik[1].* 1) Dept. Pediatrics & Adolescent Med, Charles Univ Prague, Fac Med & Gen Univ Hosp, Prague 2, Czech Republic; 2) Institute of Inherited Metabolic Dis, Charles Univ Prague, Fac Med & Gen Univ Hosp, Prague 2, Czech Republic.

Anaemia represents a heterogeneous group of hematologic diseases, which can be induced by plenty of genetic and risk factors. Although the association of mitochondrial dysfunction and hematologic pathology is generally known, the underlying pathologic factors with regard to evolving anaemias are poorly characterized. In this study, we report on three patients, who manifested profound cytochrome c oxidase deficiency combined with anaemia. Genetic causes of their OXPHOS deficiency were found with the use of targeted sequencing of mitochondrial exome and SNP microarray analysis. A rarely occurring pathological 6-kbp homozygous deletion was identified in two unrelated patients affecting the *PUS1* gene, which remarkably leads to different disease-phenotype in both patients. Two previously characterized deleterious missense sequence variations of *COX10* gene (p. Asn204Lys; p. Pro225Leu) were identified in the third patient, however, their combination has not been reported yet, which may imply the variant patient disease-phenotype. Based on our results and current knowledge, we suggest the infantile deficiency of *PUS1* and *COX10* to be classified as the early fatal and slowly progressive forms. To conclude, mitochondrial disorders manifest poor phenotype-genotype correlation even in the patients with the same causal mutations. *This study was supported by grants RVO-VFN64165/2012, IGA NT13114-4, IGA NT14156–3 and GAČR 14-36804G.* .

2303T

Clinical and molecular characterization of glutaric aciduria type 1 patients from India. *P. M. Tamhankar[1], L. Vasudevan[1], S. Thomas[2], S. Niazi[2], R. Christopher[3], D. Solanki[4], P. Dholakia[4], M. Muranjan[5], M. Kamate[6], P. Singhi[7], N. Sankhyan[7], U. Kalane[8], J. Sheth[9], V. Chennuri[10], V. Joshi[11], S. Patil[12], P. Gadgil[12], R. Gulati[13], M. Vasikarla[14], C. Datar[15], R. Cariappa[16], S. Jagadeesh[17], S. Danda[18].* 1) Genetic Research Center, National Institute for Research in Reproductive Health, Mumbai, Maharashtra, India; 2) Department of Bio-informatics, National Institute for Research in Reproductive Health, Jehangir Merwanji Street, Parel, Mumbai, Maharashtra, India; 3) Department of Neurochemistry, National Institute of Mental Health and Neuro Sciences, Bangalore, Karnataka, India; 4) Mantra Child and Neurology and Epilepsy Clinic, Bhavnagar, Gujarat, India; 5) Department of Pediatrics, Seth GS Medical College and KEM hospital, Mumbai, Maharashtra, India; 6) Department of Pediatric Neurology, KLE University's J N Medical College, Belgaum, Karnataka, India; 7) Pediatric Neurology and Neurodevelopment Unit, Department of Pediatrics, Post Graduate Institute of Medical Education and Research, Chandigarh, India; 8) Pediatric Neurology Centre, KEM Hospital, Pune, India; 9) FRIGE's Institute of Human Genetics, FRIGE House, Satellite, Ahmedabad, Gujarat, India; 10) Department of Pediatrics, Employee State Insurance Scheme Hospital, Thane, Maharashtra; 11) Navigene Genetic Science, Mira Road, Thane, Maharashtra, India; 12) Department of Neurology, Kokilaben Dhirubhai Ambani Hospital and Research Center, Mumbai, Maharashtra, India; 13) Department of Pediatrics, Jawaharlal Institute of Postgraduate Medical Education and Research, Pondicherry, India; 14) Fernandez Hospital, Hyderabad, Andhra Pradesh, India; 15) Sahyadri Medical Genetics, Pune, Maharashtra, India; 16) Department of Medical Genetics, Christian Medical College, Vellore, Tamil Nadu India; 17) Neogen Labs Pvt Ltd, Bangalore, Karnataka, India; 18) Fetal Care Research Foundation, MediScan, Chennai, Tamil Nadu, India.

Glutaric acidemia type 1 (GA-1, OMIM 231670) is an autosomal recessive inborn error of metabolism caused by the deficiency of glutaryl-CoA dehydrogenase. We present a retrospective analysis of clinical profile and mutation analysis in twenty six unrelated families with affected children (27 cases total). The mean age at onset of illness was 9. 7 months (+/-11. 07) whereas the mean age at referral for molecular diagnosis was 35. 81 months (+/- 32. 9). Patients were referred from different states of India (Karnataka, Maharashtra, Tamil Nadu, Gujarat, Andhra Pradesh, Uttar Pradesh, Punjab and Jammu and Kashmir). Clinical presentation varied from asymptomatic, acute encephalitis followed by neuro-regression and chronic developmental delay. Macrocephaly was present in 16/27 cases, dystonia was present in 23/27 cases. Patients underwent blood tandem mass spectrometry and/or urine gas chromatography mass spectrometry. Neuroimaging demonstrated batwing appearance in 95% cases. Sequencing of GCDH gene was done covering all exons and exon-intron boundaries. Mutations could be identified in nineteen families, whereas no mutations could be identified in four families. Molecular analysis is ongoing in four families. Mutations identified include novel ones such as p. P217S, p. L179V, p. L345P, c. 1167delG, p. K377R, p. M100T, p. T344T, p. W225X, p. G269S, p. H403Y, p. L424P, p. S255P, p. N373K and known mutations such as p. R402W, p. L179R, p. M339V, p. A433V, p. R402Q, c. 1173delG and p. R128Q. Using *in silico* analysis, mutations were shown to be affecting the residues responsible for homotetramer formation of the glutaryl coA dehydrogenase protein. This is the largest mutation proven series of glutaric aciduria type 1 from India till date.

2304F

A synonymous splice site variant in *MTHFR* gene causing hyperhomocysteinemia with severe neurological dysfunction that was successfully reversed with early initiation of medical therapy. *D. Reed[1], H. Dai[2], A. Conner[1], J. Zhang[2], L. Wong[2], K. Oishi[1].* 1) Genetics & Genomic Sciences, Mt. Sinai School of Medicine, New York, NY; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Introduction: Methylenetetrahydrofolate reductase(MTHFR) is an enzyme encoded by the *MTHFR* gene to form 5-methyltetrahydrofolate, a cofactor for methionine synthase. Loss of function mutations cause a rare autosomal recessive severe neurological dysfunction characterized by sudden respiratory failure, coma, neonatal seizures and death. We identified a homozygous synonymous splicing variant in *MTHFR* in an infant with sudden severe neurological depression. **Case:** The patient was a 3 week old female born to consanguineous Pakistani parents. At 17 days, she had a sudden onset of lethargy, poor feeding and cyanosis. She was intubated in the ER and at 20 days old started having seizures. She was non-dysmorphic, edematous and comatose with flaccid hypotonia, absent reflexes and a markedly distended bladder. Biochemical analyses demonstrated elevated plasma homocysteine and decreased methionine levels with absence of methylmalonic acid accumulation, consistent with hyperhomocysteinemia. SNP microarray did not show any microdeletions or duplications. Treatment for presumptive MTHFR deficiency with folinic acid, vitamin B6 and hydroxycobalamin was started. Within days her neurological function was fully recovered. **Methods:** Genomic DNA and total RNA were isolated. Next generation and Sanger sequencing were performed. RT and Real-time PCR were used to verify the transcription patterns of *MTHFR* in the proband and her parents. **Results:** We identified a homozygous synonymous splicing variant c. 1530G>A (p. Lys510Lys). With RT-PCR a significant reduction of wild type transcript in the parents and abolished expression in the proband was seen. The exon 9 skipping form was the major transcript in the proband and was absent in the control. A transcript with a 5bp GTGTG insertion at the 3' end of exon 9 and a novel shorter transcript skipping exon 9 and 10 was identified. Her heterozygous parents expressed normal and truncated transcripts. The variant was seen not in the region of homozygosity on microarray, indicating a possible founder effect for this variant. **Conclusions:** We identified a homozygous c. 1530G>A variant in the *MTHFR* gene in a newborn with hyperhomocysteinemia and sudden severe neurological dysfunction. The variant results in much higher expression of abnormal transcripts that may lead to non-functional MTHFR protein. Rapid initiation of treatment can reverse the severe neurological presentation of MTHFR deficiency and is essential for improving the prognosis.

2305T

Prevalence and severity of non-motor manifestations associated with parkinsonism in patients with Gaucher disease and *GBA1* mutation carriers. *G. Lopez[1], J. Kim[1], D. Cintron[3], E. Wiggs[2], C. Groden[1], E. Sidransky[1].* 1) NHGRI/NIH, Bethesda, MD; 2) NINDS/NIH, Bethesda, MD; 3) PHSU, Ponce, PR.

Objective: To identify the prevalence and severity of non-motor symptoms in a group of patients at risk for developing Parkinson disease (PD). **Background:** The link between mutations in the glucocerebrosidase gene (*GBA1*) and the risk for the development of parkinsonism is well-established in the literature. The prevalence of non-motor symptoms associated with sporadic PD that may be present years before motor manifestations has not been explored in subjects affected with Gaucher disease or carrying *GBA1* mutations without clinical evidence of parkinsonism. **Design/methods**: Patients with Gaucher disease or mutation carriers with a family history of Parkinson disease were longitudinally evaluated at the National Institutes of Health, National Human Genome Research Institute Genetics clinic. A series of validated scales assessing depression, fatigue, restless legs symptoms, olfactory function, anxiety, and cognition were administered during their evaluation. Complete physical and neurological evaluations were done to confirm the absence of parkinsonian motor features. **Results:** Sleep disturbances were more frequently reported by patients with Gaucher disease and mutation carriers compared to unaffected individuals, but less frequently than by patients with parkinsonism. Interestingly, anxiety scores were high in subjects with Gaucher Disease and mutation carriers. **Conclusions:** Continued longitudinal evaluation of patients with Gaucher disease and carriers of *GBA1* mutations may help identify a pre-motor phase in this cohort at risk of developing parkinsonism. This pre-motor profile may suggest specific neurotransmitters involved in neurodegeneration and contribute to a better understanding of the association between lysosomal dysfunction and alpha-synuclein aggregation.

2306F

Cognative assessment of patients with type 3 Gaucher disease. *E. Wiggs[1], G. Lopez[1], J. Kim[1], O. Goker-Alpan[1,2], R. Schiffmann[1,3], E. Sidransky[1].* 1) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD; 2) Baylor Institute of Metabolic Disease, Dallas Tx; 3) O&O Alpan, LLC. Fairfax, VA.

BACKGROUND: Neuronopathic Gaucher disease presents with diverse neurological manifestations including cognitive impairment. Standardized tests of intellectual ability were individually administered to monitor change over time, as well as to probe for patterns between subjects and within subjects. METHOD: Twenty-seven patients (13 males) with type 3 Gaucher disease were assessed using either a Wechsler IQ scale (N=25) or a structured interview of the parent when the patient was either too young or too impaired to respond accurately to the individual IQ test (N=2). Twenty-one had repeated follow-up. All patients were being treated with enzyme replacement at the time of assessment. RESULTS: At the initial evaluation, the mean Full Scale IQ was 74. 26 +/- 25. 13 (N=25) (mean= 100 +/- 15) with a range from 40 (profound intellectual deficiency) to 124 (superior intellectual ability). Four of the 25 subjects had Full Scale IQ scores of 100 or higher. Subsequent evaluations completed 2-9 years later, revealed a mean of 78. 1 +/- 25. 9; an insignificant difference. One child had seven evaluations over a ten year period and the scores remained stable throughout this interval. There was no correlation between genotype and IQ scores, and among 20 patients known to have genotype L444P/L444P, Full Scale IQ scores ranged from 40-124. However, differences between verbal ability and nonverbal ability were identified, which were most pronounced among patients with average or near average IQ scores. Of the nine with Full Scale IQ scores over 90 (25th percentile), eight had lower nonverbal ability (mean difference = 13. 5 with a range of 4 to 38 points). DISCUSSION: Cognition, as reflected by IQ testing, was quite variable among subjects with type 3 Gaucher disease ranging from individuals with superior intellectual ability to others with profound intellectual deficiency. The degree of intellectual impairment did not correlate with the patient's age or genotype. Results tended to remain stable over the time interval studied and were not affected by enzyme replacement therapy. The observed discrepancy between verbal and nonverbal abilities identifies a potential learning issue that requires the attention of educators and caregivers.

2307T

Prospective Natural History Study of Mucopolysaccharidosis Types IIIA and B (Sanfilippo). *K. L. McBride[1, 2], K. Bain[2], K. Kunkler[3], K. McNally[4], R. Grimes[4], N. Zumberge[5], L. Martin[5], S. C. Aylward[6], L. Alfano[7], K. Berry[7], L. Lowes[7], M. Corridore[8], C. McKee[8], D. A. McCarty[3], H. Fu[3], K. M. Flanigan[2, 3, 7].* 1) Center for Cardiovascular and Pulmonary Research, Nationwide Children's Hospital, Columbus, OH; 2) Department of Pediatrics, Ohio State University, Columbus OH; 3) Center for Gene Therapy, Nationwide Children's Hospital, Columbus, OH; 4) Center for Biobehavioral Health, Nationwide Children's Hospital, Columbus, OH; 5) Department of Radiology, Nationwide Children's Hospital, Columbus, OH; 6) Section of Neurology, Nationwide Children's Hospital, Columbus, OH; 7) Department of Physical Therapy, Nationwide Children's Hospital, Columbus, OH; 8) Department of Anesthesiology, Nationwide Children's Hospital, Columbus, OH.

Mucopolysaccharidosis (MPS) type III (Sanfilippo) is a group of 4 enzyme deficiencies leading to tissue accumulation of heparin sulfate. CNS disease is prominent with initial normal development then cognitive decline, hyperactivity, and death. Somatic findings are modest. This study will prospectively assess individual longitudinal disease progression in MPS IIIA and IIIB to define suitable outcomes for future gene transfer trials. Inclusion criteria are age >2 yrs with confirmed MPS IIIA or B and neurologic dysfunction. Planned visits are at 0, 6, and 12 months. 15 IIIA (M=9, F=6; age 5. 0 +/- 1. 9 y) and 10 IIIB subjects (M=8, F=2; age 8. 7+/- 3. 2 y) have completed baseline evaluation; 10 total have finished 12 months. Anesthesia was administered safely >30 times for MRI. Baseline tests showed declines in Leiter, Mullen, and Vineland scales and the Pediatric Evaluation of Disability Inventory as a function of age. Skills peaked at 2. 5-3. 0 year level, with a slight continued mobility gain at early ages before decline. Several subjects had a measure floor effect. The Child Behavior Checklist had scattered results with little change from 0 to 6 months. The 10 meter, six minute, and four-stair walk proved difficult to administer reliably. Actigraphy data suggests reduced sleep efficiency and increased time awake. Serum ALT is high and increases with age. Urine GAG levels decline steeply as a function of age, approaching normal in older subjects. Baseline brain MRI changes (observed as young as 3 yrs) included cerebral volume loss, primarily of white matter. MRS showed decreased N-acetylaspartate:creatine ratios in white and grey matter, with a reduction in choline to creatine ratios in grey matter. Average MRI liver volume was 2. 2X normal, and spleen size was 1. 9X normal. CSF protein rose with age, and may reflect BBB damage with protein leakage or inflammatory proteins. Preliminary data analysis suggests several outcome measures could be used in a gene transfer study, including cognitive function by the Leiter and Mullen, adaptive functions via the Vineland parent interview; liver and spleen size by MRI, and MRI spectroscopy markers. The utility of the CBCL as an assessment of change over time is uncertain. 6MW tests will not be useful in this population. Urine GAG decline with age will need to be incorporated into efficacy measures to avoid incorrect interpretation, and increased serum ALT and CSF protein will need to be accounted for in safety monitoring.

2308F

Autopsy findings in Gaucher disease indicate variability in response to therapy. *D. K. Borger[1], G. Lopez[1], J. Kim[1], M. Allgaeue[2], H-W. Wang[2], E. Sidransky[1].* 1) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD. , Select a Country; 2) Laboratory of Pathology, NCI,NIH. Bethesda, MD.

Gaucher disease (GD), the inherited deficiency of the lysosomal enzyme glucocerebrosidase, exhibits considerable phenotypic diversity. Over the past two decades many patients with GD have been treated with enzyme replacement therapy (ERT), intravenous infusions of recombinant glucocerebrosidase. Clinically, ERT results in significant improvements in the visceral and boney manifestations, but there have been few pathologic studies documenting reversal of disease. We reviewed the pathologic findings of five adult patients with GD where full autopsy studies were performed. All five were referred to the NIH Clinical Center, and autopsies were performed primarily because of the development of parkinsonian manifestations. The five patients included four females and one male and age at diagnosis of GD disease ranged from 8-47 years, with most diagnosed by age 19. Four underwent splenectomy prior to the availability of ERT. The age at death was between 53-70 years (mean 60 years). All were diagnosed by deficient levels of glucocerebrosidase as well as by genotype, which included two patients with N370S/c. 84insG, and one each with N370S/L444P, N370S/N370S and D409H/L444P. All five patients received ERT for 4-22 years in doses between 60-400IU/kg/month. Remarkably, at autopsy two patients had no evidence of any Gaucher cells, the characteristic pathologic finding in GD, in the liver, spleen, bone marrow or lungs. These patients did not have the longest duration or highest doses of ERT. One patient had a relatively low disease burden, with occasional Gaucher cells in the bone marrow, but none in the liver, lungs or GI tract. The fourth patient was also noted to have a moderate disease burden, with occasional Gaucher cells seen in the liver and leptomeninges. Interestingly, the fifth patient, who received the longest duration (22 years) and highest doses of ERT (up to 100IU/Kg/week prior to her death) had significant Gaucher pathology including extensive involvement of her liver, bone marrow and lungs. Furthermore Gaucher cells were found in her pancreas, small intestine, adipose tissue of the breast, lymph nodes, and eyes. The autopsy studies demonstrate a wide range of disease involvement in treated patients, not specifically associated with age, genotype, duration and dose of therapy or splenectomy. Such studies will be important for the identification of genetic or environmental modifiers impacting disease pathogenesis, burden and response to therapy.

2309T

Erythropoietic Protoporphyria (EPP) and X-Linked Protoporphyria (XLP): Clinical, Biochemical and Molecular Characterization of North American Patients. *M. Balwani[1], H. Naik[1], K. E Anderson[2], D. M Bissell[3], J. Bloomer[4], H. L Bonkovsky[5], J. D Phillips[6], J. Overbey[7], I. Peter[1], R. J Desnick[1], Porphyrias Consortium of the Rare Diseases Clinical Research Network.* 1) Genetics and Genomic Science, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Department of Preventive Medicine and Community Health, University of Texas Medical Branch, Galveston, TX,; 3) Department of Medicine, University of California, San Francisco, CA; 4) Department of Medicine, University of Alabama, Birmingham, AL; 5) Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC; 6) Department of Internal Medicine, University of Utah, Salt Lake City, UT; 7) Department of Population Health Science and Policy, Icahn School of Medicine at Mount Sinai, New York, NY.

EPP and XLP are inborn errors of heme biosynthesis with the same phenotype, but resulting from autosomal recessive ferrochelatase (*FECH)* gene mutations or erythroid-specific 5-aminolevulinate synthase (*ALAS2*) gene mutations, respectively. They are characterized by severe, painful photosensitivity and markedly elevated erythrocyte protoporphyrin (PP) levels. Liver dysfunction/liver failure can occur in ~5% of patients. Here, we present clinical, biochemical and genetic features in 211 (190 EPP, 21 XLP) North American patients. Biochemically and/or mutation confirmed patients were enrolled into the Longitudinal Study of the Porphyrias [clinicaltrials. gov: NCT01561157] from 2010 to 2015. Median age of the patients was 36 years (range: 3-77 years; males: females, 1:1) and 94 % were Caucasian. Median age of symptom onset for EPP patients was 3 years (n=149) and XLP males was 2. 3 years (n=6). 61% of EPP patients and all XLP males reported symptom onset in ≤30 minutes of sun exposure. XLP females (n=11) had variable degrees of sun sensitivity, with 5 reporting symptoms ≤30 minutes of sun exposure and others with no symptoms. 45% of patients reported 3-10 painful phototoxic episodes in a year with a majority of patients taking 1-3 days to recover. Most described the phototoxic pain as “burning” (95%), “tingling” (82%) and/or “itching” (80%). Elevated liver function tests were reported in 30% of patients. Two patients had liver transplants. Anemia was noted in 47% of EPP, and 30% and 78% of XLP male and female patients, respectively. Of the 181 patients genotyped, all XLP patients had an *ALAS2* mutation and 158 had a *FECH* mutation in *trans* with the common low expression allele; 4 patients did not have an identifiable mutation in *FECH* or *ALAS2*. *FECH* mutations included missense (28%), deletions (27%), consensus splice site (23%), and nonsense (17%) lesions. Mean erythrocyte PP levels were significantly higher in XLP males (mean 4251 ug/dL) as compared to EPP patients (mean 1759 ug/dL). XLP females had markedly variable PP levels (range 201 to7544 ug/dL) consistent with clinical symptoms. Baseline data from EPP and XLP patients shows that most patients presented clinically with acute cutaneous photosensitivity in early childhood; the most common symptom being burning pain. Higher PP levels were associated with decreased time to burn and increased risk of liver dysfunction indicating that PP levels are a major determinant of disease severity in these patients.

2310F

Novel report of phosphoserine phosphatase deficiency presenting in an adult. *H. M. Byers[1], R. L. Bennett[1], E. Malouf[1], M. D. Weiss[2], J. Feng[1], C. R. Scott[1], S. Jayadev[1,2].* 1) Division of Medical Genetics, University of Washington, Seattle, WA; 2) Department of Neurology, University of Washington, Seattle, WA.

Serine is a non-essential amino acid that plays a vital role in proper development and functioning of the central nervous system. Serine deficiency syndrome consists of a group of autosomal recessive, inborn errors of metabolismresulting from a deficiency in ofone threeenzymes in the serine biosynthesis pathway: phosphoglycerate dehydrogenase (PHGDH), phosphoserine aminotransferase (PSAT) or phosphoserine phosphatase (PSPH). Serine deficiencycan cause microcephaly, intellectual disability, seizures, and psychomotor retardation in children and severe axonal neuropathy in the adult. Of critical importance to clinical care of these patients, serine deficiency syndrome is treatable. Although the first case of phosphoserine phosphatase deficiency (MIM #172480) was described nearly twenty years ago, only eight cases have been reported, all in children. We describe the novel presentation of phosphoserine phosphatase deficiency in an adult, expanding the currently understood phenotype. The patient is a 37-year-old woman with lifelong intellectual disability, childhood onset epilepsy (absence seizures) and microcephaly. At age twenty, she developed lower extremity hypertonia, abnormal gait and progressively severe axonal neuropathy. Contractures in her hands and fingers were noted from age thirty-two. Clinical presentation was concerning for metabolic neuropathy. Electromyogram and nerve conduction studies confirmed axonal neuropathy. Brain magnetic resonance imaging showed no abnormalities. Plasma amino acid evaluationshowed markedly low serine and moderately low glycine. Molecular testing revealed two novel, compound heterozygous mutations in phosphoserine phosphatase (*PSPH*, NM_004577. 3)*:* c. 131T>G (p. Val44Gly) and c. 421G>A (p. Gly141Ser). Treatment with oral serine resulted in improvement of plasma amino acid laboratory values and a subjective increase in energy. This is the first report of phosphoserine phosphatase deficiency presenting in an adult.

2311T

Pancreatitis and Maple Syrup Urine Disease. *J. Duis[1], A. Hamosh[1], C. Koerner[1], A. Pipitone[1], V. Kumbhari[2], G. Maegawa[1], D. Valle[1], M. Gunay-Aygun[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins, Baltimore, MD; 2) Department of Medicine, Gastroenterology, Johns Hopkins, Baltimore, MD.

Maple syrup urine disease (MSUD) is an autosomal recessive disorder of branched-chain amino acid degradation caused by deficiency of the branched-chain α-keto acid dehydrogenase complex. Classically, MSUD presents in the neonatal period with poor feeding, irritability, lethargy, and progressive neurological deterioration. Treatment consists of a diet restricted in natural, branched-chain amino acids (BCAA) containing protein, and supplementation of non-BCAA, valine and isoleucine. Individuals with the disorder present with acute metabolic decompensation in the setting of increased catabolism. With the implementation of newborn screening, early detection of children with MSUD likely will redefine what is known about the natural history and complications of the disease. Rarely, pancreatitis, defined as the presence of the reversible inflammation within the pancreatic parenchyma, is reported as a complication of MSUD. Unlike the organic acidemias, in which pancreatitis is a more clearly described complication, in MSUD, the experience with this complication is more varied. Some report this complication at day two to three of hospitalization in the setting of improved metabolic control after an acute presentation with markedly elevated leucine levels. Reports in the literature are varied and include patients with chronic pancreatitis, acute hemorrhagic pancreatitis, and pancreatitis that developed during an inpatient hospital stay for gastroenteritis. We report a case of a 13-year-old male with MSUD with recurrent pancreatitis in the setting of ansa pancreatica, a rare anatomical variant of pancreatic ducts. We draw on our experiences from his multiple recurrences including leucine levels at the time of presentation, clinical presentation, imaging, and lab findings to help understand pancreatitis as a complication of MSUD. In addition, we discuss other patients with MSUD who have presented to our center with pancreatitis and compare these experiences to further characterize pancreatitis in this patient cohort. We discuss the possible underlying mechanisms and attempt to develop an understanding of the circumstances under which patients with MSUD are likely to develop this complication to optimize clinical care of these patients.

2312F

Phenotypes and genotypes of 35 cases with betaketothiolase deficiency in a Vietnamese referral center. *K. Nguyen[1], D. Vu[1], H. Nguyen[1], S. Yamaguchi[2], T. Fukao[3].* 1) Department of Endocrinology, Metabolism, Genetics, National Hospital of Pediatrics, Hanoi, Viet Nam; 2) Shimane University School of Medicine, Japan; 3) Graduate School of Medicine, Gifu University, Japan.

Betaketothiolase (T2) deficiency is rare inherited metabolic disease worldwide. But it is the most common organic aciduria in Vietnam with 35 cases. *Objectives:* to describe phenotypes, genotypes of T2 deficiency. *Methods:* descriptive study of 35 patients with T2 deficiency at National Hospital of Pediatrics - Hanoi - Vietnam from 2005 to 2012. 21 cases from 19 families were analyzed of *T2* gene. *Results:* 35 patients were born to 30 unrelated families and unconsanguinity parents. 33/35 patients presented crisis of ketone acidosis. One case was diagnosed without symptom at 3 days of age but developed the crisis at 6 months of age. One case was asymptomatic until now (5 years old). Mean onset of the first crisis was 13. 1 months. Clinical features of the crises were dehydration, tarchypnea, lethargy/coma and triggered by infections. Laboratory of the crises showed 100 % severe ketone metabolic acidosis (pH: 6. 5 - 7. 05), leukocytosis. 94% cases recovered from the 1st crisis and 79% cases had normal development. 20/35 patients had recurrent crisis. 21/21 patients were found homozygous/compound heterozygous mutations of *T2* gene. Five different mutations have been identified. And R208X is the most common mutation in Vietnam (70% of mutant allele) as well as in the world. *Conclusions:* The biggest T2 deficiency patients were identified in a Vietnam center (account 1/3 total cases in the world). Worldwide, there are no common mutations of *T2* gene found except common R208X mutation in Vietnam.

2313T

Mice lacking Cox10 in T-lymphocytes recapitulate the immune phenotype of a cohort of patients with primary mitochondrial disease. *P. J. McGuire[1], S. Pacheco[2], M. K. Koenig[2], E. Barca[3], S. DiMauro[3], T. Tarasenko[1].* 1) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Department of Pediatrics, The University of Texas Health Science Center, Houston, TX; 3) Department of Neurology, Columbia University Medical Center, New York, NY.

Immune cells undergo major changes in cellular energy metabolism during proliferation and differentiation. This area of immunometabolism has recently seen a resurgence, however, studying primary mitochondrial disease (pMD) as a translational model system for understanding immune cell bioenergetics is unique. Immune dysfunction in patients with pMD is under appreciated. In a cohort of patients with pMD (N=55) clinical and laboratory data suggested defects in cellular and humoral immunity marked by frequent respiratory tract viral infections (64%) including upper respiratory viral infections, pneumonia and sinusitis. Twelve patients (22%) had functional immunodeficiency requiring IVIG replacement therapy. Flow cytometry revealed elevated naïve T-cells concomitant with low memory T-cells (CD45RO+). Via enzymatic and bioenergetic evaluation, we demonstrated that mitochondrial dysfunction can be traced to peripheral blood mononuclear cells in pMD, suggesting that the immune defects seen may result from a cell-intrinsic energy deficiency. In order to study cell-intrinsic defects in bioenergetics, we created a mouse model with an isolated cytochrome c oxidase (COX) deficiency by disrupting the *Cox10* gene in T-cells using a cre-recombinase system. Cox10flox/flox;CD4Cre/+ mice show good viability and display isolated COX deficiency in T-cells, manifesting as decreased cell numbers and proliferation, enhanced cell death (24% increase), and abnormal differentiation with reduced memory T-cells (22% decrease). Immunization with the T-cell dependent protein antigen (NP-CGG) resulted in decreased antigen specific IgG titers (60% decrease). Infection with influenza lead to the reduced production of viral specific memory T-cells in Cox10flox/flox; CD4Cre/+ mice. In addition, Cox10flox/flox; CD4Cre/+ mice were unable to clear influenza infection as reflected by significantly increased lung titers (P = 0. 03). Mechanistically, these defects in cellular function could be traced to the early stages of activation and are reflective of T-cell bioenergetic deficiency. In summary, we have successfully created an animal model system of T-cell intrinsic mitochondrial dysfunction, which recapitulates components of the immune phenotype seen in a subset of patients with pMD. These studies will not only help define the role of mitochondrial bioenergetics in immune cell function, which is of great interest in immunology, but will also expand this under-recognized phenotype of pMD.

2314F

Novel TATA Box Promoter Mutation in SUCLG1: Expansion of the Phenotype. *J. L. Fraser[1, 2], J. L. Sloan[2], S. I. Berger[1], I. Manoli[2], G. M. Rice[3], C. P. Venditti[2].* 1) Medical Genetics Training Program, National Human Genome Research Institute, Bethesda, MD; 2) Organic Acids Research Section, National Human Genome Research Institute, Bethesda, MD; 3) Department of Pediatrics, University of Wisconsin School of Medicine and Public Health, Madison, WI.

We present a 12 year old female, born to second cousin consanguineous parents, with mild methylmalonic acidemia, failure to thrive, and multisystemic disease. She presented in early infancy with severe axial hypotonia and developmental delays. She had lactic acidosis and hypoglycemia on fasting as an infant; in studies from muscle biopsy and fibroblasts, respiratory chain enzyme activities were normal, with no ragged red fibers on muscle histology. The patient also presented in mid-childhood with Raynaud phenomenon, ulcerations, low immunoglobulins without frequent infections, and proximal tubule dysfunction including tubular proteinuria, phosphaturia, and glycosuria. At 12 years, she had severe choreoathetosis and ataxia, with cystic necrosis of the basal ganglia and lactate peaks on MRS in affected brain regions. While moderately delayed, she has required no respiratory support or g-tube feedings, has functional receptive language abilities, and is able to sit and ambulate independently, despite use of a wheelchair for long-distances due to fatigability and instability. She has significant attentional deficits and obsessive-compulsive tendencies. Her clinical and laboratory symptoms prompted an exhaustive but unrevealing evaluation, which included sequencing of SUCLG1 and SUCLA2. After thorough phenotyping was performed under a NIH natural history clinical protocol for methylmalonic acidemia (clinicaltrials. gov identifier NCT00078078), a genomic evaluation using a dense whole genome array (CGH+SNP) demonstrated 65 Mb (2. 2%) of homozygous sequence. Exome sequencing data, analyzed using VarSifter and filtered to prioritize candidate genes within regions of homozygosity, identified a probable pathogenic variant that alters the TATA box in the SUCLG1 promoter (chr2:84686610 T>C). Western analysis of patient fibroblast extracts demonstrated markedly decreased expression of all subunits of the succinate-coenzyme A ligase complex compared to a maternal heterozygote control . We suggest that a promoter mutation has resulted in reduced transcription of SUCLG1, diminished protein expression and a concomitant decrease in other succinate-coenzyme A ligase complex subunits. Trace SUCLG1 expression and activity likely accounts for the attenuated phenotype displayed by this patient as compared to others previously reported. Promoter-reporter studies and expression analyses will further define the pathogenicity of the TATA box variant identified in this study.

2315T

Inborn errors of metabolism spectrum of MS/MS target diseases in Vietnam. *V. Dung[1], K. Nguyen[1], M. Nguyen[1], T. Bui[1], N. Can[1], H. Nguyen[1], H. Le[1], T. Fukao[2], S. Yamaguchi[3].* 1) Metabolism and Genetics, National Hospital of Ped. Hanoi, Viet Nam; 2) Department of Pediatrics, Graduate School of Medicine, Gifu University. Japan; 3) Department of Pediatrics - Shimane University School of Medicine. Izumo, Japan.

Vietnam is the easternmost country on the Indochina Peninsula in Southeast Asia. With an estimated 90 million inhabitants as of 2013, it is the world's 13th-most-populous country, and the eighth-most-populous Asian country. Congenital anomalies accounted about 22% of causes of deaths in children under-5 (2010). The first service for IEMs was set up at the Northern referral center of Pediatrics – National Hospital of Pediatrics, Hanoi (NHP) in 2004 officially. The NHP in Ha Noi provides services to the population of north VietNam (~30 million people). **Objectives**: The aim of this report is to highlight disease spectrum of MS/MS target disease in Vietnam. **Methods:** 2405 high-risk cases with IEMs were studied at NHP during 10 years (2005-2014). Dry blood and urine samples were analyzed using MS/MS (amino acid & acylcarnitine analysis) & GC/MS (organic acid anaysis) for screening from 2005. Organic acids analysis for fresh urine samples was performed using GC/MS and amino acid analysis for plasma samples were performed using HPLC for confirmation of diagnosis. **Results:** Oganic acidemia (OAs), amino acid disorders (AAs), urea cycle disorders (UCDs) and fatty acid oxidation disorders (FAOD) were identified in 240/2405 cases (9. 9%). 121/240 patients (50. 4%) were OAs with 12 different disorders: BKT (34 cases), PPA (21 cases), 5-oxoprolinuria (19 cases), MMA (15 cases), Glutaricaciduria type II (GA II) (11 cases), 3-methylglutaconic aciduria (4 cases), Isovaleric academia (4 cases), Multiple carboxylase deficiency (MCD) (2 cases), 3-methylcrotonylCoA carboxylase deficiency (2 cases). 43/240 patients (17. 9%) were amino acid disorders including 36 cases with MSUD, 7 cases with PKU and 1 case with Tyrosinemia type 1. 38/240 patients (15. 8%) were UCDs including OTC deficiency (15 cases), Citrulinemia type 1 (10 case) and Argininosuccinic aciduria (1 case). 39/240 patients (16. 3%) were FAOD including SCAD (3 cases), MCAD (3 cases), VLCAD (8 cases), LCAD (2 cases), CPT 2 (8 cases), CPT 1 (1 case) and Primary carnitine deficiency (14 cases). Mortality rate was reduced from 50% in 2005 to 9% in 2014. **Conclusions:** Treatable conditions of IEMs were most common in Vietnamese patient identified using MS/MS. Expending newborn screening using MS/MS should be introduced to reduce mortality in Vietnamese children.

2316F

Five years of newborn screening of inherited metabolic disorders in the Czech Republic. *K. Peskova[1], J. Bartl[1], P. Chrastina[1], L. Dvorakova[1], H. Vlaskova[1], L. Fajkusova[2], D. Friedecky[3], E. Hlidkova[3], R. Pazdirkova[4], D. Prochazkova[5], P. Jesina[1], Z. Hruba[2], T. Adam[3], V. Kozich[1].* 1) Inst Inherit Metabol Disorders, General University Hospital, Prague 2, Czech Republic; 2) Faculty of Medicine and University Hospital, Center of Molecular Biology and Gene Therapy, Masaryk University, Brno, Czech Republic; 3) Laboratory for Inherited Metabolic Disorders, Department of Clinical Biochemistry, University Hospital Olomouc, Czech Republic; 4) Third Faculty of Medicine and University Hospital Kralovske Vinohrady, Department of Pediatrics, Charles University, Prague, Czech Republic; 5) Faculty of Medicine and University Hospital, Department of Pediatrics, Masaryk University, Brno, Czech Republic.

Background:Early diagnosis of inherited metabolic disorders (IMD) enables better prognosis, treatment and survival of patients. After five years, we present the results of newborn screening program in the Czech Republic including 10 IMDs. International project Region4Screening recommends quality criteria for newborn screening of 30 IMD: the total detection rate of patients better than 1:3,000, the positive predictive value higher than 20 % and false positive rate lower than 0. 3 % of testing newborns. Methods:Amino acids and acylcarnitines were extracted from dried blood spot samples and analyzed by tandem mass spectrometry using kit MassChrom Amino Acids and Acylcarnitines. Molecular genetic analysis of phenylketonuria/hyperphenyalaninemia (*PAH*), MCAD deficiency (*ACADM*), LCHAD deficiency (*HADHA*) and glutaric aciduria type I (*GCDH*) was performed using Sanger sequencing. Results:Between October 2009 and December 2014 we analyzed samples from 577,756 newborns. We detected 166 patients with subsequently confirmed IMD. We detected 110 patient with phenylketonuria/hyperphenylalaninemia (PKU/HPA), 29 patients with MCAD deficiency, 10 patients with LCHAD deficiency, 4 patients with VLCAD deficiency, 3 patients with glutaric aciduria type I, 3 patients with isovaleric aciduria, 3 patients with hydroxyprolinemia, 2 patients with leucinosis and 2 patients with CPT I deficiency. One patient with intermittent MSUD was not recognized. The positive predictive value was 26. 2 % and false positive rate 0. 08 %. The total detection rate was 1:3,500. For the most frequent diseases (PKU/HPA 1:5,200; MCAD deficiency 1:19,900 and LCHAD deficiency 1:64,200) frequencies of pathological alleles and its geographical distribution in regions were evaluated. In the patients with PKU/HPA, MCAD deficiency, LCHAD deficiency and glutaric aciduria type I we have identified 51, 8, 6 and 6 different mutations, respectively. Conclusion:The results are in agreement with Region4Screening target performance. Nevertheless, our laboratory algorithms are still being optimized in order to reduce number of false positive cases. Nowadays, we propose to extend the newborn screening panel from 10 to 15 IMDs by adding citrullinemia type I, argininemia, CBS/methylenetetrahydrofolate reductase deficiency and biotinidase deficiency. Other 10 IMDs could be detected secondarily. Supported by MZ CR – RVO VFN64165 and OPPK CZ. 2. 16/3. 1. 00/24012.

2317T

Neonatal Screening of Lysosomal Storage Diseases in a Cohort of Mexican Population. *J. Navarrete[1], D. Cervantes[1], A. Limon[2], J. Ramírez[3].* 1) Dept Gen, Hosp Sur PEMEX, Mexico City, Distrito Federal, Mexico; 2) Dept Ped, Hosp Sur PEMEX, Mexico City, Distrito Fedral, Mexico; 3) Dept. Ped, Villahermosa PEMEX, Villahermosa, Tabasco, Mexico.

The goal of newborn screening is an early detection of inborn errors of metabolism diseases. In Mexico we began newborn screening since 1977 with very few inborn errors of metabolism such as phenylketonuria, galactosemia, congenital hypothyroidism, sickle cell anemia and cystic fibrosis. (1)Petróleos Mexicanos is a big governmental institution with approximately 15,000 workers with their families. Since 2005 a larger screening has been done to all newborns in this institution through all the country. We tested a total of 69 inborn errors of metabolism including most aminoacidopathies, acidurias, hemoglobinopathies, G6PD deficiencyiy, adrenal hyperplasia, cystic fibrosis and biotinidase deficiency among the most important ones. Since August 2012 we included some lysosomal storage diseases such as Gaucher disease, Niemann-Pick type A and B diseases, Pompe disease, Krabbe disease, Fabry disease and MPS I(2). We analyzed our results from August 2012 to May 2015. We found 10 patients with a lysosomal storage disorder, 5 with Fabry disease, 4 with Pompe disease and 1 with MPS1. We confirmed diagnoses with enzyme activity and mutational analyses and we studied the whole family finding new affected members. These findings give us the opportunity to make genotype-phentype correlation for the early treatment of these diseases and close follow-up.

2318F

Newborn screening for Mucopolysaccharidoses: a pilot study in 1276 samples. *F. Kubaski[1,2], R. W. Mason[1,2], R. Giugliani[3,4], S. Yamaguchi[5], Y. Suzuki[6], J. Hanai[7], A. M. Montaño[8], S. Tomatsu[1,2].* 1) Department of Biological Sciences, University of Delaware; 2) Nemours-Alfred I duPont Children Hospital, Newark, DE; 3) Medical Genetics Service-Hospital de Clinicas de Porto Alegre-HCPA; 4) Department of Genetics/ Universidade Federal do Rio Grande do Sul; 5) Shimane University; 6) Gifu University; 7) Sapporo University; 8) Saint Louis University.

Mucopolysaccharidoses (MPSs) are genetic disorders caused by deficiency of lysosomal enzymes responsible for the catabolism of glycosaminoglycans (GAGs). There are 11 known enzyme deficiencies leading to seven main types of MPS. Together, they are estimated to occur in 1 of 22,000 live births. The phenotype is variable and progressive being the patients are asymptomatic at birth and diagnosed only when they already have signs and symptoms of the disease. Specific treatment is already available for several MPSs, and there are evidences that better outcome of therapy occurs when it is started early. Thus, early diagnosis is extremely relevant to allow better treatment as well as prognosis. The current study aims to develop a pilot study with 1276 samples for newborn screening of MPSs. We developed a protocol for GAG measurement using liquid chromatography mass spectrometry (LC-MS/MS). Dried blood spots from 1276 controls were used to determine cutoffs for two subclasses of sulfation of heparan sulfate (0S, NS). Disaccharides were produced by specific enzyme digestion with heparitinase and were quantified by negative electron ion mode of multiple reaction monitoring. Dried blood spots obtained at birth from patients who were subsequently diagnosed with MPS (MPS I= 6, MPS II= 2, MPS III= 5) were also examined. The average values for each subclass of HS in controls was HS-0S= 33. 3 ng/ml (+26. 3), HS-NS= 9. 2ng/ml (+16. 5), 14. 4 (+28. 7). Average for MPS patients for HS-0S=205. 09 ng/ml (range: 113-312 ng/ml), HS-NS: 58. 69 ng/ml (range: 31-80. 4 ng/ml). Cutoffs were defined as 3SD values (HS-0S>112. 2ng/ml, HS-NS>51. 4ng/ml). All MPS I, II and III newborn patients had higher HS-0S than control samples. We conclude that HS-0S and HS-NS are biomarkers potentially useful for the newborn screening of newborn of MPS I, II, and III.

2319T

The use of next generation sequencing in metabolic disease screening and diagnosis of newborns (California NBSeq). *R. L. Nussbaum[1,6], S. E. Brenner[1,2], F. Chen[1], R. J. Currier[3], B. Dispensa[1], B. A. Koenig[1], M. Kvale[1], P. Kwok[1], R. Lao[1], S. D. Mooney[4], J. M. Puck[1], N. Risch[1], J. Shieh[1], S. Rajgopal[5], B. Zerbe[1].* 1) UCSF, San Francisco, CA; 2) UC Berkeley, Berkeley, CA; 3) Newborn Screening Program California Department of Public Health Richmond, CA; 4) University of Washington, Seattle, WA; 5) TATA Consultancy Services Innovation Laboratories, Hyderabad, India; 6) Invitae Inc.

Purpose: NBSeq is the California component of the NSIGHT project, a NICHD/NHGRI funded program designed to explore the use of next generation sequencing in screening and diagnosis of newborns. California NBSeq is asking (1) can next-generation sequencing (NGS) improve the sensitivity or specificity of current newborn screening tests, (2) can NGS improve current newborn screening programs by shortening time-to-diagnosis following a positive screening test, (3) can NGS extend the scope of newborn screening programs to new disorders for which there are currently no screening modalities, and (4) how would NGS impact the legal, social and ethical frameworks that underpin newborn screening? **Methods:** We are using archived newborn blood spots screened for an amino acid, organic acid, or fatty acid oxidation disorder, and unaffected controls, and performing whole-exome sequencing in a blinded fashion to determine sensitivity and specificity. Our project will examine ~1600 newborn spots, consisting primarily of all confirmed true positives and false negatives, as well as a sample of false positive and true negative newborn blood spots, selected from among ~2. 4 million samples (2. 2 million regular nursery and ~216,000 NICU samples) screened in California between 2009-2013 inclusive using MS/MS. Sequence analysis will be limited to genes relevant to the metabolic disorders targeted by current screening programs. **Preliminary Results:** Two 3 mm punches from newborn blood spots stored for as long as 35 years yielded 620 ng (+ 238 ng) of DNA, with some outliers. Library fragment size, NGS quality and gene/exon coverage for the first 188 NBS samples were adequate for interrogating genes relevant to the metabolic disorders targeted by current screening programs. Our results demonstrate that dried blood spots provide a reliable DNA source for whole exome sequencing.

2320F

Improved quality of life with dietary interventions in a biochemically stable child with maple syrup urine disease. *M. Rasberry[1], B. C. Lanpher[2], K. Cusmano-Ozog[1], D. S. Regier[1].* 1) Children's National Medical Center, Washington, DC; 2) Mayo Clinic, Rochester, MN.

Maple Syrup Urine Disease (MSUD) is caused by a deficiency in the first stepof the metabolism of the branch chain amino acids (BCAA), leucine (leu), isoleucine, and valine. This leads to a build-up of these three amino acids and accumulation of the side-reaction metabolite alloisoleucine. The build-up of leu has been associated with lethargy, encephalopathy, coma, and death in the most severe cases. Case reports have identified patients with mild and intermittent forms of MSUD. Here we report a five year old girl with a new diagnosis of MSUD after a prolonged hospitalization for dehydration secondary to a viral illness. She presented to a community hospital when she was found to be overly sleepy following a night of gastrointestinal illness. She was admitted to a pediatric unit when she appeared too sleepy to maintain her hydration state. Over the following 3 days she developed ataxia, which improved with intravenous fluids. The pediatric hospitalist team contacted our referral center on-call physician due to her slow recovery and prolonged ketonuria. She was found to have massively elevated leu, isoleucine, and alloisoleucine levels. Prior to diagnosis, she was adverse to high protein foods such as meat and had a strong preference for fruits and vegetables. At the time of diagnosis, BCAA-free formula was started to provide 0. 8 g/kg/d of medical protein and started a protein restriction with a goal protein intake of 1. 0 – 1. 2 g/kg/d of natural protein. After one month on diet, she had improved emotional lability, increased concentration, improved school performance, and commented "my brain feels better with my new milk". After 6 months on diet, the family noted significant improvements in socialization, behaviors, school outcomes, and overall mood. During this same time period, her BCAA profiles stayed in the low-normal range, based on monthly evaluations. This case suggests that patients with prolonged ketonuria or dehydration-type symptoms should raise a red flag for inborn errors of metabolism and have screening with plasma amino acids and urine organic acids. Her newborn screening results were normal at that time; however, she would have been flagged as having leucine elevation using current state standards. Furthermore, minimal dietary intervention leads to significant improvement in patient quality of life, even when laboratory values remain unchanged with the intervention as in this case of mild/intermittent MSUD.

2321T

Neurotransmitters and Blood Phenylalanine in Phenylketonuria. *S. Yano[1], K. Moseley[1], X. Fu[2], C. Azen[3].* 1) Genetics Pediatrics, USC, Los Angeles, CA; 2) Pathology, Children's Hospital Los Angeles, USC Los Angeles, CA; 3) Clinical and Translational Science Institute, USC Los Angeles, CA.

Background Phenylketonuria (PKU) is a common genetic metabolic disorder causing intellectual disability due to phenylalanine (Phe) hydroxylase (PAH) deficiency. Newborn screening has been introduced for early diagnosis and dietary intervention based on blood Phe levels. Sapropterin, a synthetic form of tetrahydrobiopterin, is also used as a cofactor for PAH. PKU individuals with excellent blood Phe control recently have been reported to have a higher prevalence of neuropsychological deficits including decreased executive functioning and internalizing disorders. Abnormal neurotransmitter metabolism, particularly decreased serotonin and dopamine synthesis in the brain, is believed to be involved. Melatonin, a serotonin metabolite synthesized in the brain, and dopamine which originates from systemic neuronal cells including the brain, reported as useful peripheral biomarkers for brain serotonin and dopamine metabolism, respectively. *Methods* To evaluate brain serotonin and dopamine metabolism, nine adults with PKU completed our study consisting of four 4-week phases: (1) Large neutral amino acid (LNAA) supplementation, (2) Washout, (3) Sapropterin therapy, and (4) LNAA with sapropterin therapy. An overnight protocol measured plasma amino acids, serum melatonin, and urine 6-sulfatoxymelatonin and dopamine in first void urine after each phase. *Results* The relationship between serum melatonin and Phe showed a significant negative slope (p=0. 0005) with a trend toward differing slopes among individual subjects (p=0. 066). However, serum melatonin levels varied widely at a given Phe level. There was a negative association overall between Phe and urine dopamine (P=0. 047). *Conclusion* Blood Phe levels alone do not determine the CNS neurotransmitter metabolism among individuals with PKU. Monitoring with these peripheral biomarkers may individualize treatment to optimize brain neurotransmitter metabolism.

2322F

Combination therapy in a patient with chronic neuronopathic Gaucher disease: potential positive effect of early treatment. *F. Ceravolo, M. Grisolia, M. T. Moricca, F. Falvo, S. Sestito, A. Nicoletti, D. Concolino.* Department of Pediatric, University of Catanzaro, Catanzaro, Italy.

The chronic neuronopathic form of Gaucher Disease (NGD) is characterized by a later onset of neurological symptoms and a more protracted neurological and visceral involvement. The first clinical signs of neurological involvement appear at the median age of 2 years and the major manifestation is early development of horizontal supranuclear gaze palsy followed by cognitive impairment, myoclonic epilepsy, ataxia and spasticity which develop as the illness progresses. Treatment consists in enzyme replacement therapy (ERT) with recombinant enzyme: glucocerobrosidase (GBA). ERT has shown to improves haematological and bone disease but the enzyme is not able to cross the blood-brain barrier (BBB), preventing the neurological manifestation. The substrate reduction therapy (SRT) is based on the iminosugar, miglustat, an inhibitor of the glucosylceramide synthase, the enzyme that catalyzes the first step in glycosphingolipids synthesis. According to its chemical/physical properties, miglustat can cross the BBB. Some reports have shown the efficacy of miglustat, used in combination with ERT, in treating or preventing the neurological symptoms in patients affected by chronic NGD. We report of a 7 years old boy with chronic NGD who received an early treatment with combination therapy. Diagnosis was suspected at the age of 16 months for anemia, thrombocytopenia and splenomegaly. Biochemical exams revealed an elevated level of Chitotriosidase activity: 508 nmol/mg prot (n. v. 5. 9 - 41 nmol/mg prot), and a reduced Beta Glucosidase activity : 2nmol/mg/prot (n. v. 4. 5 – 18 nmol/mg/prot). Molecular analysis showed an homozygosity for L444P mutation in Beta-Glucosidase gene confirming the diagnosis of chronic NGD . The patient began ERT at the dosage of 60 U/Kg/every 2 weeks and at the age of two he also started the SRT with miglustat. At present the patient doesn't show any sign of neurological impairment: no saccadic movement velocity reduction; normal visual evoked potential (VEP) and auditory brain responses (ABR). Finally he has no cognitive impairment and regularly attends at school. This finding is in contrast with the natural evolution of homozygous L444P in NGD patients worldwide and, more interestingly, with our personal records (data not published) consisting in 5 older patients with the same genotype, all comings from the same restricted geographical area , who were only treated with ERT and presented neurological impairment in the first 5 years of life.

2323T

Strategy to assess the pharmacokinetics of a proprietary human acid α-glucosidase with high mannose 6-phosphate during its development as a potential next-generation treatment for Pompe disease. *E. R. Benjamin, R. Hamler, R. Khanna, D. Hilliard, S. Xu, K. Valenzano, R. Gotschall, H. Do, F. K. Johnson.* Translational Science, Amicus Therapeutics, Cranbury, NJ.

Pompe disease is a lysosomal storage disorder that results from deficiency in acid α-glucosidase (GAA) activity, and is characterized by lysosomal glycogen accumulation. The current treatment is enzyme replacement therapy (ERT) using recombinant human GAA (rhGAA). The targeting and uptake of rhGAA into lysosomes requires mannose-6-phosphate (M6P), a specialized carbohydrate which binds to cation-independent M6P receptors (CIMPR) at the cell surface. The quality and quantity of M6P on existing ERT is not optimal, which limits lysosomal targeting. While existing ERTs provide some clinical benefit, unmet needs still exist due to inadequate muscle uptake. We have developed a novel rhGAA enzyme with significantly higher M6P content compared to existing products. This new rhGAA (designated as ATB200) binds the CIMPR with high affinity *in vitro*, and is more efficiently internalized by skeletal muscle myoblasts. Co-administration of a pharmacological chaperone to stabilize the ERT and prevent inactivation provides additive benefits. To support the development of ATB200, a comprehensive approach to assess the pharmacokinetics was developed. The measurement of circulating active enzyme by rhGAA activity assay is complemented with an absolute quantitative method to measure total rhGAA protein. The latter utilizes LC-MS/MS quantification of ATB200-specific "signature" peptides after trypsin digestion of proteins present in plasma samples. The method has been analytically validated in rat and monkey plasma. Pre-specified assay acceptance criteria for precision [% Coefficient of Variation ≤20% (25% for LLOQ)], accuracy [% Difference from Nominal ≤20% (25% for LLOQ)] [LLOQ=0. 500 µg/mL], and other parameters (e. g. , selectivity in matrix±high concentration anti-rhGAA antibodies, dilution linearity, stability) were met. In a single-dose research study in rats, total ATB200 protein levels were measured in plasma up to 24 hours after 10-minute intravenous (IV) infusion. Total ATB200 protein concentrations closely matched those of active ATB200 at time points up to 4 hours. At later time points, the concentrations of active ATB200 were progressively lower compared to total ATB200 protein. The systemic exposure (AUC0-24h and Cmax) increased in a greater than dose-proportional manner. These data demonstrate that characterization of the pharmacokinetics of both active and total ATB200 is relevant for this high M6P rhGAA, a potential next-generation treatment for Pompe disease.

2324F

Diagnosis of a mild case of mucopolysaccharidosis type II (Hunter disease) in a 3-year old girl: to treat or not to treat? *A. M. Laberge, P. Allard, G. A. Mitchell.* Service de genetique medicale, CHU Sainte-Justine, Montreal, Quebec, Canada.

A 3-year old girl was admitted for respiratory distress following tonsillectomy. Language delay, frontal bossing, and midface hypoplasia were noted, prompting genetics referral. Medical history included recurrent pneumonia, and an episode of febrile seizure. Growth and head circumference were at the 50th centile. Family history was unremarkable. She had mild hepatosplenomegaly, bilateral mild to moderate mixed hearing loss and chest X-ray showed broadened ribs and ovoid vertebral bodies, suggesting a lysosomal storage disorder. Cardiac and ophthalmological evaluations were normal. Urine MPS were elevated: 52. 9 mg/mmol creatinine (control 6. 9 +/- 3. 7). Qualitative analysis showed a majority of dermatan and heparan sulfate. Alpha-iduronidase activity was normal. Iduronate sulfatase activity was very low: 0. 05 nmol/h/mg prot (median 10. 7 (8. 8-17. 7)). Sequencing of IDS identified a heterozygous pathogenic mutation: c. 1122C>T. Karyotype was normal. Methylation analysis for X-inactivation showed complete inactivation bias in leukocytes. Brain MRI was normal except for wide Virchow-Robin spaces. Skeletal survey was consistent with dysostosis multiplex, but joint mobility was normal. Neurological assessment revealed only deafness and associated language delay. There are only a handful of reported cases of MPS II in girls, with variable severity. Studies in males have shown that enzyme replacement therapy (ERT) may improve/stabilize visceromegaly, lung function, joint mobility, and cardiac valve regurgitation. ERT requires weekly infusions and is expensive, which can represent a burden for the family. Evidence about ERT in females is limited to 3 cases, all more severely affected than our case at the same age (Manara et al, 2010; Jurecka et al, 2012). All had neurological involvement and started ERT after age 5 (5. 8 yrs, 9 yrs, and 7 yrs, respectively), which led to disease stabilization, except for progression of neurological involvement in 2/3 cases. The pros and cons of ERT were discussed with the mother and it was mutually agreed to postpone ERT for the time being because the patient currently has no functional limitations that would be expected to respond to ERT. We will continue to closely monitor diseaseprogression and reconsider ERT accordingly. This case highlights how trying it can be to decide when to start treatment (i. e. when the treatment burden is worth the benefit)in mild cases of a disease, when evidence is limited.

2325T

Human Recombinant Arginase Enzyme:A Potential New Therapy for Arginase Deficiency. *L. Burrage[1,2], Q. Sun[1], S. Elsea[1], M. Jiang[1], S. Nagamani[1,2], A. Frankel[3], E. Stone[4], G. Georgiou[4], B. Lee[1,2], Members of the Urea Cycle Disorders Consortium.* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Texas Children's Hospital, Houston, TX; 3) Simmons Comprehensive Cancer Center, UT Southwestern Medical Center, Dallas, TX; 4) Department of Chemical Engineering, The University of Texas at Austin, Austin, TX.

Arginase deficiency is characterized by elevated plasma arginine levels as the result of a deficiency of arginase 1, a cytosolic urea cycle enzyme. Clinically, patients with arginase deficiency present with spastic diplegia, developmental delay or intellectual disability, seizures, and growth deficiency. In contrast to other urea cycle disorders, recurrent, severe hyperammonemia is a less frequent complication. Normalization or near-normalization of plasma arginine levels is the consensus goal of treatment in arginase deficiency because arginine and its metabolites are believed to contribute to the neurologic features of this disorder. However, no studies have evaluated the efficacy of current treatment strategies for the disorder. To this end, we analyzed longitudinal data from patients (n=22) with arginase deficiency who are enrolled in a natural history study conducted by the Urea Cycle Disorders Consortium. Our analyses revealed that 97% of plasma arginine levels in subjects with arginase deficiency were above the normal range despite standard treatment, which typically consisted of protein restriction and nitrogen-scavenging agents. Moreover, only 55% of patients with arginase deficiency reported one or more hyperammonemia episodes with most patients (mean age: 14. 7 years, range: 0-50 years) reporting fewer than five total lifetime reports of hyperammonemia. These data indicate that normalization of plasma arginine is challenging in this disorder. Recently, recombinant arginine-degrading enzymes have been introduced as a therapeutic strategy for various forms of cancer. We tested whether one of these enzymes, a pegylated human recombinant arginase 1 (AEB1102), reduces plasma arginine in two murine models of arginase deficiency. In both the neonatal and adult arginase deficient mouse models, AEB1102 administration resulted in reduced plasma arginine levels after repeated dosing. Despite improved plasma arginine levels, enzyme administration did not result in improved survival likely because this pegylated enzyme does not enter the liver and thus, does not ameliorate urea cycle dysfunction and resulting hyperammonemia that accounts for lethality. Given that elevated plasma arginine rather than hyperammonemia is the major treatment challenge in patients with arginase deficiency, we propose that AEB1102 may have therapeutic potential as an arginine-reducing agent in patients with arginase deficiency.

2326F

Familial Hypercholesterolemia (FH): treatment of an Italian family by LDL-apheresis (LDL-a). *A. Saluto[1], F. Napoli[1], T. Beltramo[1], O. Camerini[1], K. Bonomo[2], P. Ottone[1].* 1) Laboratory of Immunohematology, S. Luigi Gonzaga University Hospital, Orbassano (TO), Italy; 2) Division of Metabolic Disorders, S. Luigi Gonzaga University Hospital, Orbassano (TO), Italy.

FH is a common genetic disorder inherited in an autosomal dominant way with a prevalence of approximately 1 in 300-500 for heterozygotes in North America and Europe. It is characterized by elevated Low-Density Lipoprotein Cholesterol (LDL-C) which causes the accumulation of cholesterol in the arterial wall resulting in accelerated atherosclerosis and premature cardiovascular disease; typical external signs are xanthomas and arcus lipoides. FH is genetically heterogeneous as it can be caused by defects in at least three different genes that encode proteins involved in the hepatic clearance of LDL-C. These defects may be due to mutations in the gene coding LDL-R (classic FH or ADH-1), in the gene coding for the APOB (ADH-2) or in the gene coding for the enzyme PCSK9 (ADH-3). Mutations in LDL-R gene are the most frequent cause of FH (from 90% to 95%); more than 1. 700 different mutations have been characterized (Vogt et al, 2015). We have a family, referred to our centre, with the mutation, c. 68-?_1845+?del,p. (V23Gfs*29), a 24 kb deletion eliminating exons 2-12, named "FH-Pavia" (Bertolini et al, 1992): the father (subject I-1), heterozygous is 51 years old, the son (subject II-1) and the daughter (subject II-2) homozygous are respectively 18 and 16 years old. The three patients were treated with Rosuvastatin at maximum levels and underwent LDL-a every two weeks. As for the subject I-1 we used the HELP method (heparin-induced extracorporeal LDL precipitation), whereas for the subjects II-1 and II-2 the DSA method (dextran sulphate adsorption) was applied. Total cholesterol, LDL-C, HDL-C and triglycerides were calculated before and immediately after apheresis. We observed an average reduction of total cholesterol (57% in the heterozygous subject, 79% in the homozygous subjects), of LDL-C (61. 6% in the heterozygous subject, 83% in the homozygous subjects) of HDL-C and triglycerides (similar values for the three patients). Our work was performed to get more information on lipid apheresis and to evaluate the effectiveness in lowering lipid values in heterozygous and homozygous subjects; it shows that LDL-a notably lowers total cholesterol, LDL-C and, to a much lesser degree, HDL-C and triglycerides. The follow-up of the patients reveals the complete regression of xanthomas and suggests a reduced risk of cardiac events after apheresis.

2327T

Utility of citrulline supplementation in treating hypoargininemia in children with MELAS syndrome. *A. W. El-Hattab[1, 2], L. T. Emrick[1, 3], S. Chanprasert[1, 3], J. W. Hsu[4], M. Almannai[1, 3], W. J. Craigen[1, 3], F. Jahoor[4], F. Scaglia[1, 3].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas, USA; 2) Division of Clinical Genetics and Metabolic Disorders, Tawam Hospital, Al-Ain, United Arab Emirates; 3) Texas Children's Hospital, Houston, Texas, USA; 4) Children's Nutrition Research Center, Baylor College of Medicine, Houston, Texas, USA.

Mitochondrial encephalomyopathy, lactic acidosis, and stroke-like episodes (MELAS) is a frequent mitochondrial disorder. There is growing evidence that nitric oxide (NO) deficiency occurs in MELAS and results in impaired blood perfusion that can contribute to several complications including stroke-like episodes, myopathy, and lactic acidosis. The etiology of NO deficiency in MELAS is multifactorial with the decreased availability of NO precursors, arginine and citrulline, playing a major role. Previous studies demonstrated hypoargininemia and hypocitrullinemia in adults with MELAS. In this study we initially measured plasma amino acids in 5 children with MELAS at baseline and then after 48 hours of oral arginine supplementation. After a 2-week wash-out period, plasma amino acids were measured again at baseline and after citrulline supplementation. We also measured plasma amino acids in 5 healthy control children once at baseline. The age of children with MELAS ranged from 4 to 16 years and of the control group from 7 to 16 years. The baseline mean plasma arginine in children with MELAS was significantly lower than the control group (61. 3±5. 3 vs. 151. 3±18. 4 μM, p<0. 05). However, the baseline mean plasma citrulline was not significantly different between both groups. In children with MELAS arginine supplementation resulted in raising plasma arginine (58. 5±5. 2χ184. 3±14. 2 μM, p<0. 05) but not plasma citrulline. Interestingly, citrulline supplementation resulted in a more significant raise in plasma arginine (64. 1±5. 7χ257. 1±21. 3 μM, p<0. 05) and plasma citrulline (21. 3±7. 0χ155. 6±37. 7, p<0. 05). We have also used isotope infusion technique to assess NO production in the children with MELAS before and after arginine or citrulline supplementation. The measurement of isotopic enrichment is currently ongoing. Hypoargininemia and hypocitrullinemia have been previously described in adults with MELAS syndrome and can contribute to the NO deficiency in this syndrome. Arginine and citrulline supplementations have been suggested to be of therapeutic utility in MELAS through their ability of restoring NO production. In this study we provide evidence that children with MELAS also exhibit hypoargininemia. Citrulline supplementation was superior to arginine supplementation in raising both plasma arginine and citrulline in children with MELAS. This study adds more evidence to the utility of citrulline supplementation as one of the therapeutic options in MELAS syndrome.

2328F

Aggresomes formation and negative gain of function as alternative molecular mechanism in patients affected by phenylketonuria: implications for the therapy. *G. Bonapace[1], M. T. Moricca[1], F. Ceravolo[1], I. Mascaro[1], P. Strisciuglio[2], D. Concolino[1].* 1) Pediatrics, Faculty of Medicine University Magna Graecia, Viale Europa Catanzaro, Italy; 2) Dept of Pediatrics University Federico II Naples.

Phenylketonuria (PKU MIM 261600) is a human metabolic disease caused by mutations in phenylalanine hydroxylase gene (PAH) and is inherited in an autosomal recessive Mendelian fashion. Phenylalanine hydroxylase (PAH also known as phenylalanine 4-monooxygenase EC 1,14,16,11) catalyzes the rate-limiting step in L-Phenylalanine (L-Phe) catabolism in liver, using tetrahydrobiopterin (BH_4) and dioxygen as additional cosubstrates. PKU mutations are associated with impairment of PAH activity, leading to accumulation of L-Phe in plasma, and neurological damage in untreated patients Even though an early dietary treatment, consisting of a low protein diet with L-Phe restriction, can prevent the development of metabolic and pathological sequelae, the social burden of this treatment and the reduced patient'compliance, have led to search for new therapeutic approaches. To set a novel terapeutic approach for treating this disease we studied the aggregosome formation and the negative gain of functionusing is the c. 143C>T (L48S) PAH mutation as model. L48S is a well documented PAH misfolding mutation characterized by a progressive clinical worsening, and described as inconsistent in BH_4 responsive patients. The experimental hypothesis we testedis based on the formation of intracellulat toxic soluble aggregates induced by the mutant protein. Based on this mechanisms and on the observation that most of PKU patients are compound heterozigous for two different mutations, we hypothesize that in the patients carrying the L48S misfolding mutation in association with a second mutatedallele that retains some residual enzymic activity the formation of solubile aggregates can take pace and that these aggregates couldtrigger a generalized cell disfunction explaining the observed worsening of the clinical phenotype. By using electroporation we transfected HEK 293 cells with an expression vector carrying PAH WT, or PAHL48Sand demonstratedthe aggregosome formation and the loss of biological activity in the L48S transfected cells. By using 1-Deoxynojirimycin hydrocloride also described as a HSP70 inducers, we were able todecrease the amount of aggregated L48S protein in a dose response in comparison to untreated cells, to rescue the biological activity without any effect onthe HEK cells based on morphology and biochemical evaluation. Wepropose aggresomes formation as novel molecular mechanism in PKU with very interesting implications for the therapy

.

2329T

A Pilot Open Label Trial Assessing the Safety and Efficacy of Betaine in Patients with a Peroxisome Biogenesis Disorder (PBD) and PEX1-Gly843Asp (G843D) Genotype. *N. E. Braverman[1, 2], F. Plourde[2], R. Cooper[3], R. O. Jones[4], A. B. Moser[4], W. Rizzo[3].* 1) Human Genetics, McGill University, Montreal, QC, Canada; 2) Department of Pediatrics, McGill University Health Center, Montreal, QC, Canada; 3) Department of Pediatrics, University of Nebraska Medical Center, Omaha, NE, USA; 4) Department of Neurogenetics, Kennedy Krieger Institute, Johns Hopkins Hospital, Baltimore, MD, USA.

Introduction: The PBD are a heterogeneous group of disorders with a wide clinical spectrum and no therapy. A common mutation encodes a misfolded, degraded protein with residual function, PEX1-G843D. Betaine was identified as a non-specific chemical chaperone drug that rescued peroxisomal matrix protein import and plasmalogen levels in PEX1-G843D patient fibroblast lines. Thus we sought to determine if betaine was safe and effective as a treatment for PEX1-G843D patients. *Study description:* Two centers participated in this study. Ten patients with a PEX1-G843D genotype received betaine orally for 6 months at a dose of 3, 6 or 12 gms daily, depending on age or body weight. Age range was 11 months - 31 years; two patients were PEX1-G843D homozygotes and 8 were PEX1-G843D/null genotypes. Patients were examined at baseline and 6 months. Betaine effect was assessed by measurement of peroxisome metabolites that reflect clinical disease: very long chain fatty acids, pipecolic acid, plasmalogens and C27 bile acids. Peroxisome function tests were done twice before the trial (to provide baseline values), and at 6 months. *Results:* Betaine was tolerated well in all patients. There were large intra-and inter-individual variations in peroxisome functions at each time point. Statistically significant trends (paired sample t-tests) were observed for plasma pristanic/phytanic acid ratio decreased by 32%, pipecolic acid decreased by 20%, RBC membrane C26:1 decreased by 23% and C26/C22 ratio decreased by 10%. In the two PEX1-G843D homozygotes, RBC membrane plasmalogens increased by 50%. *Conclusions:* Although promising trends were observed, betaine did not overall improve peroxisome functions in a 6 month trial at the dosage used. Blood metabolite markers may be less sensitive than direct measurements of cell responses and plasma betaine levels achieved were below that used in the cell lines. To improve future clinical trial designs, baseline variation in peroxisome functions in PBD patients must be established. Other biomarkers with less intrinsic variation and clinical endpoints should be investigated. Larger patient numbers and crossover or placebo-controlled designs are indicated to overcome the effect of variation.

2330F

Long term follow up of 17 patients with methylmalonic acidemiafollowing solid organ transplantation. *J. L. Sloan, I. Manoli, C. P. Venditti.* NIH/NHGRI/GMBB, Bethesda, MD.

Methylmalonic acidemia (MMA) is a devastating disorder, often recalcitrant to medical management. Solid organ transplantation, including liver (LT), combined liver-kidney (LKT) or kidney transplantation (KT), has been used as a therapy since the 1990's yet, there is limited information on long- term outcomes of transplant recipients and no guidelines regarding indications for transplant. We describe the natural history up to 16. 9y post-transplant in 17 individuals with MMA (15 mut, 1 cblA, 1 cblB) followed at the NIH Clinical Center as part of a dedicated natural history study (clinicaltrails. gov identifier NCT00078078). There were 3 LT, 9 LKT and 5 KT recipients in our cohort with mean follow up since transplant of 5. 5 yrs (range 0. 8-16. 9 yrs). LT recipients were transplanted younger (mean 2. 1y) than LKT (mean 15. 2y). 10/22 of the MUT alleles in LT and LKT patients were null. Although all 12 patients status post LT or LKT remained metabolically stable (77. 3 yrs combined follow-up), two LT patients displayed renal disease requiring subsequent KT. Two LKT recipients developed severe neurological complications (2 months, 5 yrs later) despite adequate metabolic control. In addition, one mut0 patient with isolated KT suffered repeated metabolic decompensations and developed optic nerve atrophy 3 yrs post-transplant. Improvement in metabolic parameters was observed following transplantation (72-93% decrease), primarily due to correction of renal function, but all had persistently elevated methylmalonic acid in plasma (71-1383 μM, nl <0. 4) and urine (166-4446 mmol/mol cr, nl<3). Despite the massive MMA elevations, two LKT patients >15 yrs post transplant had eGFR >40 ml/min/1. 73m2. Studies in this patient group, the largest assembled to date, afford the following conclusions: 1. LT and LKT, but not isolated KT, completely prevent ketoacidotic crises. 2. After LT, patients can develop renal failure and therefore require careful monitoring and this is also a theoretical risk for LKT and KT patients. 3. Although solid organ transplantation improves biochemical parameters, adequate metabolic control is required following the procedure due to the risk for MMA related complications e. g. optic nerve atrophy, neurological complications. Further studies are needed on the association of genotype and metabolic correlations pre- and post-procedure and the outcomes achieved, to develop guidelines about the optimal timing and procedure indicated for each patient.

2331T

Auditory pathway findings in Smith-Lemli-Opitz syndrome (SLOS). *T. Pointon[1], D. Hayes[2], E. R. Elias[3].* 1) University of Colorado School of Medicine, Aurora, CO; 2) Department of Physical Medicine and Rehabilitation, University of Colorado School of Medicine, Aurora, CO; 3) Department of Pediatrics and Genetics, University of Colorado School of Medicine, Aurora, CO.

Introduction: Smith-Lemli-Opitz syndrome (SLOS) is an autosomal recessive condition caused by mutations in the*DHCR7*gene which codes for 7-dehydrocholesterol reductase, the final enzyme in the cholesterol biosynthetic pathway. This results in cholesterol deficiency and accumulation of sterol precursors 7- and 8-dehydrocholesterol (7-DHC and 8-DHC). Patients with SLOS present with a complex phenotype including cognitive disabilities, autism, hearing loss, and progressive retinal dystrophy. Both sensorineural and conductive hearing loss are observed in SLOS. While retinal dysfunction has been reported, variations in the auditory pathway have not previously been described in SLOS. The goal of this study was to analyze neural conduction through the auditory pathway to better understand the pathophysiology of hearing loss in SLOS. **Methods**: Auditory brainstem response (ABR) was measured serially in 16 SLOS patients being treated with cholesterol supplementation and the antioxidant preparation AquADEKS® under an IRB approved protocol. Data was collected at baseline, then annually after beginning treatment. ABR data was assessed to determine hearing status and time required for neural conduction through the auditory brainstem pathway. **Results**: Prolonged neural conduction through the auditory brainstem pathway was observed in 6 out of 16 patients with SLOS. Of these patients, 3 had sensorineural and/or conductive hearing loss. Normal auditory brainstem neural conduction was observed in 9 out of 16 SLOS patients. Of these, 6 had sensorineural and/or conductive hearing loss. Complete sensorineural deafness was found in one patient, thus no ABR data could be collected. **Conclusion**: These data suggest that the metabolic alterations in SLOS negatively impact neural conduction through the auditory system in a subset of patients. Variation in auditory neural conduction seen among patients with SLOS is likely multifactorial. It is hypothesized that cholesterol deficiency contributes to neuronal membrane instability, altered functioning of ion channels and decreased myelination of axons. In addition, oxidation of 7-DHC and 8-DHC into toxic oxysterol byproducts, which cause neuronal death and retinal dysfunction, may also contribute to altered function in the auditory system. Current studies are underway to further investigate the effect that treatment has on these auditory findings and to determine if age at treatment initiation plays a role in outcomes.

2332F

A Replication study of 49 Type 2 Diabetes Risk Variants in a Punjabi Pakistani Population. *A. Zia[1], X. Wang[2], A. Bhatti[1], F. Y. Demirci[2], W. Zhao[3], A. Rasheed[4], M. Samuel[4], A. K. Kiani[1], M. Ismail[5], J. Zafar[6], P. John[1], D. Saleheen[4,7], M. I. Kamboh[2].* 1) Atta ur Rahman School of Applied Biosciences (ASAB, National University of Sciences and Technology (NUST), Islamabad, NA, Pakistan; 2) Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, PA, USA; 3) Institute of Translational medicine and human genetics, Department of Medicine, University of Pennsylvania, PA, USA; 4) Center for Non-communicable diseases, Karachi, Pakistan; 5) Institute of Biomedical and Genetic Engineering (IBGE), Islamabad, Pakistan; 6) Pakistan Institute of Medical Sciences (PIMS), Islamabad, Pakistan; 7) Department of Biostatistics and Epidemiology, University of Pennsylvania, PA, USA.

Background: The burden of type-2 diabetes is alarmingly high in South Asia, a region that has many genetically diverse ethnic populations. Genome-wide association studies (GWAS) conducted largely in European populations have identified a number of loci predisposing to T2D risk, however the relevance of such genetic loci in many South Asian sub-ethnicities remain elusive. The aim of this study was to replicate 49 SNPs previously identified through GWAS in Punjabis living in Pakistan. **Methods:** We examined the association of 49 SNPs in 853 T2D cases and 1,945 controls using additive logistic regression models after adjusting for age and gender. **Results:** Of the 49 SNPs investigated, 16 SNPs were associated with T2D risk in Punjabis in a direction consistent with prior published reports. The most significant association was found for rs7903146 at the *TCF7L2* locus. For a per unit increase in the risk score comprising of all the 49 SNPs, odds ratio in association with T2D risk was 1. 16 (95% CI 1. 13-1. 19, P<2. 0E-16). **Conclusion:** These results suggest that a large number of T2D susceptibility loci are shared between Europeans and Punjabis living in Pakistan.

2333T

Gene Expression Profiles of LPL Deficiency Reveal LPL-Independent Pathways of Triglyceride Metabolism. *D. Gaudet, D. Brisson, K. Tremblay.* Université de Montréal Community Genomic Medicine Centre and ECOGENE-21, Chicoutimi, QC, Canada.

Background: Lipoprotein lipase deficiency (LPLD), a rare autosomal recessive disease, is a major cause of extreme hypertriglyceridemia (hyperTG). However, patients with LPLD do not exponentially accumulate TGs, and thus, must have access to a LPL-independent rescue pathway to survive. Recent data support APOC3 as a key regulator of LPL-independent pathways of TG metabolism. These pathways remain however to be identified. **Aim:** To investigate gene expression profile of patients presenting a wide spectrum of LPL activity. **Methods:** A total of 51 consenting subjects participated in this study. Prior to the analyses, patients were divided into three groups covering a wide spectrum of LPL activity: 15 homozygotes for null LPL gene mutations and less than 3% of LPL activity (HoLPL); 22 heterozygotes for null LPL mutations and 50% of normal LPL activity (HeLPL); and 14 normolipemic controls (wild-type LPL). The gene expression profiles of the HoLPL and HeLPL was compared to the control subjects profile. RNA samples were extracted from whole blood PAXgene and hybridized on Affymetrix Human Gene 2. 0 ST Array according to manufacturer procedures. Robust Mutli-Array normalization has been applied on the probe raw intensities. Differential expression moderated T-tests between studied groups were performed using the Bioconductor package Limma. False discovery rate (FDR) estimation was carried out using the Benjamini-Hochberg method. Biological pathways of the differentially expressed probes have been analyzed using the Qiagen Ingenuity Pathway Analysis software. **Results:** The studied groups allowed to conduct three differential analyses on a total of 48,226 detected probes. At a p-value < 0. 01, a FDR of 5% and a >2-fold down-expression (down) or over-expression (over) significance levels, 143 probes have been found differentially expressed in HoLPL and 65 in HeLPL compared to healthy controls. Of the identified probes, 59 are shared by both HoLPL and HeLPL; 84 are specific to HoLPL [70 down (38 annotated;32 non-annotated) and 16 over (8 annotated;6 non-annotated)]; and 6 are specific to HeLPL [5 down (1 annotated; 4 non-annotated) and 1 over annotated probe]. Most of the annotated genes are involved in circadian, inflammatory, immune or signaling pathways, docking systems or receptor-mediated clearance mechanisms. **Conclusion:** These results reveal gene expression signatures of a wide spectrum of LPL activity and suggest LPL-independent mechanisms of TG metabolism.

2334F

Novel *NPC* variants causing severe Niemann Pick C deficiency. *K. Chao[1], C. Adams[1], L. Fleming[3], L. A. Wolfe[1,2], W. A. Gahl[1,2], C. Toro[1], D. R. Adams[1,2].* 1) NIH Undiagnosed Diseases Program, National Institutes of Health, Bethesda, MD; 2) Office of the Clinical Director, National Human Genome Research Institute, Bethesda, MD; 3) National Human Genome Research Institute, Medical Genetics Branch, Bethesda, MD.

Introduction: We present a 5-year-old African-American male who was diagnosed with Niemann Pick C deficiency (NPC). His case is remarkable for some unusual features of NPC and for a set of DNA mutations that were challenging to detect with the exome analysis pipeline at the time. The clinical course began with normal development until 15-18 months of age, when he was noted to have difficulty walking. By age 2 he was ataxic; by 2-1/2 he could not stand independently. Communication skills also deteriorated: at 26 months he showed language comprehension and expression to the 9-12 month range. He became non-verbal by age 3. At age 4, he developed recurrent episodes of respiratory failure and had at least one episode of cardiopulmonary arrest secondary to aspiration. Extensive evaluations were performed prior to admission at NIH including muscle biopsy, chromosomal microarray, lysosomal storage panel, and biochemical measurements. Muscle pathology showed abnormal but non-diagnostic findings including abnormal mitochondrial morphology. MRI brain imaging showed non-specific white-matter changes. **Methods:** Abnormal results detected during his NIH admission included a nerve conduction study with decreased CMAP amplitude and decreased conduction velocity in several nerves. Needle EMG showed decreased insertional activity in all muscles and muscle atrophy. EEG revealed nearly continuous diffuse posterior predominant 1-2 Hz high voltage irregular spikes and slow wave discharges. Muscle mitochondrial DNA copy number was 29% suggestive of a mitochondrial depletion syndrome. Leukocyte coenzyme Q-10 levels were elevated at 188 picomole/mg. Abdominal ultrasound showed borderline large liver and spleen sizes (12. 2 cm and 9. 5 cm). **Results:** Research exome revealed two novel variants in *NPC1:* a c. 3266A>G mutation inherited from the mother and a de novo c. 2119C>T mutation. Follow up fibroblast cholesterol esterification was 0%. **Discussion/Conclusions:** This case of Neiman Pick C disease highlights a severe disease course. In addition, it was a challenge for our exome analysis pipeline. Mendelian-inconsistent mutations make up a substantial portion of the noise seen in exome genotype results. We use consistent segregation as a component of variant filtration. This case featured one allele that was Mendelian-inconsistent due to a true-positive new-dominant mechanism and has prompted us to develop new approaches to filter variants that may form compound heterozygous pairings.

2335T

Erythropoietic Protoporphyria: Investigation and Evaluation of Potential Pathogenicity of Ferrochelatse (FECH) Mutations in Genomic Databases. *W. Qiao, M. Balwani, R. Srinivasan, M. Yasuda, R. Desnick.* Icahn School of Medicine at Mount Sinai, New York, NY.

Erythropoietic protoporphyria (EPP) is an autosomal recessive inborn error of heme biosynthesis resulting from loss-of-function mutations in the ferrochelatase (*FECH*) gene, which encodes the enzyme that inserts iron into protoporphyrin IX (PP) to form heme. EPP is characterized by the accumulation of PP in the erythron, and patients experience severe phototoxic pain when exposed to sunlight. The prevalence of EPP in Western Europe ranges from 1:75,000 to 200,000. Most EPP patients identified in North America are Caucasians. ~4 % of EPP patients carry two *FECH* loss-of-function mutations, while the majority carry one *FECH* loss-of-function mutation and the common *FECH* low expression allele, IVS3-48T>C, which occurs in about 10% of Caucasians. These mutations reduce overall *FECH* activity to < 35% of normal, causing symptoms. To date, over 190 pathogenic *FECH* mutations causing EPP have been reported (Human Gene Mutation Database). However, the effect of *FECH* missense mutations on enzyme activity have not been characterized for most. Interrogation of the 1000 Genomes and the Exome Sequencing Project (*ESP*) databases revealed 33 *FECH* variants in all racial and ethnic groups, 29 missense variants and 4 at or near consensus splice sites, including 10 reported pathogenic mutations. In Caucasians, there were 21 missense and 3 intronic variants, including 9 reported pathogenic mutations. Overall frequency in Caucasians was about 16% for both databases, with the frequency of R96Q being ~15%, while the others ranged from 0. 01 to 2%. All 29 missense variants were evaluated by *in silico* pathogenicity prediction programs and eukaryotically expressed *in vitro* to determine their *FECH* catalytic activity. The *in silico* studies predicted 6 variants as “pathogenic”, 17 as “ambiguous” and 6 as “likely benign”. The *in vitro* expression assays revealed that 4 variants (encoding D274N, Q285R, P334L, L405F) had < 10% of expressed wild-type (WT) activity, confirming their pathogenicity. 13 Variants, including 2 previously reported mutations (G55C, P62R), had > 40% WT activity, suggesting that they are likely benign. The combined frequency of the G55C and P62R alleles among Caucasians is ~2. 0%, further supporting that they are not pathogenic. Taken together, these studies emphasize the importance of functional characterization of missense variants. .

2336F

Topic: Mucopolysaccharidosis type IIID (Sanfilippo syndrome type D): Initial characterization of the murine model. *M. Jamil[1], E. M. Snella[1], S. Q. Le[2], S. Kan[2], B. C. Birtcil[3], P. I. Dickson[2], N. M. Ellinwood[1], J. D. Smith[3].* 1) Iowa State University College of Agriculture and Life Sciences, Ames, IA USA; 2) Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, CA USA; 3) Iowa State University College of Veterinary Medicine , Ames, IA USA.

The mucopolysaccharidoses (MPSs) are lysosomal storage enzymopathies that result in storage of glycosaminoglycans (GAGs). Symptoms include somatic and CNS signs. Storage of the GAG heparin sulfate (HS) cause MPSs with fatal pediatric neurodegenerative disease. The MPS with predominate CNS disease is MPS III (Sanfilippo syndrome), a heterogeneous HS storage disorder of 5 subtypes (IIIA-IIIE). Of the four MPS IIIs in humans, all but IIID have mouse models. Herein we describe the first N-acetyleglucoseamine-6-sulfatase (GNS) deficient murine model of MPS IIID. A knockout mouse was generated at Taconic Artemis GmbH that removed exons 2 to 13 of the 14 exons in GNS. Both the founding the embryonic stem cells and the chimera mates were from a C57BL/6 background. The production of GNS -/- mice, diagnosed by PCR, confirmed to the expected numbers for a simple recessive Mendelian pattern. Initial characterization involved comparison of affected mice at 4 and 13 months of age with litter and sex matched GNS+/+ mice. At 13 months of age, the GNS-/- mouse had an enlarged urinary bladder (1. 1 ml urine), moderate hepatomegaly, decreased abdominal fat, and an unkempt pelage. Hematoxylin and eosin (HE) stained histological sections showed extensive vacuolation of hepatocytes, foamy vacuolation of segments of renal cortical tubular epithelia, vacuolation of neurons in the brain, and extensive loss of Purkinje cells seen in the cerebellum. Additionally, numerous neurons had cytoplasmic granular storage in Luxol fast blue (LFB) stained sections of brain. Four month old affected mice had no significant abnormalities seen on gross exam, mild vacuolation of hepatocytes on HE, and infrequently observed LFB positively stained neurons in the CNS. Biochemical analysis of 4 month old mouse tissues found no detectable GNS activity in affected versus control mice. The murine model of MPS IIID showed similar gross and histological findings to other MPS III murine models, including urine retention, and marked liver, renal, and CNS storage. Neuronal vacuolation and severe Purkinje cell loss is comparable to findings in other MPS III models. Extensive hepatocellular vacuolation is consistent with heparan sulfate storage, and the LFB positive staining of neurons is consistent with the known secondary CNS storage of glycosphingolipids in MPS III. This new model will be useful to further study the pathogenesis and therapy of human MPS IIID.

2337T

Large scale screening for lysosomal storage disorders in minority groups. *R. Iimgala[1,2], C. Sejpal[1,2], C. Cross[3], M. Gondre-Lewis[3], O. Goker-Alpan[3].* 1) Lysosomal and Rare Disorders Research and Treatment Center, Fairfax, VA; 2) O and O Alpan, LLC, Fairfax, VA; 3) Howard University School of Medicine, Washington, DC.

Introduction: Lysosomal Storage Disorders (LSDs) are a group of over 40 inherited disorders, where affected individuals exhibit symptoms that become progressively worse, often prior to access to testing and diagnosis. Although many LSDs are considered panethnic, the incidence for individual disorders is highly variable. LSDs challenge the current healthcare system, not only because of diagnostic delays, but also variable access to very expensive therapies. To highlight diagnostic challenges originating from ambiguous clinical manifestations, we initiated a large scale screening for "treatable forms" of LSDs with a special focus on under-represented minority groups. **Subjects:** The study population is comprised of anonymized patients in the Washington, D. C. metro area seeking healthcare for varied reasons. 85% of the patients are African-American and 7% of are Hispanic, and the remainder report as white, other, or American Indian. The aim is to compile results from at least 10,000 unique samples and include subjects of both genders in the screening. **Methods:** Under IRB approved protocols (NCT02120235 and IRB-14-MED-09), dried blood spots are prepared from anonymized peripheral blood samples within 24 hrs of draw and miniaturized fluorometric enzyme assays are performed using 4-Methylumbelliferyl substrates specific for a-galactosidase (Fabry disease), a-glucosidase (Pompe disease), b-glucosidase (Gaucher disease), chitotriosidase (Gaucher and Niemann-Pick C disease) and b-galactosidase (Gangliosidosis) enzymes. All assays are performed in triplicates in 384-well format and fluorescence is measured using a spectrophotometric plate reader. Reference ranges for all the enzyme assays are established using n=40 random normal subjects with no known symptoms of any of the LSDs. **Results and Conclusions:** Preliminary findings resulting in positive diagnosis are highlighted to indicate prevalence rates for individual LSDs in minority groups primarily represented by those of African-American and African and Afro-Caribbean descent. Incidence rates are compared to results from newborn screening for LSDs. Since our study focuses on adults and children already under clinical care for various symptoms, the study is expected to result in differential prevalence rates for the LSDs tested, and will highlight the importance of such large scale screening on early diagnosis of LSDs and prompt intervention in minority groups with varied access to healthcare and genetic testing.

2338F

Whole-exome sequencing identifies *ECHS1* mutations in Leigh syndrome. . *M. Tetreault[1,2], H. Antonicka[3], S. Fahiminiya[1,2], GA. Mitchell[4,5], MT. Geragthy[6], M. Lines[6], KM. Boycott[6], EA. Shoubridge[1,3], JJ. Mitchell[7], JL. Michaud[4,5,8], J. Majewski[1,2], Care4Rare Canada Consortium.* 1) Department of Human Genetics, McGill University, Montreal, QC H3A 1B1, Canada; 2) McGill University and Genome Quebec Innovation Center, Montreal, QC H3A 1A4, Canada; 3) Montreal Neurological Institute, McGill University, Montreal, QC H3A 2B4, Canada; 4) Department of Pediatrics, Universite de Montreal, Montreal, QC H3T 1C5, Canada; 5) CHU Sainte-Justine Research Center, Montreal, QC H3T 1C5, Canada; 6) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, ON K1H 8L1, Canada; 7) Department of Pediatrics, Montreal Children's Hospital, Montreal, QC H3H 1P3, Canada; 8) Department of Neurosciences, Universite de Montreal, Montreal, QC H3C 3J7, Canada.

Leigh syndrome (LS) is a rare heterogeneous progressive neurodegenerative disorder usually presenting in infancy or early childhood. Clinical presentation is variable and includes psychomotor delay or regression, acute neurological or acidotic episodes, hypotonia, ataxia, spasticity, movement disorders, and corresponding anomalies of the basal ganglia and brain stem on magnetic resonance imaging. To date, 35 genes have been associated with LS, mostly involved in mitochondrial respiratory chain function and encoded in either nuclear or mitochondrial DNA. We used whole-exome sequencing to identify disease-causing variants in four patients with basal ganglia abnormalities and clinical presentations consistent with LS. Compound heterozygote variants in *ECHS1,* encoding the enzyme enoyl-CoA hydratase were identified. One missense variant (p. Thr180Ala) was common to all four patients and the haplotype surrounding this variant was also shared, suggesting a common ancestor of French-Canadian origin. Rare mutations in *ECHS1* as well as in *HIBCH*, the enzyme downstream in the valine degradation pathway, have been associated with LS or LS-like disorders. A clear clinical overlap is observed between our patients and the reported cases with ECHS1 or HIBCH deficiency. The main clinical features observed in our cohort are T2 hyperintense signal in the globus pallidus and putamen, failure to thrive, developmental delay or regression, and nystagmus. Respiratory chain studies are not strikingly abnormal in our patients: one patient had a mild reduction of complex I and III and another of complex IV. The identification of four additional patients with mutations in *ECHS1* highlights the emerging importance of this pathway in LS.

2339T

Measurement of markers of urinary oxidative stress in methylmalonic aciduria, propionic aciduria and Barth syndrome. *K. Victor, H. Vernon.* McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

Oxidative stress is an increasingly recognized mechanism of cellular pathology in disorders of intermediary mitochondrial metabolism, and defining its' role in specific disorders could have implications for targeted therapy. Therefore, in order to explore the possible imbalance between the production of reactive oxygen species and antioxidant reserve in three disorders involving mitochondrial function, methylmalonic aciduria (MMA) and propionic aciduria (PA), Barth syndrome (BTHS), we assessed multiple urinary markers of oxidative stress. These disorders are classified as organic acidurias, but have different roles in mitochondrial intermediary metabolism. MMA and PA are closely related disorders of branch chain amino acid metabolism, and Barth Syndrome is a disorder of cardiolipin remodeling. We measured protein oxidation (via urine di-tyrosine levels), lipid oxidation (via urine 15-F2t-isoprostane level) and urine antioxidant capacity in 8 individuals with MMA, 4 individuals with PA , 8 individuals with BTHS, and 5 controls. The average urine di-tyrosine level in MMA was 4. 87x105 FU/ mg Cr +/-1. 28x105 (SD), in PA was 9. 39x105 FU/ mg Cr +/-2. 72x105(SD), in BTHS was 1. 17x106 FU/ mg Cr +/-9. 83x105(SD), and in controls was 1. 43x105 FU/mg Cr +/-5. 74x104(SD). MMA, PA and BTHS samples all had significantly higher di-tyrosine levels when compared to controls (p-values <0. 05). The urine 15-F2t- isoprostane levels, as determined by a competitive enzyme-linked immunosorbent assay (ELISA), was not significantly different between MMA, PA, BTHS or control groups. The urine antioxidant capacity in MMA was 0. 83 mM Trolox +/-0. 53 (SD), in PA was 1. 34 mM Trolox+/-1. 37 (SD), in BTHS was 1. 21 mM Trolox+/-0. 72 (SD), and in controls was 2. 71 mM Trolox+/-1. 12 (SD). The controls had significantly higher urine antioxidant capacity than the MMA cohort (p-value of 0. 007) and the BTHS cohort (p-value of 0. 03). These results suggest that oxidant stress may play a significant role in the pathophysiology of these three mitochondrial-associated disorders, but with differing biochemical consequences, and that correction of antioxidant capacity could be a viable target for therapeutic intervention in MMA, BTHS, and possibly PA.

2340F

Can a heterozygous c. 1880_1881del (p. S627fs) novel pathogenic mutation in the NDUFS1 gene cause mitochondrial Complex I deficiency? *S. Yang, D. Regier, K. Chapman.* Genetics & Metabolism, Children's National Health System, Washington, DC.

Liver involvement can be seen as a common feature in childhood mitochondrial disorder; it may appear as neonatal or early childhood liver failure. The numbers of genes that have been associated with mitochondria disorders have continued to expand. NDUFS1 is a nuclear-encoded mitochondrial gene which encodes the largest subunit of respiratory chain complex I. Complex I deficiency is the most frequently encountered in mitochondrial disorders, and NDUFS1 is one of several mutational hot spot genes for isolated complex I deficiency. We present a family with two children who both had acute fulminant liver failure around 10 months of age. The older girl recovered, however her younger brother died before a liver transplant could be performed. The 7-year-old girl also has had elevated transaminase levels and abnormal LFTs in conjunction with illness. Otherwise, she is healthy and developmentally normal in between illness episodes. Family history is positive for other paternal family members (including father, uncle, and possible grandfather) with asymptomatic elevated liver enzyme. Along with many differential diagnoses for these relatives, mitochondrial dysfunction was suspected. The Whole Exome Sequencing was offered to the family after an extensive amount of genetic testing was done without providing a diagnosis. A paternally inherited novel variant in the NDUFS1 gene was reported, and this variant is predicted to be pathogenic. Targeted mutation analysis was subsequently performed on the deceased brother's tissues as well as the paternal uncle who also has had abnormal LFTs. Both of them were found to carry the same pathogenic mutation. Since only one pathogenic variant in NDUFS1 gene was identified in multiple family members, either with acute liver failure or a history of asymptomatic elevated liver enzyme, the assumption we have is that this is an autosomal dominantly inherited disease. We propose that the heterozygous c. 1880_1881del (p. S627fs) mutation represents a novel pathogenic variant in NDUFS1 and acts in a dominant manner.

2341T

Evaluating Neutral Lipid and Phospholipid Contents of Lipid Droplets in Fibroblast Cells from Patients with Peroxisomal Disorders. *W. Fallatah[1,2], P. Dranchak[3], JG. Hacia[4], NE. Braverman[1].* 1) Department of Human Genetics, McGill University-Montreal Children's Hospital Research Institute, Montreal, QC, Canada; 2) Department of Genetic Medicine, King Abdul-Aziz University, Jeddah, Saudi Arabia; 3) National Center for Advancing Translational Sciences, National Institutes of Health, Rockville, MD, USA; 4) Department of Biochemistry and Molecular Biology, Broad Center for Regenerative Medicine and Stem Cell Research, University of Southern California, Los Angeles, CA, USA.

Peroxisomes and lipid droplets (LDs) are discrete intracellular organelles that play a significant role in regulating cellular lipid metabolism. While the main functions of peroxisome are fatty acid beta-oxidation and ether phospholipid biosynthesis, LD has a critical role in regulating intracellular lipid storage. During cellular stress like starvation the cell can survive by utilizing the fat stored in LDs as a source of energy. Accumulation of LDs has been reported in C. elegans with defective peroxisomal b oxidation (MAOC-1/DHS- 28/DAF-22 genes) and defective peroxisome assembly (PRX10, the ortholog of the human PEX10 gene). LDs also accumulate in liver tissue from mice with defective peroxisome assembly (Pex11α). We previously studied the neutral lipid and phospholipid contents of LDs in peroxisome-deficient Chinese Hamster Ovary (CHO) cells and we found that the neutral lipid content of the LDs is significantly higher in PEX2-null CHO cells during normal cellular growth and under starvation compared to wild type CHO cells. Therefore peroxisomes are likely to play a role in regulating LDs. Our objective of the current study is to evaluate the neutral lipid and phospholipid contents of LDs in fibroblast cell lines from patients with peroxisomal disorders under normal and starvation condition. We are studying LD numbers by their content of either neutral lipids or phospholipids in fibroblast cell lines from patients with defective peroxisome assembly in 13 different PEX genes. We will use the green neutral lipid and red phospholipid specific immunofluorescence staining (LipidTOX™) to compare the contents of LDs between control and PEX gene-deficient fibroblast cell lines under normal and starvation conditions. This study highlights another pathway of intracellular organelle communication between peroxisomes and LDs, and helps us to understand pathophysiology in the peroxisomal biogenesis disorders.

2342F

Acute intermittent porphyria (AIP): high incidence of pathogenic hydroxymethylbilane synthase non-synonymous variants in genomic databases suggests other predisposing genetic and/or environmental factors for acute attacks. *B. Chen, J. Hakenberg, C. Solis-Villa, R. Srinivasan, D. Doheny, I. Peter, R. Chen, D. Bishop, M. Yasuda, R. Desnick.* Genetics and Genomics Sciences, Mount Sinai School of Medicine, New York, NY.

AIP is an autosomal dominant hepatic porphyria due to the half-normal activity of the heme biosynthetic enzyme, hydroxymethylbilane synthase (HMBS). Symptomatic individuals experience life-threatening acute neurovisceral attacks that are triggered by factors that induce the up-regulation of hepatic 5'-aminolevulinic acid synthase (ALAS1) with resultant accumulation of neurotoxic porphyrin precursors, aminolevulinic acid (ALA) and porphobilinogen (PBG). To date, over 360 *HMBS* pathogenic mutations have been reported in the Human Gene Mutation Database (HGMD). The disease prevalence in Western Europe was estimated at 1 in 160,000 to 220,000 (Elder et al, 2013). However, the frequency of pathogenic *HMBS* mutations that markedly reduce enzyme activity and the actual disease penetrance remain unknown. To estimate the prevalence of likely pathogenic *HMBS* mutations, we interrogated the 1000 Genomes Project and the NHLBI Exome Sequencing Project (ESP) databases. Among Caucasians, 13 non-synonymous (NS) and 9 synonymous (S) variants were identified. Allele frequencies for the 13 NS variants ranged from <0. 01% to 0. 28%. Of the 9 S variants, 4 encoding T35T, S45S, L161L, V202V were predicted to affect splice site enhancers or silencers, possibly causing splicing defects. Of the 13 NS variants, 4 encoding E86V, R195C, R225Q, R321H were reported as pathogenic in the HGMD. All NS variants were evaluated by 14 *in silico* prediction programs and prokaryotically expressed *in vitro* to assess their pathogenicity. *In silico* programs predicted 5 variants as "pathogenic" (encoding D65H, I71T, A122P, R195C, R225Q), 2 (R246H, R355Q) as "ambiguous", the remaining 6 as "likely benign". *In vitro* HMBS expression revealed that A122P and R195C mutant proteins are pathogenic with ~ 3% of expressed wild-type (WT) activity**.** All other NS variants, including 3 previously reported in HGMD (E86V, R225Q, R321H), had 60 – 110% of expressed WT activity and had > 60% of WT thermostability**.** Notably, the combined frequency of the 2 pathogenic *HMBS* alleles, A122P and R195C, is ~ 0. 02% (1 in 2500). Thus, the penetrance of AIP would be low, ~ 2% or lower among AIP heterozygotes, as the prevalence of patients with acute attacks is estimated ~1 in 200,000 in Europe. These results reiterate the importance of functional studies of gene variants and suggest that modifying genes and/or environmental factors are involved in the pathogenesis of acute attacks of AIP.

2343T

X chromosome inactivation in 82 female patients with Fabry disease. *L. Echevarria[1,2], I. E. Jurca-Simina[2], K. Benistan[2], S. Ben Khaled[2], C. Boucly[2], L. Cuisset[3], L. Gadi[2], F. Jabbour[2], C. Jauny[2], S. Medaouri[1], R. Y. Carlier[2], D. P. Germain[1,2,4].* 1) University of Versailles, Division of Medical Genetics, 78180 Montigny, France; 2) French Referral Center for Fabry disease and inherited disorders of connective tissue, 92380 Garches, France; 3) University Paris V Descartes, Laboratory of Biochemistry and Molecular Biology, CHU Cochin, 75014 Paris, France; 4) University of Versailles, UMR 1179, 78180 Montigny, France.

Background: Fabry disease (FD) is an X-linked genetic disorder caused by the deficient activity of lysosomal alpha-galactosidase. While males are usually severely affected, clinical presentation in female patients may be more variable ranging from asymptomatic to, occasionally, as severely affected as male patients. **Aims**: To evaluate the existence of skewed X chromosome inactivation (XCI) in females with FD, its concordance between tissues, and its contribution to the phenotype. **Methods**: 82 females with FD were enrolled. Clinical and biological work-up included two global scores (MSSI and DS3), cardiac magnetic resonance imaging, measured glomerular filtration rate using the 51Cr EDTA method, cerebral magnetic resonance imaging, pain assessment, and measurement of alpha-galactosidase activity. XCI was analysed in 4 tissues using *HUMARA*, *PCSK1N* and *SLITRK4* genes methylation studies. **Results**: Skewed XCI was found in 26% of the study population. A correlation was found in XCI patterns between blood and the other analysed tissues. Significant differences in residual alpha-galactosidase levels, severity scores, progression of cardiomyopathy, deterioration of kidney function and distribution of angiokeratoma, with the direction and degree of skewing of XCI were evidenced. Analyses of neurological involvement are ongoing. **Conclusion**: X chromosome inactivation is a significant modifier of the clinical phenotype in female patients with Fabry disease.

2344W

Trisomy 8 mosaicism in Colombia. *P. Paez[1], AM. Gómez[2].* 1) Universidad el Bosque, Facultad de Medicina, grupo IINGM, Bogotá, Colombia; 2) Hospital Universitario Clínica San Rafael, Bogotá, Colombia.

Introduction. We describe a case of trisomy 8 mosaicism (T8M) in Colombia; this is the second case from Colombia reported in the literature and the first one of an adult female that has been followed up for a long time. We compare the findings of this case with those reported previously, including the first Colombian patient. Case description. A 19-year-old female born from a normal first pregnancy and from healthy parents was referred for evaluation because of developmental delay. The child showed speech difficulties and cognitive delay. She presents facial dysmorphism and orthopedic anomalies: camptodactyly of second through fifth fingers, scoliosis, and deep plantar creases; urogenital anomalies: left double urinary collecting system and right double renal pelvis. Her karyotype revealed a trisomy 8 mosaicism (47,XX,+8[40]/46,XX[10]), and neuroimaging showed agenesis of corpus callosum. Discusion. Warkany syndrome or trisomy 8 is a rare clinical condition with a frequency of about 1:25000 to 1:50000 births that more frequently appears as trisomy 8 mosaicism (85%). The prevalence of this syndrome in Colombia is unknown. This patient presents the classical phenotype: facial dysmorphism, camptodactyly, deep plantar furrows, cognitive delay, urogenital anomalies, and agenesis of corpus callosum. These findings were less flourished in the first Colombian case that only presented facial dysmorphism, developmental delay, and scoliosis. This is in agreement with the high phenotypic variability described among affected patients. About 25% of patients with T8M present congenital heart defects and there is an increased incidence of hematologic malignancies. Nevertheless, Colombian patients are negative for cardiac disease or malignances. The condition arises post-cygotically through a mitotic error with the consequence of a dosage effect. However, the phenotypic variability between patients (like that of Colombian patients) has not yet been understood. Recently, it has been described in trisomic fibroblasts a characteristic expression and methylation pattern even on genes at other chromosomes, which in cooperation with the gene dosage effect of genes on the chromosome 8 may explain the clinical features and elevated cancer risk seen in these patients. Finally, the deregulation of some genes like AGPAT6, CPXM2, CRYAB, has been reported as an explanation for cognitive delay.

2345T

A genetic liability model explains the sex bias in neurodevelopmental disorders. *A. Polyak[1], J. Rosenfeld[2,3], S. Girirajan[1,4].* 1) Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Signature Genomic Laboratories, PerkinElmer, Inc. , Spokane, WA; 4) Department of Anthropology, Pennsylvania State University, University Park, PA.

Neurodevelopmental disorders such as autism and intellectual disability/developmental delay (ID/DD) have been associated with a significant sex bias with a diagnosis skewed towards males. A systematic assessment of this bias has been confounded by factors such as phenotypic and genetic heterogeneity, and a history of neuropsychiatric phenotypes often reported in the families of affected individuals. To assess the extent of sex bias within these factors, we analyzed the frequency of comorbid features, magnitude of genetic load, and the existence of family history within 32,155 individuals ascertained clinically for autism or ID/DD, including subset of 8,373 individuals carrying rare copy number variants (CNVs). We find that females are more likely to manifest comorbid features within both autism (p=2. 9×10-6, OR=1. 34) and ID/DD (p=7. 2×10-4, OR=1. 08) cohorts. The male to female ratio of ascertainment for ID/DD and autism was contingent upon the presence of specific comorbid features, ranging from 8:1 for autism with psychiatric comorbidities, to 2. 5:1 for autism with epilepsy. This ratio was also contingent on specific CNVs, ranging from 8:1 observed in autistic individuals carrying 22q11. 2 duplication and 1. 3:1 in those carrying 16p11. 2 deletion. Males between ages 12 and 18 years showed a higher frequency of comorbid features than those younger than 5 years for autism (p=0. 004, OR=1. 10) or ID/DD (p=3. 4×10-4, OR=1. 20). However, females showed little change in the frequency of comorbid features across the developmental stages. Further, females showed a higher frequency of large CNVs compared to males when comorbidities were considered separately or in aggregate, a phenomenon, which diminished when applied to severely ascertained cohorts. Affected males showed family histories such as autism (p=0. 04) or specific learning disabilities (p=0. 05) while females showed family histories of growth abnormalities (p=0. 03). Our results suggest a model where the genetic liability for the manifestation of neurodevelopmental phenotypes exist at different thresholds between males and females. These thresholds can be altered by the frequency and type of comorbid features, presence of disease-associated deletions and duplications, and a pre-existing family history of related phenotypes.

2346F

De novo deletion 2 (q 14. 1q21. 1) and association with attention deficit and hyperactivity. *J. G. Pappas[1], K. E. Daley[1], M. J. Messito[2], E. Brent[3], K. M. Hopkins[3].* 1) Dept Pediatrics, Clin Genetic Scvs, New York Univ, Sch Med, New York, NY; 2) Dept Pediatrics, New York Univ, Sch Med, New York, NY; 3) Dept Pediatrics, Division of Developmental and Behavioral Pediatrics, New York Univ, Sch Med, New York, NY.

De novo interstitial deletions of chromosome 2q that include band 2q14. 3 are rare. A published case report that includes a review of eight cases with variable breakpoints revealed common facial features, cognitive disability and attention deficit hyperactivity disorder (ADHD) (Grealy MT et al, 2014). We present an eight year old boy with de novo interstitial deletion 2q, language and learning difficulties (LLD) and severe ADHD. There is only one published case with deletion breakpoints similar to ours (Baker KL et al, 2001). Our case was born full-term to 21 year old. Pregnancy and delivery were uncomplicated. His birth weight was 3650 grams at the 75th centile, birth length 52 cm at the 75th centile and his head circumference 35. 5 cm at the 50th centile. He achieved his early motor milestones normally and he walked at age 15 months. Speech delay was ascertained at age two because he was unable to combine words and speech therapy was initiated. He had clinical and laboratory genetic evaluation at age three and a half. His weight was 19. 6 kilograms at the 97th centile, height 106. 5 cm at the 97th centile and his head circumference 50. 5 cm at the 50th centile. He had elongated face, large forward facing ears, single midline hair whorl and likely familial large forehead and midface hypoplasia present also in his mother. Routine peripheral blood karyotype reveled 46,XY,del(2)(q13q21. 1). Parental karyotypes were normal. Cardiac evaluation including echocardiography and renal ultrasound were normal in our case. Follow-up at age 7 revealed severe ADHD and LLD. Microarray (oligonucleotide, single nucleotide polymorphism, Affymetrix CytoScan HD) defined the breakpoints of the deletion as 2q14. 1q21. 1(114,949,093-131,610,682)x1 [hg19]. Our case shares ADHD, LLD, facial features and tall stature with the only published case with similar deletion (Baker KL et al, 2001). Other cases with deletions overlapping our case's deletionpresented with cognitive disability and various craniofacial abnormalities with large prominent forehead in six out ofnine published cases (Grealy MT et al, 2014). The ADHD was a striking feature of our case and it is shared with the published case with similar breakpoints. The CNTNAP5 gene is included in the deletion and its haploinsufficiency may be associated with ADHD (Pagnamenta AT et al, 2010). Our case contributes to delineation of the phenotype of the deletion 2(q14. 1q21. 1) and the association of CNPNAP5 haploinsufficiency with ADHD.

2347W

10-year-old Female with *KANSL1* Mutation and First Reported Case of Normal Intelligence. *C. Keen[1], C. Samango-Sprouse[1, 2, 3, 4], H. Dubbs[5], E. H. Zackai[6].* 1) The Focus Foundation, Davidsonville, MD; 2) Neurodevelopmental Diagnostic Center, Davidsonville, MD; 3) George Washington University, Department of Pediatrics; 4) Florida International University, Department of Molecular Genetics; 5) Children's Hospital of Philadelphia, Division of Neurology; 6) Children's Hospital of Philadelphia, Division of Clinical Genetics.

Introduction: Koolen-de Vries syndrome (KDVS), first described in 2006, occurs in 1 in 16,000 births and is caused by the microdeletion or mutation of the *KANSL1* gene located on chromosome 17q21. 3 [Koolen, 2006]. Phenotypic characterization includes epilepsy, cardiac defects, kidney problems, and skeletal anomalies [Koolen, 2006; Koolen, 2008; Tan, 2009]. Developmental delay, mild-to-moderate intellectual disability, a cheerful disposition, hypotonia, and distinctive facial features are also described. Speech and language-based difficulties, graphomotor dysfunction, and slow feeding are not uncommon in the current literature. To our knowledge, this is the first case of a child with *KANSL1* mutation with normal IQ and Childhood Apraxia of Speech (CAS). She has also been diagnosed with ADHD, and graphomotor dysfunction which is associated with Developmental Dyspraxia (Dev Dys). Methods: Comprehensive neurodevelopmental assessments and parent behavioral questionnaires were administered. All neurodevelopmental domains were probed including neuromotor, intelligence, speech and language, and executive function. All assessments were standardized on large normative samples in the United States. Results: Formal assessment indicated a low-normal IQ (VIQ=83) and several subtest scores within normal limits on the WISC-IV. She has a scaled score of 7 on similarities, comprehension and word reasoning. CAS was diagnosed early, characterized by hypotonia, slow feeding as a newborn, and speech delay. Dysarthria was apparent when longer speech patterns developed. She had below average performance on the BOT-2 in balance and strength, a reflection of truncal hypotonia. She also had below average performance on fine motor function supporting dysgraphia with graphomotor dysfunction. Conclusion: To our knowledge, this is the first reported case of normal IQ, Dev Dys, and CAS in KDVS. Dysgraphia and graphomotor dysfunction are further documentation of Dev Dys. In our patient, delayed speech, feeding issues, truncal hypotnoia are reflective of CAS, dysarthria and Dev Dys. This case expands the mild end of the developmental spectrum seen in children with KDVS and *KANSL1* mutation.

2348T

An 8. 3Mb 19p13. 2p13. 3 duplication associated with duodenal atresia and left ventricular noncompaction. *Y. Dowa[1,2], I. Imoto[3], M. Kawai[1], T. Heike[1].* 1) Department of Pediatric Neurology, Gunma Children's Medical Center, Shibukawa, Gunma, Japan; 2) Department of Neonatal Intensive Care Unit, Kyoto University Hospital, Kyoto, Kyoto, Japan; 3) Department of Human Genetics, Institute of Biomedical Sciences, Tokushima University Graduate School.

We report a new case of duplication of 19p13. A preterm baby admitted our neonatal intensive care unit because of very low birth weight, respiratory distress and suspicion of duodenal atresia. She was the first child of healthy non consanguineous parents. She was born at 35 weeks and 1day of gestation by caesarean section because her mother suffered from intrauterine infection and her cardiotocograph indicated non-reassuring fetal status. Her birth weight was 1321g (-3. 0 SD), height was 36cm (-3. 2 SD) and occipitofrontal circumference was 23. 5cm (-4. 4 SD). She presented severe intrauterine growth retardation, microcephaly, sparse hair, sparse eyebrows, epicanthal folds, flat philtrum, thin upper lip, microstomia, accessory ears, hypoplastic fourth toes and nails. The X-ray images revealed duodenal atresia. The blood cell count revealed severe thrombocytopenia and cardiac ultrasonography showed left ventricular noncompaction. She had also congenital stridor and feeding difficulty. G-banded chromosomal test and subtelomere FISH revealed direct duplication of 19p13. Her karyotype was 46,XX,dup(19)(p13. 2p13. 3). Further, we performed array-based copy number analysis (Affymetrix CytoScan HD). It showed a 19p13. 3 duplication of approximately 8. 3Mb which was consistent with the FISH results, a 5q35. 3 duplication of approximately 0. 2Mb and 11q11 deletion of approximately 0. 3Mb. As both 5p35. 3 duplication and 11q11 deletion were very small, they regarded asymptomatic changes. On the other hand, her facial appearance and the characteristic toes were compatible to previous reported duplication of 19p. But the complications of duodenal atresia and left ventricle noncompaction did not report yet. Our findings with a review of literatures allow clarification of a more precise and comprehensive phenotype-genotype correlation for 19p duplication.

2349F

Pallister–Killian Syndrome: Clinical and Cytogenetics Variability of five Egyptian patients. *M. M. Eid[1], O. M. Eid[1], S. Abdel Hadi[2], N. Farouk[3], G. El Kamah[4], H. H. Afifi[4], A. Median[1], G. M. H. Abdel Salam[4].* 1) Human Cytogenetics, National research center, Cairo, Egypt; 2) Pediatric Department, Cairo University, Cairo, Egypt; 3) Orodental Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt; 4) Clinical Genetics Department, Human Genetics and Genome Research Division, National Research Centre, Cairo, Egypt.

Background: Pallister–Killian syndrome (PKS) is a rare sporadic genetic disorder caused by a mosaic tetrasomy of chromosome 12p, which manifests with characteristic facial features, intellectual disabilities, skin pigmentation abnormalities and epilepsy. The diagnosis of PKS can be complicated due to the phenotypic variation and the overlap with other syndromes that makes the molecular cytogenetic test necessary to confirm the diagnosis. **Method:**Five patients clinically suspected to have the characteristic facial feature for PK were included in this study. Pedigree construction, thorough clinical examination, radiological investigation and EEG. Heparinized blood sample for karyotyping and buccal samples or skin biopsy for FISH analysis were collected. **Results:**The ages of our patients ranged from 10 months to 2 years. All the patients have the characteristic clinical findings that present in PK patients in form of bitemporal alopecia, hypertelorism, and abnormal ears. Further, all patients had pigmentary skin anomalies, and hypotonia. Although seizures and epilepsy are of common features in PKS and some may consider it part of the triad for the diagnosis that include also characteristic facial feature, skin pigmentation abnormalities, non of our patients exhibited any of them and their EEG examinations were normal. Also, another interesting finding in our patients is the presence of hypogenesis or very thin corpus callosum in all patients. No major organ was affected. FISH analysis using sub telomere 12p probe on skin fibroblast or buccal samples could clearly confirm the diagnosis by the presence of 4signals of the probe instead of normal 2. **Conclusion:**The patients included in this study have shared the characteristic features those previously reported for PKS, however they differed in some clinical aspects. The percentage of the tetrasomy 12p might explain the variability in the severity of the clinical condition that could be attributed to the gene dose that present in the supernumerary chromosome.

2350W

Ring chromosome 15: expanding the phenotype. *H. El-Bassyouni[1], M. Eid[2], O. Eid[2], S. Hammad[2], A. El-Gerzawy[2], M. Zaki[1], M. El-Ruby[1].* 1) Clinical Genetics, National Research Centre, Cairo, Egypt; 2) Cytogenetics Department, National Research Centre, Cairo, Egypt.

Ring chromosome 15 is a rare disorder, with less than 50 cases reported in the literature to date. We report the **c**linical and cytogenetic evaluation of a patient with ring chromosome 15. Diagnostic tests including echocardiography, abdominal ultrasound, brain computerized tomography (CT), magnetic resonance imaging (MRI) and electroencephalogram (EEG) were done. Clinical examination of the patient revealed the characteristic features of ring chromosome 15, such as growth retardation, hypertelorism, frontal bossing, a highly arched palate, small hands and feet and café-au-lait spots. In addition, the patient presented with a mild intellectual disability, a congenital atrial septal heart defect, and abnormal EEG records. We also report 2 novel findings, which to our knowledge; have not been reported before in ring chromosome 15 patients: large areas of hyperpigmentation on the front of both legs and feet and hypogenesis of the corpus callosum. Cytogenetic studies using both conventional G-banding and fluorescence in situ hybridization (FISH) with a Sub Tel 15q probe confirmed the diagnosis of ring chromosome 15. .

2351T

Puberty evolution in Jacobsen Syndrome. *C. M. Jurca[1,2], K. Kozma[1,2], O. Iuhas[1], A. Jurca[2], C. Cladovan[1,2], M. Bembea[1,2].* 1) Genetics Department, Municipal Clinical Hospital dr. Gavril Curteanu Oradea, Oradea, Bihor, Romania; 2) Faculty of Medicine and Pharmacy, Oradea.

BACKGROUND: Jacobsen Syndrome (JS) is a contiguous gene syndrome caused by partial deletion of the long arm of chromosome 11. It is a rare syndrome, first described in 1973 by Jacobsen. About 200 cases have been reported thus far. The incidence is about 1/100,000 newborns, with a female/male ratio of 2:1. The key features of the syndrome are prenatal and postnatal growth failure, mental retardation, craniofacial dysmorphism, and thrombocytopenia. Approximately a quarter of children with JS die in infancy. There are very few observations regarding the pubertal period of patients with JS. OBJECTIVES: Identify, describe and monitor pubertal characteristics of patients with JS. CASE REPORT: We present the case of a patient, now a 22 year-old female, who has been monitored in our Genetics Department since the age of 3 months. She had intrauterine growth retardation, failure to thrive, and feeding difficulties. Phenotypically she shows short stature, severe mental retardation, and characteristic craniofacial dysmorphism suggesting JS: microcephaly, trigonocephaly, hypertelorism, bilateral iris coloboma, down-slanting palpebral fissures, palpebral ptosis, strabismus, epicanthic folds, small nose, depressed nasal bridge, thin upper lip, large mouth, small lower jaw, and low set ears. Bilateral clynodactyly, camptodactyly and simian creases are present. She learned to walk at the age of 4 years; she shows spasticity and has profound delay in gross and fine motor skills. She has severe learning difficulties, communicates non-verbally, and displays heteroaggressive and autoaggressive behavior, which have been more pronounced after puberty. She has a mild hemorrhagic tendency. Primary amenorrhea and genital infantilism are evident. Hematological (pancytopenia) and immunological (IgM deficiency) anomalies, absent in infancy, became apparent at puberty. Karyotype: 46, XX del (11) (q23. 3-qter). DISCUSSION: This is an unusual case of JS by rare frequency and uncommon survival. We identified evolution through puberty, with hematologic and immunologic changes, primary amenorrhea, genital infantilism and aggressive behavior. CONCLUSIONS: Patients with JS need special monitoring during puberty. KEY WORDS: Jacobsen Syndrome, mental retardation, craniofacial dysmorphism, thrombocytopenia, puberty.

2352F

A rare case with chromosome 10p15. 1p11. 22 duplication. *E. Kirat, H. Ulucan, G. Guven, A. Koparir, N. Bilge, M. Seven.* Istanbul University Cerrahpasa Medical faculty, Istanbul, Turkey.

Trisomy 10p is a rare chromosomal abnormality which is characterized by severe growth and mental retardation, craniofacial anomalies including dolichocephaly, delayed closure of fontanels, frontal bossing, hypertelorism, up-slanting palpebral fissures, broad nasal root, distinctive mouth, and high arched and cleft palate, camptodactyly, club foot and congenital heart defects. Clinic features vary according to part of trisomic chromosomal region. Here we present a 6-year-old boy with partial trisomy 10p. He was referred to our clinic for developmental delay and dysmorphic features such as dolichocephaly, slanting palpebral fissures, broad nasal root, and micro/ retrognathia. He also had bilateral camptodactyly and club foot. Chromosome analysis showed 46,XY-,der(19)ins(19;10)(q13. 1;p15p11. 2). We analyzed this trisomic region in detail by SNP array as arr[hg19] 10p15. 1p11. 22(5,537,569-33,489,889) x3. This rearrangement with the specific phenotype is discussed reviewing the literature.

2353W

An extremely rare case of trisomy 8 and trisomy 21 mosaicism. *K. Kozma[1,2], O. Iuhas[1], C. M. Jurca[1,2], V. Filip[2], C. Nagy[1], M. Bembea[1,2].* 1) Genetics Department, Municipal Clinical Hospital dr. Gavril Curteanu Oradea, Oradea, Bihor, Romania; 2) Faculty of Medicine and Pharmacy Oradea.

BACKGROUND: Trisomy 8 mosaicism (T8M), also known as Warkany syndrome 2, is a very rare chromosomal abnormality (1:25,000 – 1:50,000 live births) defined by the presence of three copies of chromosome 8 in some cells of the organism (46/47+8). Mosaicism of trisomy 8 and another autosome is extremely rare in live births but was described in some miscarriages and in chronic myeloid leukemia. CASE REPORT: We report the case of a 2 week old baby boy evaluated for craniofacial dysmorphism. He shows a square face, frontal bossing, large anterior fontanel, hypertelorism, deep set eyes, broad nasal bridge, broad/bulbous nose, upturned nostrils, macrostomia, everted lips, micrognathia, high vaulted palate, frenulum of the superior lip adherent to the midline of gingiva, uncommon hypertrophy of the gum at the level of the superior central incisors teeth, dysplastic ears with partial absence of the helix, prominent antihelix, and large lobules. Other traits: deep palmar and plantar creases, abnormal dermatoglyphics, broad toes, prominent calcaneus, short neck with extra skin fold, narrow shoulder, elongated trunk, large mongolian spot of the back, bilateral hydrocele. Radiological findings: hypoplasia of viscerocranium, wormian bones of posterior fontanel and occiput, blurred cranial sutures, sclerosis of orbits and base of skull, large, curved ribs, vertebral anomalies. Organ anomalies: ventricular and atrial septal defect, agenesis of the corpus callosum, agenesis of right kidney, left hydronephrosis. Karyotype: 47,XY,+8(43)/47,XY,+21(10). DISCUSSION: To our knowledge no kindred have been presented with this type of mosaicism in the literature. We discuss the mechanism of the anomaly, the impact for prenatal diagnosis, as well as the need for early postnatal diagnosis, and better management. CONCLUSIONS: This unique case is a challenge in diagnosis and management of a rare trisomic mosaicism. KEY WORDS: chromosomial mosaicism, trisomy 8, trisomy 21, craniofacial dysmorphism.

2354T

Genotype-phenotype correlation of an 8p complex rearrangement. *L. Martelli[1,2], C. H. P. GRANGEIRO[1,2], A. G. GOMES[1], T. M. JOAQUIM[1], L. A. F. LAUREANO[2], J. D. GRZESIUK[1], F. G. O. GENNARO[1], J. HUBER[2], J. A. SQUIRE[1,3].* 1) Genetics Dept, Ribeirao Preto Medical School, University of Sao Paulo, Ribeirao Preto, SP, Brazil; 2) Medical Genetics Dept, Clinical Hospital, Ribeirao Preto, SP, Brazil; 3) Pathology Dept, Ribeirao Preto Medical School, University of Sao Paulo, Ribeirao Preto, SP, Brazil.

Partial duplication 8p is associated with congenital anomalies and different levels of mental retardation. However, the overall spectrum depends on the size and genomic location of the duplicated region. Inverted duplication 8p associated with 8p deletion is an uncommon complex chromosomal rearrangement, with an incidence of 1 in 10,000–30,000 liveborns. We describe a female child initially evaluated by obstetric ultrasonography with corpus callosum agenesis, single umbilical artery and ureteropelvic junction obstruction. Physical examination showed craniofacial dysmorphisms - macrobrachycephaly, arching eyebrows, upslanting palpebral fissures, small nose with anteverted nares, small mouth, hirsutism - and global hypotonia. Cranial MRI indicated corpus callosum agenesis and reduced white matter. US urinary tract showed marked left pyelocaliceal dilatation secondary to obstructive uropathy and cystourethrography with neurogenic/overreactive bladder. The proband had severe delay in psychomotor development, skeletal deformities (kyphosis, pectus excavatum and planovalgus feet) and recurrent urinary infections. Chromosomal analysis by GTG banding revealed 46,XX,add(8p),inv(9)(p13q13). Her father had the same inversion and maternal karyotype was normal. Evaluation by aCGH (2x400k Agilent®) showed a 3. 7Mb deletion at 8p22. 2-22. 3 and four gains at 8p11, 8p12 and 8p22 (two regions) corresponding to 7Mb. Deletion on 8p22. 2-22. 3 involved the *MIR596, ARHGEF10, KBTBD11, MYOM2* and *CSMD1* genes. Losses in that region have been associated with different phenotypic findings including intellectual disability and autism spectrum. The 2Mb gain at 8p12 includes the genes *NRG1* (neuregulin) and *TTI2,* both associated with cell-cell signaling in growth and development of multiple organ systems. *NRG1* is also related to the recycling and maintenance of urothelium and can be a potential biomarker of overactive bladder. The 1. 8Mb gain in 8p11. 22-p11. 23 includes four genes associated with neurogenesis: *GPR124, DDHD2, ADAM5* and *ADAM18*. The first gain at 8p22 of 1. 7Mb involved the gene *MTMR7*, a member of the myotubularin family of tyrosine/ phosphatases, that interact with *PTEN*. This interaction could explain the macrocephaly. The second gain at 8p22 of 1. 5Mb involved the *TUSC3* gene, also associated with mental disability. Exhaustive clinical evaluation and genomic characterization of the deletion and duplication size were essential for diagnosis and genetic counseling.

2355F

Breakpoint mapping by chromosome microdissection and next generation sequencing in patients with balanced chromosomal rearrangements and correlation to phenotype. *M. I. Melaragno[1], M. Moysés-Oliveira[1], M. A. P. Ramos[1], M. C. P. Cernach[1], M. Z. de Souza[1], N. Kosyakova[2], T. Liehr[2], D. Hardekopf[3], J. Stika[3], R. Plachy[3], T. Jancuskova[3].* 1) Genetics Division, Universidade Federal de Sao Paulo, Sao Paulo, Brazil; 2) Institute of Human Genetics, Jena University Hospital, Friedrich Schiller University, Jena, Germany; 3) synlab genetics s. r. o. , Laboratory synlab genetics s. r. o. , Prague, Czech Republic.

The use of next-generation sequencing (NGS) technology for breakpoint sequencing in patients with balanced chromosome abnormalities (BCA) has proven useful for identifying gene disruptions and novel candidate disease loci. We studied two new patients with balanced rearrangements: (1) a 2 year-old female patient with a pericentric inv(20)(p11. 21q12) showing mild hypotonia, wide forehead and protruding to preach in the glabella, large ears with sharply bended helices, small mouth with high and narrow palate, long and tapered fingers, anteverted nostrils, and mild micro/retrognathia; and (2) a 29 year-old woman with a balanced t(4;7)(q28. 3;p22. 1) with flat face, mild brachycephaly, telecanthus, epicanthus, high myopia, high and narrow palate, micrognathia, mild obesity, and mild intellectual disability. Chromosomal microarray showed no potentially disease-associated copy number variations. Regions around the breakpoints of the rearranged chromosomes (about 20–40 Mb) were microdissected using glass microneedles and DNA was amplified by DOP-PCR. The chromosomal origin of the microdissected segment was assessed by reverse chromosome painting (i. e. , fluorescence in situ hybridization on normal chromosome spreads). The GS Junior platform (Roche) was used for next generation sequencing. Chromosomal breakpoint positions of altered chromosomes were identified with accuracy of ~3kb, and referred according to GRCh38. The patient with t(4;7) showed the breakpoints in 4q28. 3(133,193,610bp), in an intergenic region, and in 7p22. 1(5,326,264) in the *TNRC18* gene (trinucleotide repeat containing 18). The patient with inv(20) showed breakpoints in 20p11. 21(24,813,647) and 20q12(41,844,905), both in intergenic regions. The only disrupted gene was *TNRC18* in patient 1 that codes a protein with unknown function but with a potential transcription regulatory region sequence-specific DNA binding. No phenotypic alteration was reported associated to this gene. DECIPHER shows some deletions involving *TNRC18* gene and several other genes in patients, some whom have brachycephaly and high palate. Thus, our data indicate no definitive candidate gene for the disease in the patients. We concluded that phenotype in our patients may underlie more complex pathogenic mechanisms than interruption of specific candidate genes, such as position effect flanking genes near the breakpoints, or other genetic alterations unrelated to the balanced rearrangement. Financial support: FAPESP, Brazil.

2356W

Natural history of patients with trisomy 13 receiving pediatric intensive management. *E. Nishi[1,2], M. Takasugi[3], T. Hlroma[3], M. Arakawa[1], T. Nakamura[3], Y. Fukushima[2], T. Kosho[1,2].* 1) Division of Medical Genetics, Nagano Children's Hospital, Azumino, Japan; 2) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Japan; 3) Department of Neonatology, Nagano Children's Hospital, Azumino, Japan.

Trisomy 13 (T13) is a most common autosomal trisomy syndrome, characterized by multiple congenital anomalies, severe developmental delay, and a short life span with the 1-year survival rate as 5–10 % and the median survival time as 7 days according to a large-scale population-based study. Management of neonates with T13 is controversial, palliative care or intensive care, supposedly due to the lack of precise clinical information of the natural history, especially on efficacy of treatment. To delineate natural history of T13 managed under intensive treatment, we reviewed detailed clinical data of 24 patients with full T13 admitted to Nagano Children's Hospital, providing pediatric intensive treatment to those with T13, from 1994 to 2014. Thirteen patients (54%) were found to have fetal ultrasonographic abnormalities, and five were diagnosed with T13 karyotypically. Eight patients were born by cesarean section, which was selective in one and emergent in seven. The most common indication for emergent cesarean section was fetal distress. All patients needed resuscitatation, intubation in 20, and had respiratory and cardiovascular stabilization. Mechanical ventilation was required by 21, six of whom were extubated and 12 of whom needed tracheostomy. Mean birth weight was 2273 g (range, 690 to 3758). Major complications included congenital heart defects (92%), cleft lip or cleft palate (67%), polydactyly (63%), and cryptorchidism (53% of male). Surgical operations were performed on 16 patients (67%), including tracheostomy, ileostomy, hepatic portenterostomy, lens removal for cataract, and correction of umbilical hernia, cleft lip, and polydactyly. Survival rates at age 1 week, 1 month, 1 year, 2 year, and 5 year were 100%, 96%, 46%, 25% , and 17% respectively. Median survival time for both sexes was 335. 5 days, (range, 22 to 3987). It was longer for boys (443 days) than for girls (181 days). Nine patients (38%) were discharged home. Twenty-one patients (88%) were fed enterally, including five having blender food through gastrostomy, eight using ED tube for gastroesophageal reflux, and five fed orally. In conclusion, the current study, though the sample size is small, would demonstrate much longer survival in patients with T13 having pediatric intensive management than in the cohorts described in previous population-based studies. The evidence is helpful for clinicians to offer the best information on treatment options to families of patients with T13.

2357T

Down Syndrome Resources for Clinicians and Investigators: The DS-Connect® Registry Professional Portal and an Updated NIH Research Plan. *M. Parisi[1], S. Bardhan[1], D. Jae[2], L. Kaeser[1], V. Rangel Miller[2].* 1) IDD Branch, NICHD, NIH, Bethesda, MD; 2) PatientCrossroads, San Mateo, CA.

Down syndrome (DS) is the most common genetic cause of intellectual disability. The National Institutes of Health (NIH) has supported research in DS to better understand and to develop effective treatments for those with this chromosomal disorder. As one effort to develop research resources, the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD), NIH, launched *DS-Connect®: The Down Syndrome Registry* (http://DSConnect. nih. gov) in 2013 to facilitate information sharing among persons with DS, families, and researchers. The Registry has been supported by the Down Syndrome Consortium, a public-private partnership composed of self-advocates and family members, DS researchers, leaders of professional DS organizations and foundations, and members of the Trans-NIH Down Syndrome Working Group (DSWG). As of June 2015, the registry had nearly 3000 participants. Individuals with DS and family members enter demographic and health information into the online, secure, confidential *DS-Connect®* database. Registry participants can answer surveys, see graphs of their child's height and weight, look up local health care providers, and view aggregate de-identified data from all participants. A professional portal was launched in December 2014 for clinicians and researchers with 3 levels of access depending on their needs: Level 1 users can view the de-identified data and perform detailed searches; Level 2 users can perform more complex searches and review individual-level data; and Level 3 users can post a study notice, work with a Registry Coordinator to recruit for a study, or develop a commercial application. A Research Review Committee reviews all Level 3 requests to ensure that they are scientifically sound. Examples of Registry-supported studies will be presented. A Spanish language version and mobile-device interface are planned for 2015. In addition, the DSWG has developed objectives for DS research across NIH and published *Down Syndrome Directions: The NIH Research Plan on Down Syndrome*, a 2014 update to the first plan developed in 2007. The new research plan demonstrates progress in all the research domains, but gaps remain in several longer-term objectives. The plan provides goals for the new category of "DS and Aging" in recognition of the high prevalence of Alzheimer's disease in aging individuals with DS. The updated research plan is available at [https://www.nichd. nih. gov/publications/Pages/pubs_details. aspx?pubs_id=5865].

2358F

A possible chromosome 9 new heteromorphic variant in a patient with phenotypic alterations. *A. Perez[1,2], M. E. S. Colovati[2], N. Kosyakova[3], T. Liehr[3], A. B. Hamid[3], M. I. Melaragno[2], A. C. M. Malinverni[2].* 1) Medical Genetic Center, Universidade Federal de São Paulo, São Paulo, Brazil; 2) Genetics Division, Department of Morphology and Genetics, Universidade Federal de São Paulo, São Paulo, Brazil; 3) Institut fur Humangenetik , Universitatsklinikum Jena, Jena , Germany.

Several alterations involving chromosome 9 are considered normal population variants. These variants include 9qh+, 9cenh+, 9ph+, 9phand inv(9)(p11q13). The frequency of these variants is approximately 1. 5% in the general population and they are referred to as heterochromatic variants or heteromorphisms, commonly found in routine cytogenetics. Studies with fluorescence in situ hybridization (FISH) using different probes have classified heterochromatic variants in different pattern types (Kosyakova et al, 2013). On the other hand, some studies have associated these heteromorphisms with several clinical disorders such as intellectual disability, congenital malformations and infertility. We report on a 15 year-old female patient who presented with intellectual disability delayed, language and neurodevelopment delay and facial dysmorphism. Her karyotype using GTG-banding in 20 metaphases revealed a pericentric chromosome 9 inversion: 46,XX,inv(9)(p12q13). The SurePrint G3 Human SNP+CGH Microarray, 4x180K (Agilent®) revealed a duplication of approximately 6 Mb, including 22 genes: arr[hg 19]9p13. 1p11. 2(38,869,901-44,870,714)x3. FISH was performed with pericentromeric 9 mix probes and with several chromosome 9 probes: RP11-397D12 (9p13. 2), RP11-246P17 (9p13. 1), RP11-138L21 (9p13. 1), RP11-448C16 (9p12), and RP11-138K6 (9p11. 2). An unusual rearranged chromosome was found: 46,XX,der(9)(pter→p11. 2::q21. 11→q12::p11. 2→p13. 2::q12→p11. 2::q21. 11→qter). Since this chromosome 9 pattern has not been described before, this rearrangement was considered molecular cytogenetically unbalanced. However, patient´s father´s karyotype showed a similar chromosome 9 possibly indicating a variant not described to date. Thus, chromosome 9 variants may be more complex than previously thought, and their precise characterization using a combination of different techniques (G- and C-banding, genomic array and FISH with different probes) is essential for a better knowledge of this intriguing chromosome region. Financial support: FAPESP.

2359W

Perspective of Mosaic Patau Syndrome. *J. Roy.* Amity Institute of Biotechnology, Noida, India.

Abstract: Trisomy 13 is one of the most common autosomal disorders in newborns. Popularly known as Patau Syndrome, the disorder is seen in 1 out of every 10,000 live births, out of which 5% are a result of mosaicism. Mosaicism is a condition where an extra chromosome 13 is present only in some of the person's cells. It is not inherited. With postfertilization nondisjunction, a random misdivision occurs in the cell's early mitotic stages. As a result, the infant has two different cell lines: cells with a normal complement of 46 chromosomes and cells with 47 chromosomes. This mixture of cells is referred to as mosaicism. The excess genetic material in each cell is disruptive to their course of development, leading to the characteristic features of Patau syndrome. The physical features of people with mosaic Patau syndrome are many times milder than persons with full Patau syndrome as all the cells do not have an extra copy of chromosome 13. Therefore, they have normal development in some cases with mild dysmorphic features. Few of them also turn out to have a better life expectancy than people with full Patau syndrome. Although, the phenotype and clinical outcome vary greatly in mosaic trisomy 13 often resulting in early death from major malformations. Many live born infants with mosaicism even die within the first month. Mosaic Patau syndrome causes serious physical and mental abnormalities, including heart defects, unusual facial features such as a sloping forehead, a smaller than average head (microcephaly), small or missing eyes (microphthalmia),cleft palateor extra fingers and toes (polydactyly), abnormal genitalia andmental retardation. Since it cannot be cured, genetic counselling is highly encouraged for knowing the chances of other babies being affected, as well as understanding the diagnosis options for identification at foetal stage. Genotypes and copy number information provided by Single-nucleotide polymorphism (SNP) array allow determination of parental origin and uniparental disomy status and direct quantification of mosaicism. Such information may lead to a better understanding of mechanisms of mosaicism, the observed phenotypic variability and better prediction of recurrent risk. **Key words**: Patau Syndrome, mosaicism, SNP array, uniparental disomy.

2360T

The Impact of Early Hormonal Therapy (EHT) on Anthropomorphic Measurements in 47, XXY from 4 to 96 months of age. *C. Samango-Sprouse[1, 2, 3, 4], C. Keen[2], C. Sprouse[5], T. Sadeghin[1, 2], S. Powell[6], A. Gropman[7].* 1) Neurodevelopmental Diagnostic Center, Davidsonville, MD; 2) The Focus Foundation, Crofton, MD; 3) George Washington University, Department of Pediatrics; 4) Florida International University, Department of Molecular Genetics; 5) Georgetown University, Department of Neuroscience; 6) George Washington University; 7) Children's National Medical Center, Division of Neurodevelopmental Disabilities and Neurogenetics.

Introduction: 47, XXY (Klinefelter's Syndrome) is the most commonly occurring chromosomal aneuploidy (1:650) [Chang et al. , 2015]. Though the phenotype varies substantially among those effected, it is clinically defined by small testes, hypogonadism, and neurodevelopmental dysfunction [Skakkebaek et al. , 2015]. These common features are thought to result from a decrease in testosterone levels caused by the extra X. Testosterone replacement therapies have been found to remedy many consequences associated with androgen deficiency [Samango-Sprouse et al. , 2015]. While these findings are well known, the impact of EHT on the physical growth in XXY is not well understood. It has been postulated that EHT may alter growth patterns of boys with XXY but this has not been well investigated. This study investigated the effects of EHT on head circumference (HC), height (HT), and weight (WT) in boys with XXY at 36, 72, and 96 months of age. Methods: As part of the continuing assessment and follow up at the Neurodevelopmental Diagnostic Center, anthropomorphic measurements were taken and recorded by a nurse blinded to each child's course of treatment (n=158). 59 boys were evaluated at 4-36 months; 22 had received EHT, 37 had no EHT. 68 boys were evaluated at 37-72 months; 34 had received EHT, 34 had no EHT. 31 boys were evaluated at 73-96 months; 19 had received EHT, 21 had no EHT. To maintain fidelity of all data, data was calculated and entered by someone blinded to the study. Appropriate biostatistics were completed on the anthropomorphic measurements to determine if EHT had impact between 4 and 96 months of age. Results: Findings indicate EHT does not significantly alter HC, or HT from 4-96 months in XXY boys. There was a significant difference in WT between treated and untreated boys at 73-96 months (p=. 025). Boys who had received EHT weighed 2. 3kg more than boys who received no EHT. Discussion: XXY boys have an androgen deficiency that may begin prenatally to have long-term effects on multiple areas of neurological function. To our knowledge, this is the first study demonstrating no deleterious effects of EHT on anthropomorphic measures for XXY boys from 4-96 months of age. Muscle tone improved in the EHT group, as demonstrated through the BOT-2, possibly explaining the significant difference between the two groups at 73-96 month. This suggests positive effects of EHT on selective aspects of growth and development.

2361F

Towards the identification of 9p21. 1 deletion genesis in a family with development delay. *F. J Sheth[1], J. Andrieux[2], T. Liehr[3], K. Shah[1], S. Trivedi[1], M. Desai[1], A. Shah[1], B. Patel[1], JJ. Sheth[1].* 1) Cytogenetic and Molecular Cytogenetic, FRIGE's Institute of Human Genetics, Ahmedabad, Gujarat, India; 2) Laboratory of Medical Genetics, Jeanne de Flandre Hospital CHRU de Lille, Lille Cedex, France; 3) Jena University Hospital, Friedrich Schiller University, Institute of Human Genetics, Kollegiengasse 10, 07743 Jena, Germany.

Statement of purpose Advancement in various cytogenetic diagnostic techniques has helped in detection of more cryptic segmental aneuploidies in cases with multiple congenital anomalies/dysmorphism with or without mental retardation. We report here a family; a young non-consanguineous couple reported with two spontaneous first trimester miscarriages followed by birth of two female children with unexplained mental retardation, associated dysmorphism and global developmental delay especially motor delay, absence of speech and recognition at the age of 7 and 5 years. The mother at 5thgravida was referred for further work up and advice. The family was investigated to uncover the underlying causes and remedies thereof. **Methods used**The family members (mother, father, two alive daughters) and the amniotic fluid culture cells from the gravid mother underwent conventional cytogenetics, aCGH and qPCR as a part of investigative effort. **Summary of results and discussion**Cytogenetic banding study was apparently normal [46,XX] at 550-band resolution. Oligonucleotide arrayCGH using 44K exhibited 151-283Kb deletion at 9p21. 1 [arr 9p21. 1 (28,519,569-28,670,535)x1] involving *LINGO-2 (LRRN6C)* gene. *LINGO-2 (LRRN6C)* is expressed in cortical neurons, dorsal root ganglion neurons and negatively regulates motor neuron survival and axonal length. This deletion is also reported to be associated with the expressivity of tremors and autism. Paternal inheritance was confirmed by qPCR. The deletion might be polymorphic as it was inherited from phenotypically normal father. During 5th gravida, prenatal diagnosis was carried out from amniotic fluid culture at 16 weeks showed rob(13;14)(q10;q10)pat and 151-283Kb deletion at 9p21. 1 {arr 9p21. 1(28,519,569-28,670,535)[hg19]x1} in the fetus. Looking to the clinical presentation of both the children, paternal UPD14 was suspected. Later, it was ruled out after analyzing five different STR loci. Since both the alterations were inherited from the phenotypically and mentally normal father and no rearrangements were detected in the mother; it is highly likely that the when the child inherit deletion encompassing 9p21. 1 along with homozygous deleterious mutation/s in the LINGO2 (LRRN6C) gene, may give rise to the phenotypic expression of facial dysmorphism, global developmental and motor delay, absence of speech and recognition.

2362W

A Predictive Model for Obstructive Sleep Apnea in Patients with Down Syndrome. *B. G. Skotko[1,2], E. A. Macklin[3], M. Muselli[4,5], L. Voelz[6], M. E. McDonough[1], E. Davidson[2,6], V. Allareddy[7], Y. S. N. Jayaratne[8], R. Bruun[9], N. Ching[10], G. Weintraub[11], L. Albers Prock[2,6], D. Gozal[12], D. Rosen[2,13].* 1) Down Syndrome Program, Division of Medical Genetics, Department of Pediatrics, Massachusetts General Hospital, Boston, Massachusetts; 2) Department of Pediatrics, Harvard Medical School, Boston, Massachusetts; 3) Biostatistics Center, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts; 4) Rulex, Inc. , Boston, Massachusetts; 5) Institute of Electronics, Computer, and Telecommunication Engineering, Italian National Research Council, Genoa, Italy; 6) Down Syndrome Program, Division of Developmental Medicine, Department of Medicine, Boston Children's Hospital, Boston, Massachusetts; 7) Department of Orthodontics, The University of Iowa College of Dentistry and Dental Clinics, Iowa City, Iowa; 8) Division of Orthodontics, Department of Craniofacial Sciences, University of Connecticut School of Dental Medicine, Farmington, Connecticut; 9) Department of Dentistry, Boston Children's Hospital, Boston, Massachusetts; 10) Children's Dentistry, El Cerrito, California; 11) David Geffen School of Medicine at University of California, Los Angeles, California; 12) Department of Pediatrics, The University of Chicago, Chicago, Illinois; 13) Division of Respiratory Diseases, Department of Medicine, Boston Children's Hospital, Boston, Massachusetts.

Obstructive sleep apnea (OSA) is a frequent condition in Down syndrome (DS) with prevalence ranging of 55–97% vs. 1–4% in the general pediatric population. OSA is associated with significant morbidity. Sleep studies, the gold standard for diagnosing OSA, are noninvasive, but, they are costly and often poorly tolerated by individuals with DS. Our objective was to construct a clinically useful predictive model to screen for OSA in patients with DS using a combination of signs, symptoms, craniofacial anatomy, and metabolic markers. We enrolled 130 patients with DS, ages 3–21 yrs, for an observational, prospective cohort study. Exclusion criteria included previous adenoid or tonsil removal, a recent sleep study, or treatment for OSA. This study included a physical examination/medical history, lateral cephalogram, 3-D photography, validated sleep-related questionnaires, an overnight sleep study, and night and morning urine samples. Data were used to construct a statistical predictive model using the Logic Learning Machine technique. This model predicts three levels of OSA severity based on apnea-hypopnea index (AHI) (none: AHI $\leq$ 1/hour total sleep time (hrTST); mild: 1 < AHI $\leq$ 5/hrTST; moderate to severe: AHI > 5/hrTST). The best model had a PPV by cross validation of 55% for an AHI > 1/hrTST and 25% for AHI > 5/hrTST; the cross-validated NPV were 73% and 90% respectively. In our random clinical sample, the model predicted no or mild OSA in 39% and 50% of participants, respectively. The model included survey questions, medication history, anthropometric measurements, vital signs, patient's age, and physical examination findings. Data from the study suggests that a negative finding by our model would predict with 90% accuracy that a given individual truly did not suffer from moderate to severe OSA. Positive results would need to be confirmed with a sleep study given the relatively high false positive rate. Nevertheless, using simple procedures that can be collected at minimal cost in a primary care setting, the proposed model predicts when patients with DS are unlikely to have moderate to severe OSA, obviating the need for many sleep studies. With half of patients with DS predicted to be free of moderate to severe OSA and 90% accuracy of those predictions, the model could have broad clinical application. We are currently launching a validation study to confirm the performance of the model prior to clinical implementation.

2363T

A girl with recombinant X chromosome and Turner like stigmata: report of clinical and cytogenetic findings and review of the literature. *N. Yachelevich[1], M. Regelmann[2], A. Babu[3], J. Reiner[3], S. Scott[3], L. Edelmann[3], N. Cohen[3].* 1) Clinical Genetics Services, Department of Pediatrics, New York University School of Medicine, New York, NY; 2) Division of Pediatric Endocrinology and Diabetes, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Mount Sinai Genetic Testing Laboratory, Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

The patient was a product of an IVF pregnancy with donor gametes delivered at 35 weeks. In infancy she had gastro esophageal reflux. She had recurrent urinary tract infections and was found to have horseshoe kidney and vesico-ureteral reflux which was surgically repaired. Soft tissue cervical mass has been stable in size. Cardiac evaluation for murmur showed mildly dilated ascending aorta (Z score 2. 59). Blood pressure was normal. Her height was close to the 10th percentile until her fifth birthday, after which she was noted to have a deceleration in height velocity. Her height at 9 years 5 months was -3. 19 SD, which was the reason for referral. She has ADHD treated with dexmethylphenidate, starting at age 8 years. She receives regular education. Physical exam is notable for short stature, palpable soft tissue lateral cervical mass, minimally increased carrying angle of forearms, posteriorly rotated ears, mild right eye ptosis, and mild neck webbing. Ovaries were not visible on pelvic ultrasound at age 9. 5 years. Low anti Mullerian hormone level was concerning for gonadal dysfunction. Celiac screen was negative and thyroid antibodies negative. IGF-I, thyroid function tests, sedimentation rate, complete metabolic panel and complete blood count were normal. Peripheral blood chromosome analysis revealed a recombinant chromosome X with karyotype 46,X,rec(X)dup(Xp)inv(X)(p11. 3q21. 1), which results in partial trisomy and partial monosomy of the short and long arms of the X chromosome, respectively. The banding pattern of the recombinant X chromosome suggests it was derived from an X chromosome with a pericentric inversion with likely breakpoints at Xp11. 3 and Xq21. 1. Array comparative genomic hybridization (180K CGH+SNP, Agilent Technologies, hg19) confirmed a duplication of 43. 0 Mb on chromosome Xp22. 33-p11. 11. 3 between breakpoints 169,796-43,251,684 and a deletion of 78. 8 Mb on chromosome Xq21. 1-q28 between breakpoints 76,124,793-155,208,244. Interestingly, the recombinant X chromosome resulted in total three copies of the *SHOX* gene despite her short stature. *DIAPH2* gene, which increases the likelihood of gonadal dysfunction/ovarian insufficiency was deleted. The XIST gene at the X inactivation center (XIC) on Xq13. 2 was not altered. Recombinant X chromosomes are rare and patients vary in regards to their adult stature, gonadal function and external stigmata. We review published cases. The patient's initial response to growth hormone therapy is reported.

2364F

Blood Count Alterations in Newborns with Down Syndrome: A Case-Control Study. *FJ. Martínez Macías[1], MJ. González-Cruz[3], C. Peña-Padilla[1], EL. Mellín-Sánchez[1], E. Zapata-Aldana[1], L. Bobadilla-Morales[1,2], JR. Corona-Rivera[2].* 1) Research and Registration Center of Congenital Anomalies, Clinical Genetics, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca", Guadalajara, Mexico; 2) Instituto de Genética Humana "Dr. Enrique Corona Rivera", CUCS-Universidad de Guadalajara; 3) Neonatology, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca".

Introduction: Down Syndrome (DS) is the most frequent chromosopathy in living newborns. The principal complications in DS are those that alter blood tissues, and the most frequent findings are: thrombocytosis, thrombocytopenia (TP), polycytemia, neutrophilia, transitory myeloproliferative disorder, and congenital leukemia. These alterations endanger the stability and considerably reduce life expectancy. Consequently this study was executed at Hospital Civil de Guadalajara "Juan I. Menchaca" to determine the frequency of these alterations within the population of western Mexico, and with it to create awareness of the necessity of dealing with blood count in patients with DS as soon as possible. **Objective:** To describe the principal alterations in blood counts during the hebdomadarian period in children with DS born in our Hospital. **Methods and materials:** After a clinical diagnosis (later confirmed cytogenetically) of DS, a blood count was taken during the first days of life, both of the patients and of the healthy controls. A non-coupled case control study was performed, based at the hospital, retrospective analytic at the Hospital Civil de Guadalajara "Dr. Juan I. Menchaca", Guadalajara, México during the period betweenJanuary the 1st of 2009- Decemberthe 31stof 2014. The proportion and frequency of the categorical values were analyzed, and averages and standard deviations of quantitative variables were found. Chi squared test and the fisher exact were used, the quantitative variables were compared with a *t* student test for two independent samples. The association between alterations in blood count and DSe was measured with the Odds ratio (OR). **Results:**TP was the most common hematological anomaly detected in 67/108 cases (62. 03%), followed by high CHCM 36/108, anisocytosis 38/108, monocytosis 26/108, and limphocytosis 26/108. The hematological alterations that were associated to DS, comparing with controls: TP (OR 30. 50). The severity was: mild TP (OR 9. 81) and moderate TP (OR 8. 06); anisocytosis (OR 19. 78), linfocytosis (OR 23. 84), leucemode reaction >30,000 leu (OR 10. 43), leucemoide reaction >35,000 leu (OR 17. 7) and policytemia (OR 15. 88). **Discussion:** The hematological alterations are vital to patients with DS, which is why a complete blood count is suggested during initial hospital stay. As it is taking a smear of peripheral blood in patients who require it. These actions will lead to early detection and a timely treatment of these alterations.

2365W

Hexasomy for Chromosome 15q11. 2q13. 1. *a. Tsai.* pediatrics, University of Colorado , Aurora, CO.

Duplications of chromosome 15q11-13, both interstitial and as a supernumerary marker chromosome, are common with well-known clinical features. However, higher order duplications are considerably rarer. We describe an 8-month-old girl with a hexasomy (6 copies) of 15q11. 2q13. 1 and tetrasomy (4 copies) of 15q13. 2q13. 3 on an isodicentric marker chromosome. The patient has infantile epilepsy, cortical visual impairment, dysphagia, and global developmental delay.

2366T

Aarskog-scott(saa) with cleft lip and palate in patients from operation smile colombia; clinical description. *I. Briceno[1], j. Martinez[1], A. Collins[2], A. Zarate[1], L. Arias[1], I. Briceno[3].* 1) Bioscience, U Niversidad de La Sabana, Chia, Colombia; 2) University of Southampton, UK; 3) Pontificia universidad Javeriana, Colombia.

AARSKOG-SCOTT(SAA) WITH CLEFT LIP AND PALATE IN PATIENTS FROM OPERATION SMILE COLOMBIA; CLINICAL DESCRIPTION.

Briceño, Ignacio. Martínez, Julio Cesar. Collins, Andres. Zarate, Andrés Arias Liliana, Briceño Ignacio.

Aarskog-Scott syndrome (ASS), is a rare disease characterized by short stature, hypertelorism, shawl scrotum, and brachydactyly, there can be a wide phenotypic variability including other features as neurobehavioral disorders, short nose, widow's peak, and inguinal hernia. Although the presence of cleft lip and palate (CLP) has been described in patients with ASS, it is of rare occurrence. During the clinical genetics examination of a population of 1350 patients from "Operation Smile Colombia" 25 patients have been found with ASS and CLP. After signing the informed consent, the patients were examined and blood samples for DNA extraction was obtained and molecular genetic analyses were performed. The clinical signs are described and the phenotype–genotype correlation is analysed.

2367F

Clinical Face Phenotype Space: dysmorphology photographs to aid diagnosis of genetic syndromes. *C. Nellaker[1], M. Alvi[1], Q. Ferry[1], J. Steinberg[1], C. Webber[1], D. R. FitzPatrick[2], C. P. Ponting[1], A. Zisserman[1].* 1) University of Oxford, Oxford, United Kingdom; 2) MRC Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, United Kingdom.

Clinical dysmorphologists require an enormous breadth of experience to correctly classify and diagnose ultra-rare diseases. Developments in computer vision research now enable computational phenotyping based on ordinary photographs. Computational phenotyping tools are poised to become a transformative technology in healthcare, bringing objective quantification and learned models to help clinicians diagnose, prioritise interventions, and monitor outcomes. We recently introduced Clinical Face Phenotype Space (CFPS), which locates patients in the context of known syndromes, and thus can help to generate disease hypotheses. Moreover, CFPS improved the clustering of patients by phenotype by 27 fold over random chance, even when no known syndrome diagnosis exists. This holds promise as an impartial means by which to narrow the search space for suspected rare diseases, and could augment the prioritisation of testing in clinical investigations. The CFPS algorithm automatically detects faces in photographs, annotates locations of key anatomical parts, and extracts machine readable feature descriptions of the facial gestalt. From a proof-of-principle CFPS we have further improved the face detection, feature point annotation and feature descriptors. We increase accuracies of facial phenotype representations and test how novel models of phenotype variation can aid narrowing the search space to diagnoses. More precise face detections and feature point annotations improve the fidelity of the phenotypes captured in photographs, which can be used to generate life-like representations of syndrome characteristics useful for training. Furthermore we show how the choice of feature descriptor has a large impact on CFPS modelling. Three critical criteria must be met to bring these promising technologies to clinical utility: 1. The ethical and legal framework for working with inherently identifiable medical data across international systems, 2. Training the models on all possible patient examples with molecular validation, and 3. The means for clinician researchers to extract clinically relevant data from the phenotype models. Importantly we are launching the CFPS Consortium for clinical collaborations. We have established a legal, ethical and secure framework for clinical collaborations which safeguards the interests of institutions, clinician researchers and patients. Together we will bring computational phenotyping to clinical utility.

2368W

Delayed diagnosis in turner syndrome is associated with social anxiety. *C. McDougall[1], L. Conway[1], D. Bousquet Moore[2].* 1) Arcadia University, 450 S Easton Rd, Glenside, PA 19038; 2) Turner Syndrome Foundation, PO Box 726, Holmdel, NJ 07733.

Previous studies have indicated that women and children with Turner syndrome may experience psychosocial and behavioral challenges, such as anxiety, depression and decreased self-esteem. This study surveyed women over 18 years of age with Turner syndrome to assess the prevalence of social anxiety and whether there is an association between social anxiety and either current height or age at diagnosis. A link to an online survey was posted to the Turner Syndrome Foundation's (TSF) webpage and the TSF Facebook page. One hundred twenty-one women over 18 years of age responded and eighty-six submitted a complete survey between July 2014 and June 2015. The Liebowitz Social Anxiety Scale—Self Report (LSAS-SR) was used to assess social anxiety. We found that 76% of the participants are classified as having social anxiety disorder. Twenty-six women (30%) had scores between 30 and 60 indicating a non-generalized, or subtype of, social anxiety disorder and 40 women (46%) had scores >60 indicating a diagnosis of generalized social anxiety disorder. The Mantel-Haenszel chi-square test was used to assess the correlations. There was no correlation between current height and social anxiety (p=0. 985). There was a significant association between age at diagnosis and level of social anxiety (p=0. 043), with a higher prevalence of social anxiety in women diagnosed at older ages. This study provides further evidence that social anxiety is a common psychosocial difficulty for women with Turner syndrome. Recognition of social anxiety as a feature associated with Turner syndrome may help healthcare providers give appropriate guidance and services to individuals with Turner syndrome and their families. The increased risk of social anxiety associated with delayed diagnosis provides evidence that early diagnosis may lead to improved psychosocial outcomes in individuals with Turner syndrome.

2369T

An adult with intellectual disability, ADHD, autistic featrues, and an interstitial microdeletion involving CNTN4. *E. Chow[1,2,3], P. Desarkar[3,4].* 1) Clinical Genetics Service, CAMH, Toronto, ON, Canada; 2) Campbell Family Mental Health Research Institute, CAMH, Toronto, ON, Canada; 3) Department of Psychiatry, University of Toronto, Toronto, ON, Canada; 4) Dual Diagnosis Service, CAMH, Toronto, ON, Canada.

Contactin 4 (CNTN4) is one of theaxon-associated cell adhesion molecules of the immunoglobulin superfamily that play important roles in the formationof neuronal networks. CNTN4 hasbeen considered as acandidate geneforautism and developmentaldelays, based on its functions and case reports of individuals with interstitial or terminal 3p26 deletions. Most individuals affected by 3p26 deletions reported in the literature are children. We present the case of a45-year-old womanfound to have a microdeletion involving CNTN4to add to the literature on CNTN4 and 3p26 microdeletions. The proband was referred to a psychiatric genetics clinicby her dual diagnosis psychiatrist for genetic investigations of her mild intellectual disability. She was of Chinese ethnic origin and her family history was negative for autism, intellectual disabilities, and ADHD. She was the first child of her middle-class parents and was born after an uneventful pregnancy, laborand delivery. She was of average birth weight and had no obvious birth defects. Early developmental milestones were reportedlymet, but her development slowed down after the first year. She began to haverecurrent febrile and non-febrile seizures from age 4 on. Sheshowed hyperactivity, problems with expressive language, poor impulse control, and socially inappropriateand repetitive behaviors when she started school. She wasdiagnosed withmild intellectual disability in grade school, and ADHD as a young adult. As an adult, she was treated forhypertension and hypercholesterolemia. On examination as an adult,she had short stature (148 cm) and had relative microcephaly (51. 3 cm, ~3% tile for height). Her BMI was 18. 4. She had a small uvula and a thickened left pinna, but otherwise was not dysmorphic. She did not have other typical features of3pdeletion syndrome. Results: Fragile X Syndrome and Rett Syndrome testing were negative. Clinical microarrayidentified a very smallheterozygous deletion at 3p26. 2 (3,0983,55-3,120,958) involving the proximal end of CNTN4gene. Genetic testing of her family members is being pursued. Conclusions: While it cannot be definitely concluded that CNTN4 disruption by microdeletion was casually related to this patient's presentation,itlikelycontributed to herneurodevelopmental phenotype.

2370F

Expanding the clinical spectrum for 10q26. 2q26. 3 deletion. *V. Fairchild[1], A. Shrimpton[2], C. Stein[2], K. Werner[3], R. R. Lebel[1].* 1) Section of Medical Genetics, SUNY Upstate Medical University, Syracuse, NY; 2) Department of Pathology, SUNY Upstate Medical University, Syracuse, NY; 3) Department of Neurology, SUNY Upstate Medical University, Syracuse, NY.

This female was born to a 24 year old primigravid woman by emergency cesarean route at 28 weeks following a pregnancy complicated by infections. The neonatal period was significant for low Apgar scores, poor suck, episodic bradycardia, hyperbilirubinemia, and anemia. At age 9 years 4 months her course is significant for ADHD, autism spectrum disorder, developmental coordination disorder, impulsivity, central auditory processing disorder, spastic diplegia, seizures, cerebral palsy, scoliosis, high frequency hearing loss in the right ear, and encopresis. We also noted slender body habitus, long narrow face, prominent eyebrows and eyelashes, small flat feet, and scoliosis. Family history is significant for seizures and learning disability in her maternal uncle's daughter. Microarray analysis revealed a 5. 26 million base pair deletion at 10q26. 2q26. 3 (130,163,870-135,427,143) encompassing 41 genes, 19 of which are annotated in OMIM. The 10q26 deletion syndrome as a broad entity is well established in the literature, however there have been few reports describing 10q26. 2q26. 3 deletion specifically. We believe the present case will serve to further define the phenotype.

Features	10q26 Deletion Syndrome (OMIM)	Del 10q26. 2q26. 3Case Study(Iourov et al. , 2014)	Our Patient
Developmental delay	+	+	+
Hearing loss	+	-	+
Seizures	+	-	+
Hyperactivity	+	-	+
Flat feet	-	+	+
Syndactyly	+	+	-
Prominent auricles	+	+	-
High forehead	-	+	-
Central auditory processing disorder	-	-	+
Scoliosis	-	-	+
Spasticity	-	-	+
Slender body habitus	-	-	+
Prominent eyebrows	-	-	+
Prominent eyelashes	-	-	+
Encopresis	-	-	+

I. Y. Iourov, S. G. Vorsanova, O. S. Kurinnaia, Y. B. Yurov, "An Interstitial Deletion at 10q26. 2q26. 3", *Case Reports in Genetics*, vol. 2014, Article ID 505832, 3 pages, 2014. doi:10. 1155/2014/505832.

2371W

11q13. 4 deletion disrupting SHANK2 gene in a male with ASD and moderate ID. *M. Grisolia[1], F. Scionti[1], E. Elisa[1], S. Ferraro[1], S. Sestito[1], A. Novelli[2], D. Concolino[1].* 1) Department of Pediatrics, University Magna Graecia, Catanzaro, Italy; 2) Department of Medical Genetics, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy.

SHANK2 (MIM 603290) is a member of a family of scaffold proteins that localize to the postsynaptic site of excitatory synapses in the central nervous system. Mutations and copy number variations in *SHANK2* gene have been reported in individuals with autism spectrum disorders (ASDs [MIM 209850]) and intellectual disability (ID). Here we report on a male patient arrived to our attention at age 11 years for ASD and moderate ID. The proband is the second-born child of healthy consanguineous parents. Family history was unremarkable. He was born at full term after an uneventful pregnancy with a normal neonatal period. Developmental milestones were delayed: he walked unsupported at age 2 years. He showed speech delay and used gestures associated with vocalization to indicate and formulate questions. His behaviour was friendly and sociable. The physical examination showed a mild obesity (BMI=35). No dysmorphic features were observed except synophrys. No history of other relevant health problems were reported. The same phenotype was observed in his sister. Array-CGH analysis identified an interstitial deletion of 11q13. 4 [arr 11q13. 4(70,505,060-70,588,559)x1 pat; assembly hg19] disrupting *SHANK2* gene as confirmed by qPCR. Parental chromosome analysis revealed the paternal inheritance of the deletion. Sister has not yet been analyzed. Despite the paternal origin of 11q13. 4 deletion described in our patient, previous functional studies have shown that inherited variants in *SHANK2* are associated with a statistically significant reduction in the density of synapses, although not as severe as the reduction caused by the *de novo* or truncating mutations. Genetic and functional data suggest that, although present in healthy parents, some inherited *SHANK2* mutations might contribute to the development of ASD. There are also evidences that additional genetic/epigenetic factors might be necessary to develop ASD. It's not excluded that other genes involved in the *SHANK2* synaptic pathway, such as neurexin and neuroligin genes, can contribute to the phenotype observed in our patient.

2372T

Further delineation of the phenotypic spectrum of *ZBTB18* mutations. *S. A. de Munnik[1], M. W. Elting[2], S. García-Miñaúr[3], J. Schoots[1], E. A. Sistermans[2], M. M. Weiss[2], N. V. A. M. Knoers[4], H. G. Brunner[1], E. M. H. F. Bongers[1].* 1) Department of Human Genetics, Radboud University Medical Centre, Nijmegen, Netherlands; 2) Department of Clinical Genetics, VU University Medical Centre, Amsterdam, Netherlands; 3) Department of Clinical Genetics, Hospital Universitario La Paz, Universidad Autónoma de Madrid, Madrid, Spain; 4) Department of Medical Genetics, University Medical Centre Utrecht, Utrecht, Netherlands.

ZBTB18 haploinsufficiency was proposed to contribute to some of the key features of the 1q43q44 microdeletion syndrome, encompassing intellectual disability with speech delay, microcephaly and short stature. This hypothesis was based on the first publication in 2014 of a female with a global developmental delay, short stature, microcephaly, dysmorphic facial features, and a *de novo* nonsense mutation in *ZBTB18*, a gene located in the 1q43q44 microdeletion region. Here, we present a second individual with a *de novo* missense mutation in *ZBTB18*. This 15 year old boy has a severe developmental delay, seizures in infancy, and facial dysmorphic features. These two individuals both have a severe developmental delay with marked speech delay, supporting our proposition that *ZBTB18* plays an important role in the intellectual development present in the 1q43q44 microdeletion syndrome. Both individuals lack corpus callosum abnormalities on brain MRI, whereas agenesis or hypogenesis of the corpus callosum is seen in approximately 80% of individuals with a 1q43q44 microdeletion. We hypothesize that haploinsufficiency of other genes in the 1q43q44 microdeletion region contribute to these corpus callosum abnormalities. Overlapping facial characteristics include arched eyebrows, a full nasal tip, and a pointed chin. The first individual had a short stature and microcephaly, while the second individual had a normal height (-1 SD) and borderline headcircumference (-2 SD). In conclusion, the phenotype of these individuals with *ZBTB18* mutations comprises intellectual disability with profound speech delay, and overlapping facial characteristics. To further delineate the phenotypic spectrum and variability of *ZBTB18* mutations, the collection of clinical data of additional individuals with *ZBTB18* mutations is mandatory. Furthermore, functional studies are necessary to determine the specific effects of missense and nonsense mutations on protein formation.

2373F

A Novel 17q24. 2 Microdeletion Localized to the PRKCA Gene. *B. Shayota, P. Gupta.* St Joseph's Medical Center, Paterson, NJ.

Introduction: Several chromosome 17 deletions and duplication syndromes have been described in the current literature, including the rare 17q24. 2 microdeletion. Of those 17q24. 2 microdeletions reported, all have had multiple genes affected at once, making it difficult to determine the relationship of each feature with a specific gene. One study of 17q24. 2 microdeletions reported 4 unrelated cases of various deletions and clinical features including intellectual disability, speech delay, and truncal obesity. The affected genes included some with known significance like CACNG causing seizures and PRKAR1A causing the Carney Complex, while others like PRKCA have unknown significance. This has created a complex array of phenotypic features in previously reported patients. As such, it becomes increasingly important to understand the function of each individual gene and its contribution to complex clinical presentations. Therefore, we present a case of a 17q24. 2 microdeletion localized to the PRKCA gene only. Clinical Description: The patient was referred for genetic evaluation at 2 years of age for macrocephaly and significant speech and motor delays. The patient was born at 30 weeks from an otherwise uncomplicated non-consanguineous pregnancy. The patient weighed 2. 225 kg at birth and required respiratory support only. The patient has since developed autism spectrum disorder and developmental delays. On physical examination the patient has obesity (>97%), macrocephaly (>97%), 2 hair whorls, prognanthia, and bilateral pes planus. In addition, MRI revealed delayed myelination of the periventricular white matter. Molecular Analysis: Genetic analysis was performed with SNP microarray on lymphocyte extracted DNA. The results included a 239 Kb 17q24. 2 deletion involving the PRKCA gene and a 862 Kb 2q13 duplication. Follow-up genetic testing on both parents showed that the mother, without any phenotypic abnormalities, also has the 2q13 duplication. Discussion: To the author's knowledge, no other patient with a similar 17q24. 2 microdeletion has been described before. This provides great insight to the function of PRKCA gene and phenotypic presentation when deleted. While it cannot be excluded that some of the patient's features may be due to the 2q13 duplication, it is less likely since the mother carries the same duplication without symptoms. Additionally, a previous study had shown that 1 in 876 normal controls carry the 2q13 duplication.

2374W

Primary microcephaly in a child with an increased number of prophase-like cells without MCPH1 gene mutation. *S. Lala[1], S. Ursin[1, 2], H. Chen[1, 2], L. Prouty[1, 2, 3].* 1) Department of Pediatrics, Louisiana State University - Shreveport, Shreveport, LA; 2) Department of Genetics, Louisiana State University - Shreveport, Shreveport, LA; 3) Department of Clinical Pathology, Louisiana State University - Shreveport, Shreveport, LA.

Microcephalia vera, also known as autosomal recessiveprimarymicrocephaly(MCPH), is a rare, genetically and clinically heterogeneous disorder of neurogenesis within the neurogenic epithelium, leading to a decrease in neuron number and brain size. Patients with MCPH typically exhibit microcephaly at birth (greater than three standard deviations below normal) and a degree of non-progressive mental retardation, usually without other neurological findings or malformations. This is largely due to a reduction in gray matter and subsequent brain volume, primarily affecting the cortex. There are 11 subtypes of MCPH with at least eight loci and six different genes implicated in this disorder, including autosomal recessive primary microcephaly 1 (MCPH1), abnormal spindle-like, microcephaly associated (ASPM), cyclin-dependent kinase 5 regulatory subunit-associated protein 2 (CDK5RAP2) and centromere protein J (CENPJ). A defect in the ASPM gene is the most common cause of MCPH in reported families, however MCPH1 is increasingly becoming diagnosed by finding increased numbers of "prophase-like cells" on routine cytogenetic investigation. Here, we provide, to our knowledge, the first instance in the literature of a patient with increased prophase-like cells without a pathological sequence change, deletion or duplication in the MCPH1 and ASPM genes. We conclude with a brief clinical overview of primary microcephaly.

2375T

Classical microdeletion syndromes due to chromosomal rearrangements. *S. R. Menon, D. K Das, P. M. Tamhankar.* Genetic Research Center, National Institute for Research in Reproductive Health, Mumbai, Maharashtra, India.

Classical microdeletion syndromes such as Prader Willi syndrome and Wolff Hirschhorn syndrome occur usually de novo. However, in in 5 to 10 percent cases, these could due to parental balanced translocations leading to unbalanced deletion/duplication syndromes. We report a 6 months old female child with Prader Willi syndrome due to t(X;15) and another 4 months old boy with Wolff Hirschorrn syndrome with proximal 4p duplication and distal 4pter deletion. Their clinical phenotypes and molecular findings are discussed.

2376F

Developmental Delay and Skeletal Dysplasia associated with a Maternally Inherited Deletion of 3p26. 3 with a phenotypically Normal Mother. *A. S. Angulo, E. R. Elias.* Developmental-Behavioral Pediatrics, Children's Hospital Colorado, Aurora, CO.

Introduction: Children with phenotypic abnormalities often undergo genetic testing with microarray. If a subtle abnormality is found, parents may be tested. If a parent has the same abnormality, but doesn't share the child's phenotype, the clinical significance of the genetic change becomes more difficult to interpret. **Case report:** A 30-month old female with skeletal abnormalities and unusual features presented for genetic evaluation. Her past medical history was notable for slow growth and developmental delay. On exam, she had mildly dysmorphic features including hypertelorism, low set pinnae, small pursed mouth, long philtrum, and small chin. She had global developmental delay with skills at an 18 month old level. Skeletal films were significant for short extremities, genu valgus and tibial bowing, irregular pelvis, and J-shaped sella. Family history revealed healthy parents without consanguinity. The maternal aunt and mother have both had difficulty conceiving. The mother had two miscarriages. **Genetic testing:** An 849 kb loss in 3p26. 3 (1,247,015-2,095,542) was revealed by CytoSNP-850K bead array. This loss coordinates with the CNTN6 (contactin 6) gene, which encodes a protein that is part of the NOTCH signaling pathway. This protein is involved in cell adhesion and contributes to neurogenesis and axonal formation. The deletion was confirmed by FISH studies. The mother also carries the same deletion, while the father's FISH was normal. Previous patients reported with this deletion include a child with cognitive impairment who inherited the deletion from a normal parent without mention of skeletal dysplasia. Other reports of microdeletions of the CNTN6 gene are associated with intellectual disability or autism. As the patient's features suggested Freeman-Sheldon syndrome, sequencing of MYH3 was done and was normal. **Conclusion:** The patient described in this report does not fit the phenotype of previously reported patients with 3p26. 3 deletions, other than her developmental lag. Since this patient's mother has the same deletion but neither skeletal dysplasia nor developmental delay, it is difficult to know whether this chromosomal change is clinically significant or a benign familial variant. It is also unclear whether this deletion is linked with the mother's history of infertility and fetal loss. This report of a subtle deletion in 3p26. 3 helps expand the phenotype of this rare chromosome abnormality.

2377W

Auditory phenotype of Smith-Magenis syndrome. *M. Brendal[1], K. King[1], C. Zalewski[1], B. Finucane[2], W. Introne[3], C. Brewer[1], A. CM Smith[3].* 1) Otolaryngology Branch, NIDCD/NIH , Bethesda, MD; 2) Geisinger Autism & Developmental Medicine Institute. Lewisburg, PA; 3) Office of the Clinical Director, NHGRI/NIH, Bethesda, MD.

Introduction: Smith-Magenis syndrome (SMS) is a rare disorder commonly caused by a microdeletion of chromosome region 17p11. 2, or less frequently from the mutation of *RAI1*, a gene mapped to 17p11. 2. SMS is characterized by physical anomalies, developmental delay, sleep disturbances, and a distinct behavioral phenotype. Clinical recognition typically leads to diagnosis, but genetic testing is required for confirmation. Hearing loss (HL) has been reported in the SMS population, but a comprehensive characterization of the auditory phenotype does not exist. We present a thorough examination of the auditory phenotype within a large (n=76) SMS patient population aged 1-49 years. Methods: Standard, age-appropriate hearing evaluations were analyzed for the 4-frequency (. 5,1,2,4k Hz) air-conduction pure-tone average (4F-PTA), 3-frequency (. 5,1,2k Hz) PTA for air- and bone-conduction, and tympanometry. 4F-PTA determined HL degree while 3F-PTA of air and bone conduction determined HL type. Tympanometry was used to evaluate middle ear (ME) function. Categorical and descriptive analyses of HL (degree and type) and ME function were grouped by age (0-5, 6-10, 11-20, 21-49 years) and classified according to previously published criteria. Results: Of 97 diagnostically confirmed SMS patients with audiological data, 76 (deletion: 72; *RAI1* mutation: 4) individuals (F:44; M:32), aged 1-49 years, (mean: 11. 5), provided evaluable data for cross-sectional analyses. HL occurred in 72% of ears (n=133). Mild HL most affected ears under 5 years. Hearing appeared to improve for 6-10 y/o, with more normal hearing and slight HL, before declining back to the mild range for 11-20 y/o. Ears with HL and evaluable air-bone gaps (n=54) revealed conductive HL was most endemic of 0-10 y/o. For 11-49 y/o, however, 66% of HL was sensorineural (SNHL), followed by mixed HL. Tympanometry results were reflective of the HL type, with a majority of 0-10 y/o exhibiting ME dysfunction and 11-49 y/o presenting normal ME function, typical of SNHL. Conclusion: Mild HL is most probable in the SMS population with a variable HL type. Identification of the SMS auditory phenotype is critical for effective patient management, early intervention, and patient education. Future identification of genotype-phenotype correlations may lead to a better understanding of genetic pathways involved in SMS and their influence on the auditory system.

2378T

Agenesis of the corpus callosum, developmental delay and Posterior polymorphous corneal dystrophy associated with ZEB1 gene deletion. *A. Chaudhry[1], B. Chung Hon-Yin[2], DJ. Stavropoulo[3], M. Perez Araya[4], E. Heon[4], D. Chitayat[1].* 1) Division of Clinical and Metabolic Genetics. , The Hospital for Sick Children. University of Toronto. , Toronto, Ontario, Canada; 2) Department of Paediatric and Adolescent Medicine, University of Hong Kong; 3) Department of Paediatric Laboratory Medicine, Division of Genome Diagnostics, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; 4) Department of Pediatrics, Division of Ophthalmology and Vision Sciences, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada.

Agenesis of corpus callosum (ACC) is one of the most common brain abnormalities with a prevalence of 0. 5 per 10,000 in the general population and 230-600 per 10,000 in children with neurodevelopmental disability. Etiologically, the condition is heterogeneous and in a Californian birth prevalence study, chromosomal abnormalities were detected in 17. 8% of the children with ACC. We report on a girl diagnosed antenatally with ACC on fetal ultrasound and MRI. On postnatal follow-up she was noted to have developmental delay, facial dysmorphism, autism spectrum disorder and posterior polymorphous corneal dystrophy. There was no keratoconus. Array-CGH showed a 2. 05 Mb de novo interstitial deletion in chromosome 10p11. 23p11. 22. The deleted region overlaps 1 OMIM Morbid Map gene, ZEB1 (the zinc finger E-box binding homeobox transcription factor 1), in which mutations are found to cause posterior polymorphous corneal dystrophy (PPCD). Recently Jang *et al.* (2014) reported 2 unrelated individuals with agenesis/hypoplasia of corpus callosum as well as posterior polymorphous corneal dystrophy (PPCD3). Both of these individuals were found to have frameshift mutations in ZEB1 gene. We provide the first phenotype correlation to a whole gene deletion. Gene expression arrays have shown that *ZEB1* is expressed in many tissues in vertebrates including brain. ZEB1-null mice showed mesenchymal-epithelial transition, which is associated with diminished proliferation of progenitor cells in the central nervous system. Haplo-insufficiency of *ZEB1* is likely related to the phenotype of our patient. Further studies are needed to elucidate the role of *ZEB1* our patient's phenotypic findings.

2379F

Behavioral features within adaptive behavior profiles: VABS Subdomain profiles in subtelomeric disorders. *G. Fisch[1], A. Battaglia[2], J. Carey[3], R. Falk[4], R. Simensen[5], J. Youngblom[6].* 1) Statistics and CIS, CUNY / Baruch College, New York, NY; 2) Stella Maris Clinical Research, Calambrone, Italy; 3) University of Utah, Salt Lake City, UT; 4) Cedars Sinai Hospital, LA, CA; 5) Greenwood Genetic Center, Greenwood, SC; 6) CSU, Turlock, CA.

Background: Previously, we found significant differences in both cognitive- and adaptive behavior profiles in children with subtelomeric deletions: 2q37, 8p21invdupdel, Wolf-Hirschhorn Syndrome [WHS] and Jacobsen Syndrome [JBS]. Thus the aim of this study was to determine if there were significant profile differences among these subtelomeric disorders within the subdomains of the Vineland Adaptive Behavior Scale: Communication, Daily Living Skills [DLS], and Socialization. **Method**: Forty-nine children diagnosed with WHS [n=22], JBS [n=10], 2q37 [n=9] or 8p21invdupdel [n=8] most of whom were recruited previously were assessed with the VABS. Ages at testing ranged 4 – 20 years. M:F ratio was ~ 1:1 (25:24). Children were assessed for cognitive ability [IQ] with the SBFE and adaptive behavior with the VABS [Domains and Subdomains]. **Results:** As noted previously, mean IQs differ significantly from among the several subtelomeric deletion disorders. To examine Domain profiles, raw scores were converted into Age-Equivalent [AE] years. Despite their lower raw score numbers, AEs in the Written subdomain were markedly higher than Receptive or Expressive subdomain AEs in the Communication Domain. Community subdomain AEs from the DLS were somewhat higher than Personal or Domestic subdomains, but subdomain AEs in the Socialization Domain did not differ noticeably from one another. A MANOVA of the 3 Domain AEs as a function of Genetic Disorder and Gender showed significant differences in AEs for Communication and DLS as a function of Genetic Disorder but not Gender. Neither predictor variable had a significant effect on Socialization AE. Not surprisingly, given their respective IQ scores, highest AEs were recorded among children with JBS; lowest among children with WHS (Scheffe corrected $p<0.05$). To correct for the effect of IQ, AEs for the 4 genotypes were analyzed using an ANCOVA model with IQ as covariate. As expected, IQ contributed significantly to AEs in all 3 adaptive behavior Domains ($p<0.03$). Interestingly, when AE was adjusted for by IQ, genotype did not contribute significantly to AE in any adaptive behavior domain. **Conclusion:** Although it has been well documented that IQ is strongly correlated DQ, the extent to which IQ accounts for various activities associated with adaptive behavior in genetic disorders that produce ID should probably be reconsidered.

2380W

Microdeletion at Xq13. 1 causing loss-of-function of CITED1 gene in a male newborn with multicystic dysplastic kidney disease (MCDK). *L. Mao[1], Y. Li[1,2], I. Yosypiv[2], S. Rito[2], A. Janssen[1], E. Morava[1,2], T. -J. Chen[1,2].* 1) Hayward Genetics Center, Tulane University School of Medicine, New Orleans, LA; 2) Dept. of Pediatrics, Tulane University School of Medicine, New Orleans, LA.

Genetic factors have been previously demonstrated to play a role in the etiology of multicystic dysplastic kidney disease (MCDK). However, the etiology of the majority of MCDK cases remains unknown. Here we report a case of MCDK diagnosed prenatally in a male fetus with right-sided multicystic kidney at the 24-weeks gestation by ultrasound. The infant was delivered via Cesarean-section at 35 weeks due to fetal distress. Postnatal renal ultrasound confirmed the diagnosis of right-sided MCDK and revealed mild hydronephrosis in the left kidney with no cysts. MAG3 renal scan demonstrated no function of the right kidney and normal function of left kidney. No facial dysmorphology, cardiac or hepatic defect was noted. At 40-day postnatal checkup, the patient showed normal growth and development. Biochemical profile indicated that he had mildly elevated of potassium, glucose and blood urea nitrogen, and was normal on sodium, creatine and AST/ALT levels. A 79 Kb microdeletion at Xq13. 1was identified by arrayCGH. This alteration results in full deletion of four genes (PIN4, RPS4X, CITED1 and HDAC8) in this patient. PIN4 encodes a peptidyl-prolyl cis/trans isomerase and has been suggested to play a role in the cell cycle. RPS4X encodes for the ribosomal protein S4, a component of the 40S subunit. HDAC8 belongs to the class I of the histone deacetylase family that is involved in DNA methylation and transcriptional regulation. These three genes may play a role in kidney development although there is no direct evidence so far. It is worth to note that expression of CITED1 is a well-known marker for renal progenitor cells, and the function of CITED1 has been previously implicated in fetal kidney development. In CITED1 mutant mice, renal medullary dysplasia was found to be a direct consequence of placental insufficiency resulting from the loss function of CITED1. Mutations in CITED1 have never been reported in humans, and have not been associated with any clinical phenotype, including MCDK. This is the first case to link the function of CITED1 and other gene(s) within the Xq13. 1 to MCDK. Thus, further investigation on the functions of these genes should be conducted for better understanding of the molecular etiology of MCDK.

2381T

Increased Prevalence of Malignancy in Twins with 22q11. 2 Deletion Syndrome. *A. R. Schott[1], K. D. Valverde[1], J. M. Kalish[2,3], M. P. Lambert[2,3], J. J. D. Morrissette[2,3], B. S. Emanuel[2,3], E. H. Zackai[2,3], D. M. McDonald-McGinn[2,3].* 1) Arcadia University, Philadelphia, PA; 2) Children's Hospital of Philadelphia, Philadelphia, PA; 3) The Perelman School of Medicine of the University of Pennsylvania, Philadelphia, PA.

Background: Various malignancies have been reported in association with the 22q11. 2 microdeletion syndrome (22q11. 2DS), albeit rarely, with hepatoblastoma being the most common. Associated immunodeficiency has been postulated as a possible etiology. Previous studies report that the incidence of malignancy is the same for both singleton and twin pairs but epigenetic discordance has been reported within the twin population. Both the *DGCR8* and *COMT* genes found within the deleted region of chromosome 22 have been identified as areas with DNA methylation differences in twins. A better-defined risk of malignancy within this deletion syndrome allows for more appropriate patient management and potential implementation of tumor screening. **Methods:** Records from 1232 patients with a 22q11. 2DS evaluated at the 22q and You Center at the Children's Hospital of Philadelphia (CHOP) were reviewed for evidence of malignancy. This included oncologic pathology reports. **Results:** Four cases of malignancy not previously reported as case studies in the literature were identified including: a Caucasian 29 year old with cancer of the mouth; a Caucasian 10 year old with papillary thyroid carcinoma; an African American 9 year old with labial melanoma; and a Caucasian 18 year old with B-cell acute lymphoblastic leukemia. Of note, 3 out of these 4 patients were members of a twin pair. In addition, we previously reported hepatoblastoma in a twin pair seen at CHOP. The CHOP cohort includes 21 pairs of dizygotic and 7 pairs of monozygotic twins, to date three different members of monozygotic twin sets and one member of a dizygotic twin set have been affected by malignancy. **Discussion:** Based on these cases, the prevalence of malignancy in twins in our cohort was 14% (4/28), making it 15 times higher than the prevalence of malignancy found in twin studies in the general population. Moreover, the overall prevalence of malignancy within our cohort, inclusive of our previous reported cases, was 0. 6%, which is greater than the 0. 0867% (age 0-9) and 0. 187% (age 10-19) reported 2011 prevalence percentage SEER data. These findings support further study directed at both epigenetic differences and gene functions that could lead to the increase risk for malignancy in the 22q11. 2DS patient population, specifically within the subset of affected twins.

2382F

Neonatal hemolytic anemia and multiple congenital anomalies due to deletions of 5q33, 6q24, and 14q23. *Y. Watanabe[1,2], M. Unno[1], M. Mitsuo[1], S. Nakagawa[1], S. Ozono[1], K. Ueda[1], Y. Yoshino[1], N. Harada[1], K. Fukui[1], S. Yano[3].* 1) Dept Pediatrics and Child Health, Kurume Univ, Kurume, Fukuoka, Japan; 2) Research Institute of GC/MS, Kurume Univ, Kurume, Fukuoka, Japan; 3) Genetics Division, Dept Pediatrics, LAC+USC Medical Center, Univ of Southern California, Los Angeles, CA, USA.

Background :Chromosomal microdeletions have been associated with many syndromes with variable clinical findings, and make up a fraction of copy-number variants (CNVs). Recurrent CNVs generally arise by nonallelic homologous recombination (NAHR) during meiosis. NAHR mediated rearrangement often involves local architectures, i. e. , segmental duplications or low-copy repeats. Many disease associated NAHR hot spots have been reported including chromosome 5q and 14q. **Case Report**:A 3-month-old girl was hospitalized due to failure to thrive and chronic anemia. She was a product of a healthy Japanese mother and a non-consanguineous healthy father. The pregnancy history was unremarkable. She was a term baby: birth weight 2135 g (SFD -1. 3SD) and the Apgar scores were 71/85. She was noted to have bilateral cleft lip/palate, patent ductus arteriosus, bilateral hearing loss, unilateral branchial fistula and sacral dimple. Jaundice and microcytic hemolytic anemia was diagnosed at 1 day of age. Chromosome studies showed the abnormal karyotype with 46,XX,del(5)(q33. 1q33. 3). Since the deletion of 5q31-q33, which is close to the deleted region, is known to be associated with myelodysplastic syndrome, bone marrow biopsy was performed at 3 months of age. It demonstrated mild hypocellularity with decreased erythroid colonies, suggestive of pure red cell aplasia. A SNP microarray study detected multiple chromosomal deletions at 5q33. 2q34 (6. 1 Mb), 6q24. 1 (423Kb), and 14q23. 1q23. 3 (5. 8Mb**). Discussion**: The deleted region of chromosome 5q was distal to the ribosomal protein S14 gene which is defective in a myelodysplastic syndrome. The deleted regions of the chromosome 5q33. 2-q34 and 14q23. 1-q23. 3 involve the following genes: *SGCD, HAVCR1, ITK, NIPAL4, and IL12B* for 5q *and SIX6, SIX1, PRKCH, SYNE2, MTHFD, SPTB, and MAX* for 14q. There are no known genes responsible for human disorders in the deleted region of chromosome 6q. A deletion of *SPTB* is known to cause spherocytosis and hemolytic anemia in infant. *SIX1* mutations are reported in Branchio-Oto-Renal syndrome. The other genes on the deleted regions are also likely contributing her clinical findings with multiple anomalies. Parental chromosome studies including microarray studies will be performed to investigate the cause of multiple chromosomal involvement. .

2383W

Prenatal Detection By Array CGH Of *GPC3* Intragenic Duplications Causing Simpson-Golabi-Behmel Type 1: Study Of Two Cases And Literature Review. *C. Maftei[1,2], M-A. Delrue[2], V. Desilets[2], A. Fortier[1], S. Wavrant[3], F. Rypens[4], G. Mathonnet[1], S. Nizard[1,2], F. Tihy[1], G. Maire[1], E. Lemyre[1,2].* 1) Cytogenetic Laboratory, CHU Ste-Justine, Montreal, Quebec, Canada; 2) Medical Genetics Division, CHU Ste-Justine, Montréal, Canada; 3) Obstetric Department, CHU Ste-Justine, Montréal, Canada; 4) Imaging Department, CHU Ste-Justine, Montréal, Canada.

ObjectiveWe report the first two patients diagnosed by prenatal array CGH with Simpson-Golabi-Behmel type 1 (SGBS1) syndrome caused by intragenic duplications in the *GPC3* gene, and we review the literature. **Case report**The first patient, a male foetus, was investigated at 22 weeks of gestation for foetal sexual ambiguity and ventriculomegaly. Array CGH analysis revealed a duplication encompassing exon 3 to 7 of *GPC3*. The 24 weeks ultrasound showed macrosomia, nephromegaly and hypogenitalism, which were confirmed at autopsy. The second patient presented with polyhydramnios, macrosomia, pyelectasy and suspicion of polydactily, at 28 weeks of gestation. Array CGH analysis revealed a *GPC3* duplication of exon 7 in this male foetus. The postnatal evaluation found facial dysmorphia, diastasis recti, nephromegaly and vertebral anomalies. Maternal transmission was confirmed in both cases by array CGH. **Literature review**Only two intragenic duplications in the *GPC3* gene are reported in the literature: exons 2 to 4 in one patient, and exon 2 in two siblings. These patients were diagnosed in the postnatal period by MLPA, having a classical SGBS1 phenotype. Our patients are the first to be identified in the prenatal period, and by array CGH. The pathogenicity of these duplications is supported by the clinical presentation. **Conclusion**Our cases expand the mutation spectrum associated with SGBS1 and illustrate the point that array CGH interpretation of intragenic duplications can be difficult, especially in the prenatal period.

2384T

5q12. 1 Deletion – First Report of Transmission and Adult Phenotype. *C. Bupp, A. Thompson, T. Drumheller.* Spectrum Health, Grand Rapids, MI.

Microdeletion at 5q12. 1 is a rarely reported finding with common phenotypic features of intellectual disability and ocular defects. We present two additional cases of a father and son with 5q12. 1 deletion which represent the first examples of transmission of the deletion and of the adult evolution of the phenotype. Patient 1 is a 6-year-old male with developmental delay, mild intellectual disability, hypotonia, and ocular findings of esotropia, amblyopia, and hyperopic astigmatism. Other notable findings include cryptorchidism and ventricular septal defect. Patient 2 is his father, a 43-year-old male who had mild developmental delay, has difficulty with anxiety, agorophobia, short-term memory, and also has restless leg syndrome. He completed 12 years of schooling and is actively employed. Glasses are worn for hypermetropia. Patient 2 has three additional female children, all with learning difficulties, but none have had genetic testing. Both patients presented have migraine headaches beginning in childhood. Facial features are not consistent between the two or in close approximation to features previously described in the literature (broad forehead, thin upper lip, coarseness). Deletions are sized at 4. 12Mb at 5q12. 1q12. 3 (hg18 – chr5:60945135-65063168; hg19 - 60909378-65027412) which is intermediate in size to other reports (0. 98-17. 3Mb). In our review, *C5orf64* is identified as a potential candidate gene for eye movement and vision abnormalities as it is deleted in all cases of deletion that had ocular defects as a feature. Previously, the adjacent gene *KIF2A* has been suggested as the phenotypic candidate. When factoring in other cases with larger deletions, we identified a deletion that includes *C5orf64,* but not *KIF2A,* in which there are ocular defects. This suggests *C5orf64* may be more responsible for ocular findings than *KIF2A*. These reported clinical findings show some agreement with previously published phenotype, especially in developmental delay and ocular defects, yet suggest inconsistency in facial morphology. This first adult case appears to have the most mild presentation of all those reported with 5q12. 1 deletion which may be due to ascertainment bias; it should be used cautiously in predicting future outcomes. Candidate genes for the ocular findings commonly seen in 15q12. 1 deletion should now include *C5orf64*.

2385F

Developmental Delay and Characteristic Facial Features in a Novel 7p22. 3p22. 2 Microdeletion Syndrome. *A. C. Yu[1], R. M. Zambrano[2], I. Cristian[3], C. Armour[1,4].* 1) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Division of Clinical Genetics, Department of Pediatrics, Louisiana State University Health Science Center, New Orleans, Louisiana, USA; 3) Nemours Children's Hospital Orlando, Orlando, Florida, USA; 4) Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada.

The clinical use of genomic microarrays in the evaluation of patients with developmental delay has led to the discovery of many novel microdeletion and microduplication syndromes. However, there are still a number of areas in the genome where the genotype-phenotype correlation is not well described. A search of the literature reveals that there are very few reported cases of patients with a deletion in 7p22. 3p22. 2 and only one case in which the deletion was isolated. Here we describe 3 female patients with overlapping *de novo* microdeletions of 7p22. 3p22. 2. The clinical features observed in these patients included characteristic facial features, variable developmental delays, and minor brain MRI findings. Although the deletions varied in size, there was a 0. 47 Mb common region of overlap which contained the genes: *EIP3B, CHST12, LFNG, BRAT1, TTYH3, AMZ1,* and *GNA12*. We propose that these patients represent a novel and distinctive microdeletion syndrome characterized by distinctive facial features and variable developmental delay.

2386W

A case of 2q33. 1 microdeletion syndrome resulting from a *de novo* reciprocal translocation t(2;11)(q32;q13). *F. Falvo[1], F. Scionti[1], I. Mascaro[1], A. Nicoletti[1], F. Ceravolo[1], A. Novelli[2], D. Concolino[1].* 1) Department of Pediatrics, University Magna Graecia, Catanzaro, Italy; 2) Department of Medical Genetics, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy.

Interstitial deletions of 2q involving the 2q31q33 region are associated to 2q33. 1 microdeletion syndrome [MIM 612313]. Clinical features include severe mental retardation, developmental delay, dysmorphic features, failure to thrive, thin and sparse hair, cleft or high palate and behavioral disturbances. Haploinsufficiency of one gene within the deleted region, *SATB2* [MIM 608148], has been suggested to be responsible for most of the features of the syndrome. We describe a male patient with a *de novo* deletion spanning the 2q32. 2q33. 3 region due to reciprocal translocation t(2;11)(q32;q13), whose phenotype is consistentwith 2q33. 1 microdeletion syndrome. Clinical manifestations included intellectual disability, speech delay, short stature, microcephaly, cryptorchidism, anterior muscular ventricular septa defect and an aggressive behaviour. Facial dysmorphisms included long face, frontal bossing, downslanting palpebral fissures, low-set ears, blue sclerae, cleft palade, micrognathia. Brain MRI did not reveal any abnormality. Conventional karyotype detected a reciprocal translocation between chromosome 2 and 11 involving respectively 2q32 region and 11q13 region. Parental karyotypes resulted normal. Further characterization by array-CGH analysis revealed a 15 Mb deletion on chromosome 2 [arr 2q32. 2q33. 3(190,458,579-205,973,945)x1 dn; assembly hg19] as confirmed by FISH. These genetic data demonstrate the importance and diagnostic complementarity between karyotype and array-CGH for a complete molecular characterization of non-pure rearrangements. Clinical and molecular findings are consistentwith 2q33. 1 microdeletion syndrome and support the hypothesis that the 11q breakpoint don't contribute to pathological phenotype observed in our patient.

2387T

Inherited intrachromosomal Xq28 deletion involving *MECP2* and Xp22. 33 duplication in a patient with Rett-like phenotype. *A. Nicoletti[1], F. Scionti[1], S. Sestito[1], E. Pascale[1], F. Falvo[1], M. Grisolia[1], A. Novelli[2], D. Concolino[1].* 1) Department of Pediatrics, University Magna Graecia, Catanzaro, Italy; 2) Department of Medical Genetics, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy.

Xq28 deletion encompassing the *MECP2* [MIM 300005] gene were described in affected femals with Rett-like phenotype characterized by idiopathic intellectual disability, autism, epilepsy, late regression age, intact ability to walk and mild dyspraxia of hand movement, congenital malformations and facial dysmorphisms. Genomic rearrangements involving terminal deletion of Xq28 and terminal duplication of Xp22. 33 are rarely reported in literature. These cases are presumed to be the result of either a recombination event in a parent carrying a pericentric inversion or inherited from a phenotypically normal mother carrying the abnormal X. We report on 5-years-old female carrying an intrachromosomal Xq28 deletion encompassing the *MECP2* gene and an Xp22. 33 duplication. Her phenotype included short stature, speech and developmental delay, hypotonia, cerebellar vermis hypoplasia, thinning of corpus callosum, seizures, bilateral hypoacusis, patent ductus arteriosus, areas of hypopigmentation in legs and skeletal malformations (scoliosis, clinodactyly, syndactyly). Facial features were hypertelorism, epicanthus, micrognathia, low-set ears and cleft palate. Conventional karyotype revealed a partial deletion of the long arm of X chromosome. Further array-CGH analysis confirmed a deletion of Xq28 and also identified a duplication of Xp22. 33, spanning respectively 6 Mb and 2. 4 Mb [arr Xq28(149,116,213-155,190,024)x1 mat, arr Xp22. 33(61,091-2,651,980)x3 mat; assembly hg19]. Maternal chromosome analysis, performed with FISH, detected the same abnormal X chromosome. X-chromosome inactivation (XCI) analysis, performed on peripheral blood lymphocytes, revealed a skewed pattern of XCI in the patient, along with the findings of an active derivative X chromosome and an inactive normal X, while the maternal pattern of XCI resulted random. Clinical manifestations in our case are consistentwith RTT subtype. The proband also exhibited short stature and several skeletal anomalies that we suggest are probably due to the Xp22. 33 duplication including *SHOX* [MIM 213865] gene. However X-chromosome inactivation studies in other tissues may provide a better explanation for the difference in phenotypes and X-inactivation patterns between proband and her mother.

2388F

An interstitial 2q22. 3q23. 2 deletion not involving *ZEB2* gene in a male with Mowat-Wilson syndrome-like phenotype. *F. Scionti[1], A. Novelli[2], R. Apa[1], M. T. Moricca[1], G. Bonapace[1], V. Salpietro[3], P. Strisciuglio[4], D. Concolino[1].* 1) Department of Pediatrics, University Magna Graecia, Catanzaro, Italy; 2) Department of Medical Genetics, Bambino Gesù Children's Hospital, IRCCS, Rome, Italy; 3) Department of Pediatrics, University of Messina, Messina, Italy; 4) Department of Pediatrics, University Federico II, Naples, Italy.

Mowat-Wilson syndrome (MWS [MIM 235730]) is a rare genetic disorder caused by heterozygous mutations or deletions of the *ZEB2* [MIM 605802] gene. The major symptoms include characteristic facial features, usually severe intellectual disability, epilepsy, Hirschsprung disease (HSCR), and multiple congenital malformations, including genital anomalies, congenital heart defects, agenesis of the corpus callosum, and eye defects. We report on 36-years-old male patient with an interstitial deletion of the long arm of chromosome 2 involving the 2q22. 3q23. 2 region but not *ZEB2* gene and with a phenotype consistent with MWS. The proband exhibited postnatal growth retardation, severe developmental delay, hypotonia, speech delay, seizures and behavioral problems. The physical examination showed microcephaly, frontal bossing, large and deep-set eyes, hypertelorism, broad nasal bridge, open mouth, with M-shaped upper lip, frequent smiling, prominent pointed chin and low-set ears. Additional clinical manifestations included constipation and a severe immunoglobulins deficiency. Array-CGH analysis revealed a 4. 6 Mb deletion on chromosome 2 [arr 2q22. 3q23. 2(145,664,976-150,357,530) x1; assembly hg19] as confirmed by FISH. Maternal chromosome analysis resulted normal. Father was not investigated. The deletion detected in our patient doesn't involve *ZEB2* gene but his phenotype is suggestive of MWS. However several studies have identified deletions that encompass enhancer elements, resulting in loss or gain of expression of nearby genes. Upstream of the *ZEB2* promoter is reported a regulatory region, at positions chr2:146,509,221–146,510,330 (hg19) that results deleted in our patient, so we hypothesize a positional effect of 2q22. 3 deletion. Interestingly, our patient showes immunoglobulins deficiency that have not been reported previously and is possibly an additional feature of this syndrome.

2389W

Chromosomal microarray testing identifies trichorhinophalangeal syndrome type I with cognitive developmental delay. *J. Gilman[1], C. Stein[2], K. Palomino[3], RR. Lebel[1].* 1) Section of Medical Genetics, SUNY Upstate Medical University, Syracuse, NY; 2) Department of Clinical Pathology-Cytogenetics, SUNY Upstate Medical University, Syracuse, NY; 3) Department of Orthopedic Surgery, SUNY Upstate Medical University, Syracuse, NY.

This male was born to a 20 year old G4P3>4 woman with an uncomplicated pregnancy and delivery. There were no suspected teratogen exposures and the union is not known to be consanguineous. He had significant developmental delays, walking at 3 years and speaking his first words at 4 years, but he made significant improvements with the aid of special education. Radiographs and MRI demonstrated avascular necrosis of the left femoral head, but no exostoses. At age 10, the patient was diagnosed with Osteochondrosis/Legg-Calve-Perthes disease of the left hip. He underwent left hip arthrodesis and spica casting surgery at age 13. He rapidly gained 120 pounds while immobile during post-recovery, and remained hyperphagic thereafter. The patient presents with short stature (1st percentile for height), obesity (96th percentile for weight), large tongue with prominent taste buds, high arched palate, hyperextensible joints, mild scoliosis, and a bulbous nasal tip. He has short hands and feet, brachydactyly, hypoplastic toenails and a small penis. The patient's father was reported to also be affected by short stature, early baldness, irregularly short digits, and a bulbous nasal tip.

Microarray analysis identified a 198 kbp deletion at 14q32. 11 and a 3. 08 Mbp interstitial deletion at 8q23. 3. The deletion at 8q23. 3 includes only two annotated genes. *TRPS1*, which codes for a zinc-finger protein that acts as a transcriptional repressor of GATA-regulated genes. The *TRPS1* gene is associated with trichorhinophalangeal syndrome type I (TRP1), which is inherited as a dominant trait, supported in this patient by the father's history. The lack of exostoses and the location of the deletion both indicate type I as opposed to type II, although intellectual disability and hyperextensible joints are typically features of type II. The 8q23. 3 deletion also includes the *CSMD3* gene, which is associated with changes in brain activity, but loss of one copy is of uncertain significance. The deletion at 14q32. 11 includes only the genes *TDP1* and *EFCAB1.* Loss of one copy of each is considered a variant of uncertain significance. The current literature includes three reports of TRP1 with intellectual disability; all the authors conclude that the size of the deletion is proportional to the degree of intellectual disability. We propose that our patient's 3. 08 Mbp deletion 8q23. 3 is the etiology for TRP1 with intellectual disability and hyperextensible joints in him.

2390T

Deletion of KIRREL3 causes intellectual disability in Jacobsen syndrome. *C. Hatano[1], T. Yokoi[1], Y. Enomoto[1], T. Saito[2], J. Nagai[2], K. Kurosawa[1].* 1) Genetiscs, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Department of Clinical Laboratory, Kanagawa Children's Medical Center, Yokohama, Japan.

Jacobsen syndrome is a rare contiguous gene disorder associated with multiple congenital anomalies. The major symptoms include growth retardation, developmental delay, congenital heart defects, and thrombocytopenia. The degree of intellectual disability in the patients varies widely. We report on a female patient of Jacobsen syndrome caused by a deletion of terminal 6. 5 Mb in 11q24. 3q25 representing normal development. She was born to nonconsanguineous parents at 38 weeks of gestation. Her birth weight was 2660g (-0. 7SD), length 47cm (-0. 8SD), and the head circumference 32. 5cm (-0. 5SD). She had coarctation of the aorta and was repaired at 2 months old. At 3 years old, her weight was 13. 5kg (-0. 9SD), length 81. 7cm (-3. 0SD), head circumference 47. 6cm (-0. 6SD). She also had mild thrombocytopenia. She was able to walk with support at 13 months old and without support at 15 months old. Her developmental milestones were normal. She spoke at 13 months old and two-word sentence at 18 months old. She goes to kindergarten without any support. The standard karyotyping revealed normal, 46,XX. We performed cytogenetic microarray analysis, which revealed a 6. 5Mb terminal deletion at 11q24. 3q25, extending from position128, 350,397-134,868,407 (hg19). The genotype-phenotype correlation has been established on the distinctive features in Jacobsen syndrome. The deletion of JAM3 and ETS1 cause cardiovascular complication, and the deletion of FLI1 and NFRKB cause thrombocytopenia, respectively. Although she had 6. 5 Mb deletion encompassing several genes, likely responsible for intellectual disability in Jacobsen syndrome, her development milestones are within normal range. KIRREL3, associated with neural development, is retained in her der(11), but deleted in most of the previously reported patients with severe intellectual disability. These results indicate that KIRREL3 is a candidate gene responsible for intellectual disability in Jacobsen syndrome.

2391F

Submicroscopic Deletions at 13q32. 1Cause Congenital Microcoria. *L. Fares Taie[1], S. Gerber[1], A. Tawara[2], A. Ramirez-Miranda[3], JY. Douet[4], H. Verdin[5], A. Guilloux[1], JC. Zenteno[6], H. Kondo[2], H. Moisset[1], B. Passet[7], K. Yamamoto[8], M. Iwai[9], T. Tanaka[10], Y. Nakamura[11], W. Kimura[12], Ch. Bole-Feysot[13], M. Vilotte[7], O. Odent[14], JL. Vilotte[7], A. Munnich[1], A. Regnier[4], N. Chassaing[15], E. De Baere[5], I. Raymond-Letron[4], J. Kaplan[1], P. Calvas[15], O. Roche[16], JM. Rozet[1].* 1) Imagine - Institute of Genetic Diseases, Paris Descartes University, 75015 Paris, France; 2) University of Occupational & Environmental Health, Kitakyushu, Japan; 3) Instituto de Oftalmologia "Conde de Valenciana". UNAM, Mexico City, Mexico; 4) Veterinary School of Toulouse, University of Toulouse, France; 5) Center for Medical Genetics, Ghent University, Belgium; 6) Faculty of Medicine, UNAM, Mexico City, Mexico; 7) Institut Nationale de la Recherche Agronomique, Jouy-en-Josas, France; 8) Institute of Bioregulation, Kyushu University, Fukuoka, Japan; 9) Ehime University Graduate School of Medicine, Japan; 10) RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan; 11) University of Chicago, USA; 12) Kimura Eye Clinic, Kure, Japan; 13) Genomics Plateform, IMAGINE Foundation and Paris Descartes University, 75015 Paris, France; 14) CHU Hôpital Sud, Rennes, France; 15) Hôpital Purpan, Toulouse, France; 16) IHU Necker-Enfants Malades, University Paris-Descartes, 75015 Paris, France.

Congenital microcoria (MCOR) is a rare autosomal dominant disorder characterized by inability of the iris to dilate owing to absence of dilator pupillae muscle. So far, a dozen MCOR families are reported worldwide. By using whole-genome oligonucleotide array CGH, we have identified deletions at 13q32. 1 segregating with MCOR in six families originating from France, Japan and Mexico. Breakpoint sequence analyses showed nonrecurrent deletions in 5/6 families. The deletions varied from 35 Kbp to 80 Kbp in size, but invariably encompassed or interrupted only two genes: *TGDS* **encoding the TDP-glucose 4,6-dehydratase and** *GPR180* **encoding the G protein-coupled receptor 180, also known as intimal thickness-related receptor (ITR). Unlike TGDS which has no known function in muscle cells, GPR180 is involved in the regulation of smooth muscle cell growth. The identification of a null** *GPR180* **mutation segregating over two generations with iridocorneal angle dysgenesis which can be regarded as a MCOR endophenotype is consistent with the view that deletions of this gene, with or without the loss of elements regulating the expression of neighboring genes, are the cause of MCOR.** .

2392W

Genetic analysis of autism spectrum disorder (ASD) based on developed diagnostic flows using next-generation sequencing (NGS). *Y. Enomoto[1], T. Yokoi[2], C. Hatano[2], I. Ohashi[2], Y. Kuroda[2], K. Ida[2], T. Naruto[1], K. Kurosawa[2].* 1) Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 2) Division of Medical Genetics, Kanagawa Children's Medical Center, Yokohama, Japan.

Next generation sequencing (NGS) is a new technology for detecting mutations comprehensively. NGS has a basically broad impact on many facets of biological and clinical research. We introduced NGS technology since 4 years ago and developed diagnostic flows of molecular diagnosis using NGS to identify mutations in genetic disorders and birth defects in KCMC. Autism spectrum disorder (ASD) is classified into brain dysfunction. ASD is characterized by impaired social interaction, communication disorder, stereotyped behavior and hyperactivity, etc. Division of Medical Genetics in Kanagawa Children's Medical Center (KCMC) has about 50 patients of ASD annually. The majority of ASD patients in KCMC have severe symptoms. ASD is considered to be a polygenic disease, which is influenced by many genes. 904 genes related to ASD are registered in The Human Gene Mutation Database (HGMD) as of June 2015. Their relationship is not sufficiently clear yet and pathogenic mechanism of ASD remains unclear. We infer severe ASD patients have mutations in genes related to ASD. Thus, we performed NGS analysis of severe ASD patients based on developed diagnostic flows, using exon sequencing of Mendelian disorders (TruSight One) or whole exon sequencing. We identified causative mutations in some patients, including two patients of severe ASD have mutations in CHD8 gene. We report about some causative mutations identified in this analysis and clinical feature of ASD patients.

2393T

Oro-facial-digital syndrome. (ofd), in a population from operation smile foundation in colombia. clinical, molecular and genetic characterization. *j. martinez[1], i. Briceno[1], a. Zarate[1], j. Becerra[1], s. Serrano[1], s. Idarraga[1], a. Collins[2], e. Soracipa[3].* 1) Medicine, Universidad de La Sabana, Bogota, bogota, Colombia; 2) University of Southampton. UK; 3) Colegio mayor de Cundinamarca. Bogota. Colombia.

Oro-facial-digital-syndrome. (ofd) is a ciliopathy causing phenotypic alterations such as cleft lip and palate, polydactyly, lobulated tongue, telecanthus, delayed psycomotor development and even death. SOFD has been classified in 14 different variants taking into account the clinical characteristics of the patients and some molecular findings. The actual incidence and the molecular and genetics features involved in this disease is currently unknown, therefore the goal of the project was to describe the clinical molecular and genetics characteristics present in the syndrome oro-facial-digital in patients attending the foundation “operation Smile” in Colombia. by genomic analysis. After informed consent was obtained, DNA extraction was performed. OFD! candidate genes C2Cd#, C5orf42 and TCTN3 were evaluated exome sequencing was performed and analyzed in the laboratory of genetics from “The university of Southampton”. The results show the posibility of a new variant of SOFD and genetic heterogeneity was observed.

2394F

Clinical and Molecular study of a series of 31 patients with chondrodysplasia with multiple dislocations. *E. Ranza[1,17], C. Huber[1], N. Levin[1], G. Baujat[1], Y. Alanay[2], L. Al Gazali[3], P. Bitoun[4], O. Boute[5], C. Coubes[6], N. Elcioglu[7], L. Faivre[8], D. Johnson[9], U. Kotecha[10], M. McEntagart[11], E. Mihci[12], B. G Nur[12], L. Perrin[13], C. Quelin[14], P. Terhal[15], I. C. Verma[10], B. Tuysuz[16], V. Cormier-Daire[1].* 1) Department of Genetics, INSERM UMR 1163, Université Paris Descartes- Sorbonne Paris Cité, Institut Imagine, Hôpital Necker Enfants Malades , Paris, 75015, France; 2) Acibadem University, School of Medicine, Department of Pediatrics, Pediatric Genetics Unit, Istanbul, Turkey; 3) Department of Pediatrics, College of Medicine and Health Sciences, UAE University, Al Ain, United Arab Emirates; 4) Génétique Médicale, Hôpital Jean Verdier, Bondy, France; 5) Génétique Clinique, Hôpital Jeanne de Flandre, Lille, France; 6) Département de Génétique Médicale, Hôpital Arnaud de Villeneuve, Montpellier, France; 7) Department of pediatrics-genetics, Marmara university medical school, Istanbul, Turkey; 8) Centre de Génétique, Hôpital d’Enfants,Dijon, France; 9) Sheffield Clinical Genetics Service, Sheffield Children’s hospital, Sheffield, UK; 10) Center of Medical Genetics, Sir Ganga Ram hospital, New Delhi, India; 11) Medical Genetics , St George’s Healthcare NHS Trust, London United Kingdom; 12) Akdeniz University School of Medicine, Division of Pediatric Genetics, Antalya, Turkey; 13) Unité de Génétique Clinique , Hopital Robert Debré, Paris, France; 14) Génétique Médicale, Hôpital Sud, Rennes, France; 15) Wilhelmina Childrens hospital, Lundlaan 6 3584 EA AB, Utrecht, the Netherlands; 16) Istanbul University, , Dept of Pediatric Genetics, Cerrahpasa Medical Faculty, Istanbul, Turkey; 17) Service of Genetic Medicine, University Hospitals of Geneva, Geneva, Switzerland.

The group of chondrodysplasia with multiple conditions includes 7 distinct entities all characterized by short stature, dislocation of large joints, hand anomalies and/or vertebral anomalies. Facial features, and cleft palate are also often reported. Among them, Larsen syndrome is the only one with an autosomal dominant inheritance and due to *FLNB* mutations while the 6 other recessive disorders, namely Desbuquois dysplasia (DD) type 1/Kim type and type 2, Spondylo-Epiphyseal Dysplasia with dislocations, chondrodysplasia with luxations, Larsen Like and Larsen of Reunion island syndromes are due respectively to *CANT1, XYLT1, CHST3, IMPAD1, B3GAT3,* and *B4GALT7* mutations. The aim of our study was to identify the molecular basis of a series of 31 patients presenting with these conditions. The series included 15 females and 16 males, originating from Algeria, France, India, Morocco, Netherlands, Turkey, UK and UAE. Consanguinity was reported in 20 families. 6 were fetuses and 25 were postnatal cases ranging in age from 1 to 30 years of age. Criteria for inclusion in the study were 1) Short stature < -2. 5 SD 2) Dislocation of at least one large joint 3) Hand anomalies and/or Vertebral anomalies. Referring diagnosis were DD type 1 (3), DD type 2 (6), Larsen syndrome (4), Larsen of Reunion Island (1), Catel Manzke syndrome (2) and unclassified (15). We performed exome sequencing. Our first analysis focused on the 7 known genes involved in this group but also on *B3GALT6, CHSY1, DSE, CHST14, SLC26A2* and *TGDS* involved in related conditions. All mutations were confirmed by Sanger sequencing. We identified mutations in 16/31 cases including *FLNB* (1) *CANT1* (3), *XYLT1* (2), *CHST3* (1), *B3GAT3* (3), *B4GALT7* (1), *B3GALT6* (1), *TGDS* (1), *CHSY1* (1), *DSE* (1) and *SLC26A2* (1). A complete deletion of *IMPAD1* was also observed at the homozygote state in one case. The initial diagnosis was accurate in 11/17 cases. Characteristic hand anomalies were consistently observed in patients with mutations in *CANT1, IMPAD1, CHSY1,* and *TGDS.* Intellectual disability and obesity were often observed in the course of the disease. Our study further expands the clinical spectrum of this group of conditions and also emphasizes proteoglycan synthesis impairment in recessive chondrodysplasia with dislocations. Finally, exome sequencing analysis is currently in progress for the remaining 14 cases.

2395W

A new multiple congenital anomaly syndrome due to a de novo variant in the ATP binding domain of FGFR2 detected by trio-based whole-genome sequencing. *T. Vilboux[1], D. L. Bodian[1], A. Khromykh[1], P. F. Cherukuri[1], S. Garcia[2], D. Glazer[2], R. Baveja[3], B. D. Solomon[1], J. Vockley[1], J. E. Niederhuber[1].* 1) Division of Medical Genomics, Inova Translational Medicine Institute, Falls Church, VA; 2) Personalis, Inc. 1350 Willow Rd Menlo Park, CA; 3) Fairfax Neonatal Associates - IFH, Inova Health System, Falls Church, VA.

We report a newborn with a previously undescribed severe, perinatally lethal disorder. The child presented with notable dysmorphic features (microtia, mild scaphocephaly, upslanting palpebral fissures), and other major anomalies including severe radial hypoplasia, short ulna, unilateral renal agenesis with cystic hypodysplasia. Research-based trio whole-genome sequencing identified a *de novo* variant in *Fibroblast Growth Factor Receptor 2* (*FGFR2*). Although heterozygous mutations in *FGFR2* have been reported to cause craniosynostosis syndromes such as Apert syndrome, Crouzon syndrome, Jackson-Weiss syndrome and Pfeiffer syndrome as well as Lacrimoauriculodentodigital (LADD) syndrome and Bent bone dysplasia syndrome (BBDS), the patient's phenotype appears to represent a distinct condition. Further analysis of all the previously reported variants revealed that this new variant is the first one described in the ATP binding region of FGFR2. Computational variant effect prediction tools suggest a severe consequence of this variant that would lead to a loss of function of FGFR2. Craniosynostosis syndromes are due to a gain of function whereas LADD is due to a decrease function of FGFR2 and BBDS to a more complex gain/loss of function. This new disorder is distinct from all the previously described *FGRF2* related disorders likely due to the specific functional domain in which the variant occurred and the degree to which the it is damaging.

2396T

A Three-Year-Old with ALVES syndrome: Case Review with Further Delineation of Phenotype and Inheritance. *M. Crenshaw.* Division of Genetics and Metabolism, All Children's Hospital/ Johns Hopkins Medicine, St. Petersburg, FL.

Alves syndrome (Tricho-Oculo-Dermo-Vertebral) Syndrome is characterized by ectodermal dysplasia with hypohydrosis in the setting of congenital arthrogryposis. There are fewer than five cases reported in the literature, and no causative gene is yet identified. Here I review a 3-year, 4-month-old you lady presenting at birth with distal arthrogryposis with camptodactyly. This is accompanied by mild dysmorphisms including ear anomalies and micrognathia with relative microcephaly. At 3-years of age, she has developed hypohydrosis and hyperkeratosis. Her genetic evaluation has included a chromosome microarray, serum amino acid and urine organic acid analyses, long chain fatty acid analysis, and carbohydrate deficient transferrin, which are normal. Mitochondrial DNA sequencing and deletion testing was normal. Testing for spinal muscular atrophy type 1 did not reveal a mutation. Sequencing of the genes *TNNI2, TNNT3*, *MYH3,* and *TPM2* associated with distal arthrogryposis showed no mutation in these genes. A brain MRI indicated some mild, nonspecific prominence of the ventricles. She is the product of a consanguineous union with no history of affected individuals in the family. This pattern is consistent with the proposed autosomal recessive inheritance of other described cases. Further evaluation by whole exome sequencing may be helpful in further evaluation for the causative gene and is planned.

2397F

Reverse phenotyping of a patient with *CRIPT* gene mutation and further delination of the associated phenotype. *B. Demeer[1,9], A. DADBAN[2], P. VABRES[3, 4], G. MORIN[1], B. ARAL[3,5], A. VARENTERGHEM[7], J. THEVENON[3,6], D. BREMOND-GIGNAC[8], J. St ONGE[3,5], J. RIVIERE[3,5], J. COURCET[3], C. THAUVIN[3,7], L. FAIVRE[3,7].* 1) Genetics department, Amiens-Picardie University hospital, Amiens cedex, France; 2) dermatology departement, Amiens-Picardie University hospital, Amiens cedex, France; 3) Equipe d'Accueil 4271 Génétique des Anomalies du Développement, Fédération Hospitalo-Universitaire, Université de Bourgogne, Dijon, France,; 4) Service de Dermatologie, CHU Le Bocage, Dijon, France; 5) Génétique des Anomalies du Développement, Fédération Hospitalo-Universitaire, Université de Bourgogne, Dijon, France. Laboratoire de Génétique Moléculaire, FHU-TRANSLAD, Plateau technique de Biologie, CHU, Dijon, France; 6) Centre de Référence Maladies Rares ;Anomalies du Développement et Syndromes Malformatifs; de l'Est, Centre de Génétique et Pédiatrie 1, Hôpital d'Enfants, Dijon, FranceGénétique des Anomalies du Développement, Fédération Hospitalo-Universitaire, Université de Bourgogne, Dijon, France; 7) Pediatrics department, CHU Amiens-Picardie, amiens, France,; 8) Ophtalmology department, CHU Amiens-Picardie, amiens, France,; 9) EA 4666, Université de Picardie Jules Verne, Amiens, France.

We report on a 3 ½ year old boy, with prenatal onset growth deficiency (height:-4SD), microcephaly (OFC:-3. 5 SD), transient neonatal pancytopenia, facial dysmorphism, feeding difficulties, developmental and speech delay, global hyperlaxity, significant sleep disturbance, and genital, ocular and extremities anomalies. He also presents with generalized pigmentation anomalies, and signs of ectodermal dysplasia. Array CGH (Agilent 60k), cytogenetic diagnosis of chromosomal breakage syndrome and metabolic screening are negative. The whole exome sequencing performed revealed a homozygous frame-shift mutation of the *CRIPT* gene, recently described as a novel primordial dwarfism gene (Shaheen et al, 2014). The mutation (c. 132delA), described as probably pathogenic, was confirmed in the homozygous state by Sanger sequencing. Both healthy consanguineous parents were proven to be carrier in the heterozygous state. Few available clinical data of the 2 described patients show very similar clinical appearance with strikingly facial dysmorphism, growth deficiency, microcephaly, psychomotor delay, and ocular and extremities anomalies. Mottled hypopigmentation is also described in the older patient. This report is an example of "reverse phenotyping". The first description of the *CRIPT* gene by Shaheen et al helped us to reach a diagnosis in our patient. Nevertheless the term of primordial dwarfism and its broad definition used by the authors can be confusing; and can prevent some clinicians from suggesting this diagnosis. Cutaneous signs seem also to be very specific, and need to be precisely looked at in additional patients presenting with this unique syndrome.

2398W

Whole Genome Sequencing Identifies Multiple Pathogenic Mutations in a Patient With Recurrent Infections and Multiple Congenital Anomalies. *A. Fadda[1], F. Butt[2], S. Tomai[1], A. Robay[3], S. Deola[1], R. Crystal[3], E. Wang[1], K. Fakhro[1,3], C. Cugno[1].* 1) Biomedical research, Sidra Medical and Research Center, Doha, Qatar; 2) Hamad Medical Corporation, Doha, Qatar; 3) Weill Cornell Medical College in Qatar, Doha, Qatar.

A consanguineous family of Arab ethnic origin with two unaffected siblings and a 7-year-old female child affected by recurrent infections and multiple congenital anomalies was screened by Whole Genome Sequencing (WGS) to identify the offending mutations. The disorder is marked by pan-leukopenia, recurrent severe infections since birth including pneumonia, bronchitis, pyelonephritis, urinary tract infections leading to hypoplasia and functional impairment of the left kidney, and otitis media. She is developmentally delayed with stunted growth and learning difficulties. There is microcephaly, thin hair, triangular face, small eyes, low-set ears, finger and toe clubbing, dry skin, vesicoureteral reflux and skeletal abnormalities including 11 pairs of ribs. Family history is marked by three first cousins (two males and one female, also products of consanguineous marriages) that died in early childhood from severe infections. WGS was performed for the patient, two unaffected siblings, and mother. The data was mined for homozygous damaging variants conforming to autosomal recessive inheritance. A novel *LIG4* (DNA ligase) homozygous nonsense c. T1312C/p. Y438H mutation was detected. *LIG4* pathogenic mutations cause immunodeficiency, growth and developmental delay, unusual facial features and microcephaly, justifying the clinical presentation. Two additional pathogenic mutations were detected. A *LRIG2* homozygous nonsense c. C2125T/p. R709X mutation previously reported to cause Uro-facial syndrome characterized by congenital urinary bladder dysfunction and inverted smile that are not apparent in the patient; an *LFNG* (lunatic fringe involved in Notch signaling pathways) homozygous frameshift c. 163-166dupGATG/p. E56fs*2 absent in 1000 Genomes and ESP databases. *LFNG* mutations cause spondylocostal dysostosis. The patient is being investigated for congenital vertebral anomalies. The identification of multiple pathogenic mutations by WGS might indicate the presence of multiple autosomal recessive conditions due to the extensive consanguinity, resulting in complex clinical presentation. But it clearly demonstrates that when applying WGS data to clinical diagnoses, all significant variants should be considered and scrutinized for their clinical impact, irrespective of the observed clinical symptomatology.

2399T

***PTCH* gene deletion in a individual presenting holoprosencephaly associated to bilateral anophthalmia and orbital cyst.** *L. Bicudo[1,2], B. Gamba[2], J. Arruda[1], B. Maranhão[1], A. Costa[3].* 1) Genetics, Bioscience Institute, Federal University of Goias, Goiania, Brazil; 2) Genetics, Bioscience Institute, Sao Paulo State University, Botucatu, Brazil; 3) Genetics, Sao Paulo University, Bauru, Brazil.

The SHH-PTCH pathway is a dosage-sensitive signalling network which is particularly critical in cellular growth and development during embryonic development. Inherited and sporadic mutations in genes in this pathway have been implicated in a number of human birth defects and adult cancers . PTCH haploinsufficiency in humans has been postulated to interact with the expression of other genes resulting in anomalies of the eye, brain and digits, or to explain microphthalmia and orbital cysts which are functionally active during neurulation. Human *PTCH* mutations commonly cause sporadic and hereditary NBCS; however, primary eye involvement and/or holoprosencephaly due to *PTCH* mutations are uncommon. Holoprosencephaly (HPE) is a malformation sequence where the cerebral hemispheres fail to separate into distinct left and right halves. It can be associated with midline structural anomalies of the central nervous system and/or face. *SHH* is the major gene implicated in HPE and it plays a critical role in early forebrain and central nervous system development. Eye anomalies are part of the clinical spectrum of the HPE, ranging from cyclopia to variable structural eye defects with the most extreme variant represented by bilateral anophthalmia. Eye anomalies mainly represented by microphthalmia and orbital cyst was reported in a girl with NBCS who presented a mutation in the *PTCH* gene. Here we report a patient with HPE associated to bilateral anophthalmia and orbital cyst who presented a deletion showed by MLPA analysis in *PTCH1* gene. This is the third variation in *PTCH1* gene identified in patients with ano/microphthalmia and orbital cysts. The occurrence of severe eye involvement in these patients as well as the presence of HPE in two of them and BCNS in the other, make obvious the spectrum of manifestations occurring in mutations of the *PTCH* gene that includes tumorigenesis, ocular anomalies, and HPE.

2400F

Male Newborn Presented With Hyponatremia Who Has 21-Hydroxylase Deficiency. *A. Ecevit[1], H. Verdi[2], FB. Atac[2], S. Tulgar Kinik[3], HI. Ruhi[4], FA. Tukun[4].* 1) Department of Pediatrics, Division of Neonatology, Baskent University Shool of Medicine , Ankara, Turkey; 2) Baskent University Shool of Medicine Department of Medical Biology, Ankara, Turkey; 3) Baskent University Faculty of Medicine, Division of Pediatric Endocrinology, Ankara, Turkey; 4) Ankara University School Of Medicine Department Of Medical Genetics.

One of the most common inherited metabolic disorders is congenital adrenal hyperplasia (CAH). The commonest form of CAH is steroid 21-hydroxylase deficiency (21OHD, MIM +201910) caused by sequence variants in the 21-hydroxylase gene (CYP21A2, GeneID 1589) that accounts for 90–95% of cases. The clinical features reflect the magnitude of the loss of function mutations that range from complete loss of enzyme function to partial enzyme activity. The two types of the disease are classic forms, known as the salt-wasting and simple virilizing types while the third type is called the non-classic type. Newborns with salt-wasting 21-OHD CAH are at risk for salt-wasting crises. Females are frequently diagnosed at birth because of ambigious genitalia but male newborns are undiagnosed. Females with the non-classic type of 21-hydroxylase deficiency may experience hirsutism, male pattern baldness, irregular menstruation, and decreased fertility. However some individuals have no symptoms of the disorder. Here we present a case with severe hyponatremia who was compound heterozygote for 21-hydroxylase deficiency. **Case:** One- month old baby boy presented with the vomiting and failure to weight gain. Physical examination was normal except penile length 4. 5 cm and mild hyperpigmentation. The parents were not relative. Mother and father were healthy and no remarkable family history. Laboratory findings: serum Na<100 mmol/L K:5. 7 mmol/L urinary sodium:33 mmol/L plasma renin activity:22 ng/mL/hour (0. 51-2. 64)ACTH:192 ng/mL (<46) Cortisol:5. 1 mcg/dL aldosterone:1357 pg/L (50-900)DHEAS:4074 mcg/dL (31-214) testosterone:1004 ng/dL (142-923) androstenedione:65 ng/mL (006-0. 68)17OH progesterone:815 ng/mLHe was diagnosed as congenital adrenal hyperplasia. Molecular analysis revealed that: Proband was compound heterozygote for I2A/C656G and Q318X. Mother was homozygote for I2A/C656G mutation and Father was heterozygote for Q318X. With the treatment of hydrocortisone, fludrocortisone and salt replacement, he is going well. As a result, due to lack of ambigious genitalia in male newborns with salt-wasting CAH are generally undiagnosed at birth and come to medical attention with salt losing crisis. Although the mother is homozygote for I2A/C656G, she is healthy without hirsutism, infertility or menstrual irregulatory. That case impresses the probability of discordance between genotype and phenotype in cases with CAH.

2401W

Concomitant *de novo NFIX* heterozygous deletion and truncating *EVC* mutation escaping nonsense-mediated mRNA decay in a subject with severe intellectual disability and postnatal growth delay. *D. Uehara[1], S. Hayashi[1,2,3], K. Tanimoto[1], S. Mizuno[4], J. Inazawa[1,2,5].* 1) Department of Molecular Cytogenetics, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Tokyo, Japan; 2) Hard Tissue Genome Center, Tokyo Medical and Dental University, Tokyo, Japan; 3) Department of Neurobiology, Yale University School of Medicine, CT; 4) Department of Pediatrics, Central Hospital, Aichi Human Service Center, Kasugai, Japan; 5) Bioresource Research Center, Tokyo Medical and Dental University, Tokyo, Japan.

NFIX heterozygous aberrations at 19p13. 2, namely whole-gene deletions and point mutations affecting its DNA-binding domain, have been described in subjects whose common features include intellectual disability (ID), overgrowth and macrocephaly. These features are now referred to as Sotos syndrome 2 or Malan syndrome. Among our screening of 645 Japanese subjects with ID and multiple congenital anomalies of unknown etiology, we report on a female subject with a 393-kb *de novo* deletion at 19p13. 2 encompassing *NFIX*, detected by SNP arrays. Clinical features included severe ID, hypotonia, small intestinal atresia and mesenteric malformation, low stature (-3. 5 SD) and weight (-2. 5 SD), along with a relative macrocephaly (-0. 7 SD) at the age of seven years old. Her birth size was normal. In order to identify other variants that could explain her postnatal growth delay, we performed a complementary investigation with whole-exome sequencing, which revealed a novel nonsense mutation (c. 1273C>T or Q425X) in the *EVC* gene, inherited from the mother. RT-PCR followed by Sanger sequencing in lymphoblastoid cell lines revealed that, as expected, the mutant allele in the mother was degraded by nonsense-mediated mRNA decay (NMD). However, the same mutant allele escaped from NMD in the proband. *EVC* is the causative gene of the autosomal recessive Ellis-van Creveld syndrome, characterized by short stature, postaxial polydactyly, dysplastic nails and teeth, and cardiac defects in 60% of the cases. Unlike the previous reports of *NFIX* haploinsufficiency leading to overgrowth, our case presents short stature that might be explained as a result of a dominant-negative effect exerted by the mutant *EVC* protein. The reason why the mutant allele escaped only in the proband remains unclear.

2402T

Macrothrombocytopenia and Developmental Delay with a *de novo CDC42* Mutation: Yet Another Locus for Thrombocytopenia and Developmental Delay. *K. Kosaki[1], T. Takenouchi[1], T. Uehara[1], R. Kosaki[2].* 1) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan; 2) Division of Medical Genetics, National Center for Child Health and Development.

Isolated thrombocytopenia combined with a developmental delay classically suggests a chromosome 11q deletion including the *FLI1* locus, a transcription factor that is essential for megakaryopoiesis (Jacobsen syndrome) or a 21q22 microdeletion that includes the *RUNX1* locus. Here, we report a female patient who presented with macrothrombocytopenia (64,000 μ/L)and a developmental delay. The observation suggests presence of yet another locus, *CDC42*, for thrombocytopenia with developmental delay. Thepatient hadmidfacial hypoplasia, synophrys, mild ptosis, eversion of the lateral portion of the lower eyelid, exotropia, short philtrum, thin upper lip and generalized eczema without petechia. She also hadlymphedema in her legs. At age18 years,she was unable to climb stairs without support andcommunicated with simple words. A complete blood count showed a platelet count of as low as 52,000 μ/L and normal white blood cell count of 3,000 μ/L and normal hemoglobin level. A peripheral blood smear revealed enlarged platelets. Whole exome sequencing identified a *de novo* mutation in the GTPase *CDC42*c. 191A>G, p. Tyr64Cys as the sole candidate in an autosomal dominant de novo or recessive model. The mutation was confirmed using Sanger sequencing. The strong degree of similarity between the phenotype of the proposita and that of mice lacking *Cdc42* suggests a probable causal relationship. The conditional homozygous knock-out of *Cdc42* in mice results in mild thrombocytopenia and an increase in platelet size, i. e. , macrothrombocytopenia. Phenotypes other than the macrothrombocytopenia observed in the proposita could also be ascribed to the defect in *CDC42.* First, the developmental delay could be explained by the *CDC42* mutation, since Cdc42 plays a critical role in the proliferation of neuronal progenitor cells. Second, the lymphedema could be ascribed to the *CDC42* mutation, since Cdc42 directly interacts with Rac1, the defect of which leads to lymphedema in mice. Third, the severe eczema could also be explained by the *CDC42* mutation, since CDC42 directly interacts with WASP, the defect of which leads to Wiskott-Aldrich syndrome, which is characterized by eczema accompanied with thrombocytopenia. Apparent phenotypic similarity between WAS and the proposita may reflect combinatory role of WASP and CDC42 in cytoskeletal actin polymerization in various biological processes including platelet genesis.

2403F

Holoprocencephaly and Neural tube defects: Strange bedfellows: Is there a third and even stranger bedfellow-Heterotaxy? *S. Krishnamurthi, T. Sadler, R. Stevenson.* Genetics, Greenwood Genetic Center, Greenwood, SC.

Holoprosencephaly (HPE) and neural tube defects (NTDs) are both midline defects with an embryological basis in the closely intertwined dynamic processes of gastrulation and neurulation. Coincidental occurrence of these two conditions is approximately 1:16,000,000. We report a series of 5 cases of HPE with associated NTDs and laterality defects over a 23 year period (1,284,000 births, 1992-2015), which represents a 50-fold increase over the expected combined prevalence. Six additional cases with NTDs had facial findings suggestive of HPE but insufficient information to be conclusive. Some of our cases of HPE and NTDs also exhibited laterality defects. Thus we hypothesize that the etiology of HPE, NTDs and heterotaxy may be related, possibly involving genes that regulate patterning the body by early in the third week of embryonic development. Both HPE and NTDs exhibit phenotypic variability that appear to be chronologically and topographically related to alterations in genetic determinants in highly conserved regulatory pathways, notably the Sonic Hedgehog signalling pathway. Decreased expression of *sonic hedgehog* (*SHH*) has been associated with HPE, whereas in cases of NTD, expression is increased. Of the 4 major HPE associated genes, *SHH, ZIC2* and *SIX3* are all linked to the SHH pathway. *TGIF* encodes a co-repressor for the TGFb/Nodal signaling and appears to function as a key regulator of SHH signaling. Reduced expression of *Zic2* in knockdown models results in HPE and NTD. Both defects also appear to be more frequent with *TGIF* mutations. Loss of function of *TGIF* has been implicated in failure of gastrulation. Genes in these pathways have also been implicated in establishing the left-right (L-R) axis. Mutations involving genes responsible for neural patterning and left-right asymmetry have been demonstrated to produce midline and laterality defects in knockout animal models. SHH is also a key gene involved with *NODAL* in establishing the left-right (L-R) axis. Both genes are expressed asymmetrically in the primitive node to regulate *PITX2* expression on the left side that then establishes the L-R axis. Both *SHH* and *NODAL* knockouts have cyclopia and heterotaxy, while *STIL* knockouts have decreased expression of *SHH*, bilateral expression of *NODAL* and *PITX2*, cyclopia, NTD, and heterotaxy. We hope to establish a common genetic basis for these defects by examining genes in this pathway and other regulatory contributors through whole genome sequencing.

2404W

Whole Exome Sequencing of a Syrian Family with Oral Cleft and Split Hand/Foot Syndrome Reveals Novel *de novo* Mutation in *TP63*. *C. L. Simpson[1], D. C. Kimble[2], S. C. Chandrasekharappa[2], N. I. S. C. Comparative Sequencing Program[3], K. Alqosayer[4], H. Albacha-Hejazi[5], J. E. Bailey-Wilson[1].* 1) Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Baltimore MD 21224; 2) Cancer Genetics and Comparative Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda MD 20814; 3) NIH Intramural Sequencing Center, National Human Genome Research Institute, National Institutes of Health, Bethesda MD 20814; 4) Prime Health Clinic, Jeddah, Riyadh 21511, Saudi Arabia; 5) Hejazi Clinic, P. O. Box 2519, Riyadh 11461, Saudi Arabia.

Oral clefts are a common, heterogeneous birth defect affecting 1 in 750 live births in the US, typically present as cleft lip with or without cleft palate. Split hand/foot malformations are a less common heterogeneous birth defect with 1 in 8,500 to 25,000 live births affected. Here we present results of whole exome sequencing (WES) analysis in a Syrian family with an affected proband who presented with both orofacial clefting and split hand/foot. An additional family member affected with an oral cleft but without limb abnormalities died at birth and was unavailable for analysis. We performed WES in the proband, both parents and two unaffected siblings. Quality control was performed in Golden Helix SVS and PLINK. Variant annotation, Mendelian inconsistencies and novel variants in known cleft genes were examined. Candidate variants were validated using Sanger sequencing. 185,474 SNVs and 18,058 INDELs were released by the sequencing center. Genotypes were dropped for depth < 10, quality score < 10 or genotype quality to depth ratio of < 0. 5. Variants monomorphic after filtering were dropped. There were 11,905 INDELS and 137,989 SNVs available for analysis after all filtering steps. Examining Mendelian errors revealed 35 candidate *de novo* events of which 1 was in a known oral cleft and split hand/foot gene, TP63 (c. 956G>T, p. Arg319Leu) and confirmed by Sanger sequencing. A tri-allelic SNV was also found in HLA-DRB5, which is similar to and overlaps with the known oral cleft gene HLA-DRB1. Mutations in*TP63*are associated with multiple disorders which include oral clefts and limb malformations, including Ectrodactyly, Ectodermal Dysplasia and Cleft Lip/Palate Syndrome 3 (EEC3), Split Hand/Foot Malformation 4 (SHFM4), Hay-Wells Syndrome and acro-dermato-ungual-lacrimal-tooth (ADULT) syndrome, and most mutations are dominant. The mutation presented here is a *de novo* event but is also a novel mutation. However, another known mutation 1bp upstream (rs121908839, c. 955C>T, p. Arg319Cys) affects the same codon but produces a different amino acid (AA) change and provides strong evidence that mutating this AA is deleterious. However, it is not clear whether this mutation is only responsible for the split hand/foot or whether it is also causing oral cleft. The p. Arg319Cys mutation is known only to cause split hand/foot, and not oral clefts. Further functional analysis is needed to see if this mutation is the cause of the entire phenotype or only the limb malformation.

2405T

Biallelic *SZT2* mutations in a case with early infantile epileptic encephalopathy. *A. Miyauchi[1], H. OSAKA[1], N. SHIMOZAWA[2], N. MATSUMOTO[3], H. SAITSU[3], T. YAMAGATA[1].* 1) Pediatrics, Jichi Medical University, Shimotsuke-shi, Tochigi, Japan; 2) Division of Genomics Research, Life Science Research Center, Gifu University, Gifu, Japan; 3) Department of Human Genetics, Yokohama City University Graduate School of Me dicine, Kanagawa, Japan.

(Introduction) Biallelic mutations in*SZT2*have been recently reported in patients with infantile encephalopathy, who showed severe development delay, intractable epilepsy, as well as thick and short corpus callosum and persistent cavum septum pellucidum on brain MRI. We report a fourth case with biallelic*SZT2*mutations. (Case Report) Patient is a 3-year-old-girl who is the third child of healthy nonconsanguineous parents. An elder sister, showing severe developmental delay and intractable epilepsy, died at age of 2 with acute encephalopathy. An elder brother was born at 23 weeks gestation and died soon after birth. She was born at 1,398 g for 32 wks. She had facial dysmorphic features included a high forehead and bilateral ptosis. Symptoms include severe developmental delay with hypotonia, failure to thrive, and intractable epilepsy. At the age of 2 mos, complex partial seizures and secondary generalized seizures began, which were refractory to multiple antiepileptic drugs. Interictal EEG showed bilateral frontal dominant spike-and-waves. Brain MRI showed a short and thick corpus callosum. She developed thrombocytopenia (~3,000 to 200,000 /μl) since 2 yrs and 6 mos. She did not respond to discontinuation of suspicious drugs and needed platelet transfusions. A G-banding and array CGH revealed no abnormal findings. By whole-exome sequencing and subsequent Sanger sequencing, we identified compound heterozygous frameshift mutations in *SZT2*, in which c. 3700_3716del and c. 5482del were transmitted from her father and mother, respectively. (Discussion) To date, only three families with *SZT2* mutations have been reported including our family. Clinical features of previous two families were similar with those of our case except for thrombocytopenia which is uniquely observed in our case. Expression of *SZT2* is reported to be increased by glutamic neurotoxic stimulation and co-localize with catalase at the peroxisome. In mice, *Szt2* has been reported to regulate seizure threshold and influence epileptogenesis. Association of SZT2 to epileptogenicity and thrombocytopenia will be a target for future investigation.

2406F

Analysis of the planar cell polarity regulator gene *PTK7* in neural tube defects. *Y. Lei[1], H. Zhu[1], G. Shaw[2], E. Ross[3], R. Finnell[1].* 1) Dell Pediatric Research Institute, Department of Nutritional Sciences, The University of Texas at Austin, Austin, Texas; 2) Department of Pediatrics, Division of Neonatology, Stanford University School of Medicine, Stanford, CA; 3) Center for Neurogenetics, Brain and Mind Research Institute, Weill Cornell Medical College, New York, NY.

Neural tube defects (NTDs), including anencephaly, spina bifida and craniorachischisis, are severe birth defects that affect 0. 5-1 in 1000 births. Recently, mutations in planar cell polarity (PCP) pathway genes were implicated in the pathogenesis of NTDs in both the mouse model and in human cohorts. Mouse models indicate that the homogenous disruption of *Ptk7* gene, a PCP regulator, results in craniorachischisis, while embryos that are double heterozygous for *Ptk7* and *Vangl2* Lp mutations result in spina bifida. In this study, we sequenced exons of the human *PTK7* gene in 192 spina bifida patients and 190 controls from a California population. Included were live-born cases with spina bifida. Controls were randomly selected among all liveborn infants corresponding to the same geographic area and birth time periods as cases. DNA for genotyping was obtained from newborn bloodspots. We identified three rare (MAF < 0. 01) missense heterozygous *PTK7* mutations (p. Thr186Met, p. Arg630Ser and p. Tyr725Phe) in spina bifida cases. Two of the variants (p. Arg630Ser and p. Tyr725Phe) which were predicted to be damaging by PolyPhen, were absent in well-matchedall controls, as well as in ExAC control database. No novel damaging missense *PTK7* mutation was identified in control subjects. Our study suggests that missense mutations in *PTK7* appear to contribute to the genetic risk of spina bifida in the studied population.

2407W

A six-year-old boy with compound heterozygote mutations of *ERCC6* and a splice site mutation of *ATRX*. *K. Kojima[1], A. Miyauchi[1], N. Ikeda[1], M. Yamazaki[1], Y. Monden[1], H. Osaka[1], H. Shimbo[2], T. Wada[3], H. Saitsu[4], N. Matsumoto[4], T. Yamagata[1].* 1) Pediatrics, Jichi Medial University, Shimotsuke-shi, Tochigi, Japan; 2) Department of Neurology, Kanagawa Children's Medical Center; 3) Depatment of Medical Ethics, Kyoto University Graduate School of Medicine; 4) Depatment of Human Genetics, Yokohama City University Graduate School of Medicine.

(Introduction)Cockayne syndrome (CS) is an autosomal recessive disorder, showingcachectic dwarfism, severe intellectual disability (ID),progeroid appearance,sensorineural hearing loss, intracranial calcification, etc. The responsible gene is*Excision repair cross-complementation group 6 (ERCC6)*and*ERCC8.* Alpha thalassemia X-linked intellectual disability (ATRX) syndrome is characterized byID, characteristic facial features, the genitourinary tractabnormalitiesand alpha thalassemia, caused by the defect of*ATRX*. We report a boy withmutations in two genes that werecompound hererozygote mutation of*ERCC6*and alsosplice site mutationof*ATRX*.

(Patient) Patient was a 6year-old boyborn from non-consanguineous parents with uneventful delivery. Hisbirth weight was 3. 36kg (0. 2SD). At 5 year-old, his height was 91cm (-4. 3SD) and weight 14. 7kg(-1. 4SD). He hadbird like appearance, andmoderate ID with several meaningful single words. He also hadspastic diplegiawith knee and ankle joint contracture, and could walk several steps. He did not have hearing loss and genital abnormality. His cranial CT showed calcification on basal gangliaand MRI showed mild cerebellar atrophy. His blood analysis showed no thalassemia.

(Methods and Results)Genomic DNA was obtained from this patient and hismother after the informed consent. By whole exomesequencing, we detectedcompound heterozygousmutationsin*ERCC6*; c. 1936del (p. Asp646Thrfs*52) that wasalsodetected in his mother, and c. 1627A>T (p. Ile543Phe)that was supposed to be inherited from his father. In addition,a splice site mutationin*ATRX*(c. 21-G>A)was detected. RT-PCR analysisfor*ATRX*detected the mRNA that was spliced from exon 1 to 3 with exon 2 skipping. Immunostaining of patient's fibroblastsusing rabbit polyclonal antibody for C terminal region of ATRX detected the expression of ATRX in nuclei.

(Discussion)This patient had spastic diplegia and some features of Cockayne syndromealthoughthey were mild, buthadunremarkablephenotypes of ATRX syndrome. The product of splicing variant skipping exon 2 induced by c. 21-G>A in*ATRX*was supposed to terminate early. A male patient who also had exon 2 skipping mutation of*ATRX*was reported to haveChudley–Lowry syndrome. To elucidate the reason for milder phenotype in this patient, the interference of these two genes should be examined.

2408T

Identification of a new mutation in *C2CD3* by whole exome sequencing in an oral-facial-digital syndrome family with intrafamilial variability. *S. Scheidecker[1], Y. Perdomo[2], V. Geoffroy[1], D. Chaigne[3], M. Koob[4], C. Stoetzel[1], H. Dollfus[1,2].* 1) Laboratoire de Génétique Médicale, INSERM U1112, IGMA, Faculté de Médecine, UDS, Strasbourg, France; 2) Centre de Référence pour les affections rares en Génétique ophtalmologique, CARGO, HUS, Strasbourg, France; 3) Service de Pédiatrie, HUS, Strasbourg, France; 4) Service de RadioPédiatrie/Imagerie 2, HUS, Laboratoire ICube, UMR 7357/FMTS/UDS-CNRS, Strasbourg, France.

The oral-facial-digital (OFD) syndrome is a heterogeneous group of disorders that affect almost invariably the oral cavity, the face and the digits. To date, more than 13 types have been distinguished based on characteristic clinical manifestations with other organ systems involved. Extensive overlap in the phenotype exists between the various forms with different inheritances. Recently, *C2CD3* gene was reported to be associated with an autosomal recessive OFD syndrome in a child with severe microcephaly, micropenis, severe intellectual disability, retinopathy, and cerebral MRI scan anomalies, and a fetus male with oral anomalies, polydactyly and cerebral malformations (Thauvin-Robinet, et *al*. 2014). *C2CD3* encodes a protein colocalizing with OFD1 and having an essential role in the assembly of centriolar distal appendages. We report a new *C2CD3* variant in a family with two affected siblings with a phenotype of OFD spectrum. The propositus was the second child of related healthy parents. The girl was born with hypotonia and microcephaly. Were also noted hyperplastic oral frenula, hamartomas of the tongue and partial choanal atresia. She had breath difficulties with episodes of apnea in the first days of life. She developed retinopathy and mild ataxia and had learning difficulties. Cerebral MRI scan showed cerebellar anomalies associated with neuronal migration defect. Clinical examination revealed mild dysmorphism features. Her younger brother had also microcephaly at birth with hypotonia. He developed a retinopathy without other anomalies at the age of 12 months. Whole exome sequencing performed on the family trio (the affected propositus and the healthy parents) revealed a rare variant (c. 4922T>C; p. Val1641Ala) at homozygous state in the affected child. The variant was also detected by Sanger sequencing in the younger affected brother at homozygous state. The identification of new individuals carrying *C2CD3* mutation highlights the interest to look at variants in this gene very recently described in autosomal recessive form of oral-facial-digital syndrome. Description of more C2CD3 cases could help to delineate a new subtype of OFD.

2409F

Associated non diaphragmatic anomalies among cases with congenital diaphragmatic hernia. *C. Stoll, Y. Alembik, B. Dott, MP. Roth.* Genetique Medicale,Faculte de Medecine, Strasbourg, France.

The etiology of congenital diaphragmatic hernia (CDH) is unclear and its pathogenesis is controversial. Because previous reports have inconsistently noted the type and frequency of malformations associated with CDH, we assessed these associated malformations ascertained between 1979 and 2007 in 386,022 consecutive births of known outcome. Of the 139 patients with CDH, 85 (61. 2%) had associated anomalies. These malformations included: chromosomal abnormalities (n=25, 18. 0%); recognized conditions non-chromosomal (n = 24, 17. 3 %): syndromes (Fryns syndrome, fetal alcohol syndrome, De Lange syndrome, CHARGE syndrome, Fraser syndrome, Goldenhar syndrome, Smith-Lemli-Opitz syndrome, multiple pterygium syndrome, Noonan syndrome, spondylocostal dysostosis, and Beckwith-Wiedemann syndrome), malformation sequences (laterality sequence, ectopia cordis), malformation complexes (limb body wall complex), and non syndromic multiple congenital anomalies (MCA) (n=36, 25. 9%). Malformations of the cardiovascular system (n=53, 27. 5%), urogenital system (n=34, 17. 6%), musculoskeletal system (n=29, 15. 0%), and central nervous system (n=19, 9. 8%) were the most common other congenital malformations. We observed specific patterns of malformations associated with CDH which emphasizes the need to evaluate all patients with CDH for possible associated anomalies. One should be aware that the anomalies associated with CDH can often be classified into a recognizable malformation syndrome or pattern, in 57. 6% of the cases.

2410W

An aberrant splice acceptor site due to a novel intronic nucleotide substitution in MSX1 gene is the cause of congenital tooth agenesis in a Japanese family. *T. Tadashi[1], K. Masashi[1,2,7], N. Mitsuko[3], M. Junichiro[1,4], Y. Seishi[1,5], S. Akio[1,2,7], G. Hiroki[1,7], N. Atsuo[6], H. Yujiro[7], M. Hitoshi[1], S. Kazuo[1], M. Naomichi[3], T. Yoshihito[7].* 1) Maxillofacial Surgery, Aichi-Gakuin University School of Dentistry, Chikusa-ku, Nagoya , Japan; 2) Dentistry Oral and Maxillofacial Surgery, Ogaki Municipal Hospital, Ogaki, japan; 3) Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 4) Oral and Maxillofacial Surgery, Toyota Memorial Hospital, Toyota, Japan; 5) Dentistry and Oral Surgery, Aichi Children's Health and Medical Center, Obu, Japan; 6) Embryology, Institute for Developmental Research, Aichi-Human Service Center, Kasugai, Japan; 7) Perinatology, Institute for Developmental Research, Aichi-Human Service Center, Kasugai, Japan.

Congenital tooth agenesis is caused by mutations in the MSX1, PAX9, WNT10A, or AXIN2 genes. Here, we report a Japanese family with nonsyndromic tooth agenesis caused by a novel nucleotide substitution in the intronic region between exons 1 and 2 of the MSX1 gene. Because the mutation is located 9 bp before exon 2 (c. 452-9G>A), we speculated that the nucleotide substitution would generate an abnormal splice site. Using cDNA analysis of an immortalized patient blood cell, we confirmed that an additional 7-nucleotide sequence was inserted at the splice junction between exons 1 and 2 (c. 451_452insCCCTCAG). The consequent frameshift generated a homeodomain-truncated MSX1 (p. R151fsX20). We then studied the subcellularlocalization of truncated MSX1 protein in COS cells, and observed that it had a whole cell distribution more than a nuclear localization, compared to that of wild-type protein. This result suggests a deletion of the nuclear localization signal, which is mapped to the MSX1 homeodomain. These results indicate that this novel intronic nucleotide substitution is the cause of tooth agenesis in this family. To date, most MSX1 variants isolated from patients with tooth agenesis involve single amino acid substitutions in the highly conserved homeodomain or deletion mutants caused by frameshift or nonsense mutations. We here report a rare case of an intronic mutation of the MSX1 gene responsible for human tooth agenesis. In addition, the missing tooth patterns were slightly but significantly different between an affected monozygotic twin pair of this family, showing that epigenetic or environmental factors also affect the phenotypic variations of missing teeth among patients with nonsyndromic tooth agenesis caused by an MSX1 haploinsufficiency.

2411T

The Etiological Exploration of PHACE Syndrome. *K. F. Schilter[1], D. W. Metry[2], I. J. Frieden[3], W. B. Dobyns[4], E. A. Worthy[5], B. A. Drolet[1], D. H. Siegel[1].* 1) Departments of Dermatology and Pediatrics, Medical College Wisconsin, Milwaukee, WI; 2) Department of Dermatology, Baylor College of Medicine, Houston, TX; 3) Departments of Dermatology and Pediatrics, University of California San Francisco, San Francisco, California; 4) Departments of Pediatrics and Neurology, University of Washington, Seattle Washington; 5) Departments of Pediatrics and Human and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI.

PHACE syndrome is an uncommon neurocutaneous syndrome consisting of features including Posterior fossa malformations, Hemangiomas, Arterial anomalies, Cardiac defects, and/or anomalies of the Eye (OMIM #606519). The exact etiology and pathogenesis of PHACE syndrome and the associated features are not yet known. One of the most frequently associated features of PHACE syndrome are the cerebrovascular anomalies present in 84%. In a cohort of 70 PHACE patients with known cerebrovascular anomalies, 56% were further described as having arterial dysgenesis, and another 47% having anomalous course and/or origin of the arteries. It was also noted that in all but one case, the associated features of arterial and brain anomalies are found ipsilateral to the hemangioma, suggesting factors early in development. Furthermore, the effect of the anterior rather than posterior circulation suggests involvement of neural crest lineage. Based on current knowledge of the syndrome, it is hypothesized that there is a mosaic genetic component contributing to the development of multiple congenital anomalies, most notably the cerebrovascular anomalies. The PHACE Syndrome International Clinical Registry and Genetic Repository was established in 2006, and currently has approximately 210 patients enrolled. To date, genetic studies conducted through the PHACE syndrome registry, including X-inactivation analysis, copy number variation analysis, and candidate locus analysis, have not identified a genetic etiology. While these studies did not reveal any specific genetic variant or copy number change, their findings may provide information in the future about modifying or susceptibility factors in PHACE syndrome which may account for the variability in the phenotype. Current genetic analyses include whole exome and whole genome sequencing to further investigate genetic involvement in PHACE syndrome. Analyses are being directed with somatic mosaicism as the hypothesis of the etiology.

2412F

Novel heterogenous mutation in *DYNC1H1* gene in a boy with arthrogryposis and motor neuropathy. *H. Alsharhan, A. Rajadhyaksha.* Pediatrics, Miami Children's Hospital, Miami, FL.

We present a three-year old boy, currently wheelchair bound, who was born at 35 weeks gestation, via caesarian section due to breech presentation, found to have arthrogryposis, severe clubfeet, bilateral femur fractures and hip dislocation at birth. Prenatal period was significant for for decreased fetal movements, and highly suspected arthrogryposis in fetal ultrasonography. Later, he developed knee contrctures, hypotonia, and found to have global developmental delay with further femur fractures, requiring multiple orthopedic interventions. Extensive workup was conducted, which yielded normal results. His newborn screen was normal along with his karyotype (XY). Arthrogryposis panel was negative as well. His brain magnetic resonance imaging (MRI) showed mild enlarged ventricles, but spine MRI was unremarkable. His cardiac evaluation, hearing and vision were all within normal limits. Muscle biopsy was non contributory. At age of 2 years, whole exome sequencing was performed, which identified a novel de novo heterozygous mutation in dynein heavy chain1*DYNC1H1,c. 3623G>Tp. W1208L. DYNC1H1*gene encodes for a cytoplasmic dynein heavy chain, which is a motor protein complex, presents in muscles, nerves, bones and other tissues. Cytoplasmic dyneins, large mechanochemical motor compromising heavy, intermediate, and light chains, are implicated in intracellualr motility, including retrograde axonal transport, neuronal migration, spindle dynamics and other intracellular motility functions. Mutations of*DYNC1H1*gene have been associated with autosomal dominant Charcot-Marie-Tooth type, distal spinal muscular atrophy (SMA), predominantely affecting the legs (SMA-LED), and mental retardation, indicating that dynein dysfunction is associated with a range of phenotypes in humans involving neuronal development and maintenance. Here we report the first patient with this mutation who manifested with both bone and neurological manifestations. This case further expands the set of pathological mutations of*DYNC1H1*gene and thus reinforce role of cytoplasmic dynein in both motor neuropathies and bone disorders.

2413W

The analysis of *IL36RN* mutations in subtypes of pustular psoriasis. *J. Hong[1,2], T. Wang[2], T. Tsai[2].* 1) The Graduate Institute of Medical Genomics and Proteomics, National Taiwan University Hospital and College of medicine, Taipei, Taiwan; 2) Department of Dermatology, National Taiwan University Hospital and College of medicine, Taipei, Taiwan.

Pustular psoriasis is a special form of psoriasis characterized by the occurrence of sterile pustules on normal-appearing or inflamed erythematous skin. It can be further classified into several clinical features depending on the clinical course and location: generalized pustular psoriasis (GPP), annular pustular psoriasis (APP), plaque-type psoriasis with pustules (PsOp), impetigo herpetiformis (IH), erythrodermic psoriasis (EP) and localized pustular psoriasis—palmoplantar pustulosis (PPP), and acrodermatitis continua of Hallopeau (ACH). The previous studies showed that GPP is linked to *IL36RN* mutation. However, the percentage of pustular psoriasis patients with *IL36RN* mutation vary among different studies. The variation of mutation frequency might be related to the classification of subtypes of psoriasis. To investigate the frequency of *IL36RN* mutation in different subtypes of pustular psoriasis, we enrolled 57 Han Chinese patients with pustular psoriasis, including 10 ACH and 17 GPP patients. The blood samples were collected and genomic DNA was extracted from leukocytes, and the PCR-based Sanger sequencing was used to analyze the coding exons and flanking introns of the *IL36RN* gene. The results revealed that patients with ACH and GPP had the highest mutation rates of *IL36RN*, which were about 90. 0% and 76. 5% respectively. The splicing mutation, c. 115+6T>C, was the most common variant in ACH and GPP patients with *IL36RN* mutations. For APP, the mutation rate was 40%. In comparison, the *IL36RN* mutation rates of PPP, PsOP, and EP were only about 22%, 14%, and 0% respectively. In summary, a much higher *IL36RN* mutation rate was found in some particular types of pustular psoriasis, ACH and GPP. In view of the rarity of pustular psoriasis, cooperative studies will be needed to screen for other mutations for pustular psoriasis patients without *IL36RN* mutation.

2414T

Identification of a heterozygous stop-gain mutation in *ASXL3* refines autism spectrum disorder diagnosis to Bainbridge-Ropers syndrome. *J. C. Jacobsen[1], B. Swan[1], J. Taylor[2], R. Hill[2], K. Lehnert[1], R. Snell[1].* 1) The University of Auckland, Auckland 1010, New Zealand; 2) Auckland City Hospital, Auckland 1142, New Zealand.

We have identified a cohort of New Zealand females with apparently sporadic Autism Spectrum Disorder (ASD) from the Minds for Minds database#. Exome sequencing of 6 trios from this cohort was performed in order toestablishetiologic genetic variations. In one trio we identified a *de novo* heterozygous four base-pair duplication (c. 1490_1493dup). This results in a frameshift, inducing a premature stop codon (p. NL498K*) in the gene *ASXL3* (Additional Sex Combs Like 3). Five previous observations of heterozygous stop-gain mutations in *ASXL3* have been associated with Bainbridge-Ropers syndrome (BRS). These mutations occur in a disease-causing mutation hotspot in the 5' end of exon 11. The clinical features of BRS include global developmental delay, intellectual impairment, feeding difficulties in infancy, growth retardation and craniofacial abnormalities, such as highly arched eyebrows, anteverted nares, high arched palate and abnormal position of the hands with ulnar deviation. The female proband in our cohort had many of these features including global developmental delay, feeding difficulties, growth retardation and similar craniofacial features, although she lacked ulnar deviation of the hands. The ASXL genes are putative polycomb-group proteins involved in transcriptional regulation during development. Additional *ASXL3* truncating variations have been observed in the population, suggesting that the condition may be caused by dominant gain of function rather than hemizygosity. The mechanism of *ASXL3* truncation mutations in BRS has yet to be established and is the focus of ongoing work. This research has utilized whole-exome sequencing to refine a diagnosis of ASD to BRS, vouching for the use of next generation sequencing for the diagnosis of neurodevelopmental disorders. #Minds for Minds is a New Zealand based organization focused on researching ASD and other neurodevelopmental disorders. We currently have a database of more than 1200 individuals, including 700 New Zealanders diagnosed with ASD. The database reflects the male:female bias in ASD, and the self-attributed ethnic background reflects New Zealand's population.

2415F

Focusing from Chronic Kidney Disease to Cystinosis: Discrepancy Between the Clinical Incidence of Cystinosis and Estimates Based on 1000 Genomes. *W. Wu, L. Stark.* Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT.

Background: Chronic Kidney Disease (CKD) is a complex trait with many contributory causes of which cystinosis is one. Cystinosis is an autosomal recessive metabolic disease characterized by accumulation of cystine. The primary type is the nephropathic form. Patients typically suffer from proximal tubule dysfunction which leads to glycosuria, phosphaturia, aminoaciduria, sodium wasting, polyuria and metabolic acidosis. These may result in growth retardation with rickets, and acute dehydration episodes. The glomerular filtration rate declines gradually and eventually progresses to end-stage renal disease. Extra-renal manifestations are commonly seen in other organs, too. The causative gene of cystinosis has been mapped to CTNS, a gene located on chromosome 17p13. CTNS encodes cystinosin, which is a lysosomal membrane protein responsible for cystine transport. This study explored the allele frequency and estimated the affected and carrier rate of cystinosis in the general population. **Methods:**To identify pathogenic variants, we parsed the Human Gene Mutation Database (HGMD; a manually-curated database) to collect confirmed disease-causing mutations. For identification of variants in the healthy general population, we parsed the 1000 Genomes database (1KG). CTNS variants procured from both databases were intersected and allele frequency was calculated based on the 1KG. Carrier rate (heterozygotes with a pathogenic variant) and affected rate (homozygotes or compound heterozygotes with pathogenic variants) were estimated based on Hardy-Weinberg equilibrium. **Results:**There are 75 pathogenic CTNS variants in the HGMD. These variants include missense mutations, nonsense mutations, insertions, deletions, and complex substitutions. Healthy individuals in the 1KG carry 444 CTNS variants. When we intersected the two databases, we found 36 people in the 1KG who carry pathogenic CTNS variants in a heterozygote state. From these data we calculated a frequency of 0. 016 for all disease-causing alleles. Therefore, the predicted affected rate would be 0. 00027 (approximately 1:3680), and the predicted carrier rate would be 0. 032 (approximately 1:30). **Discussion:**The clinical incidence of cystinosis is nearly 6 per 100,000. However, our estimate based on 1KG is approximately 24 times higher. Possible explanations for this large difference include under-diagnosis of cystinosis, or low penetrance of CTNS variants. Distinguishing between the two will require further investigation.

2416W

Phenotype of a 23 year old male with *FAM111B* mutation: poikiloderma, myopathy, joint contractures, and lymphedema. *M. Ganapathi[1], PL. Nagy[1], C. Ulane[2], U. Agbim[3], J. Odel[4], JB. Sampson[2].* 1) Pathology & Cell Biology, Personalized Genomic Medicine, Columbia University Medical Center, New York, NY; 2) Department of Neurology, Columbia University Medical Center, New York, NY; 3) Department of Medicine, Columbia University Medical Center, New York, NY; 4) Department of Ophthalmology, Columbia University Medical Center, New York, NY.

FAM111B mutations were first described in 8 individuals with autosomal dominant hereditary fibrosing poikiloderma with tendon contracture, myopathy, and pulmonary fibrosis (Am J Hum Genet. 93 (6):1100-7, 2013). This case broadens the phenotype previously described. Clinical presentation: At age 6 months, the patient developed skin rash and hair loss. His walking milestones were normal and comparable to his fraternal twin, but he was noted to have abnormal gait starting at that time. He is ambulatory, but his proximal and distal weakness has been progressive. Pancreatic insufficiency was diagnosed at age 14yrs. He is cognitively normal. At age 23 he presented to our institution. He has poikiloderma on the face,and sun exposed areas, lymphedema in the limbs, hypohidrosis, hypotrichosis, alopecia and nail dysplasia. Muscle strength is reduced in the biceps (3/5 MRC scale), triceps (3/5), wrist extension (4/5) and hip flexor muscles (4/5). He has bilateral elbow contractures and dysphagia. Ophthalmologic examination showed shallow orbits with mild restriction of medial rectus action OU and right macular pigmentary changes. Serum creatine kinase levels were elevated (312-372 U/L). EMG studies revealed normal nerve conductions and absence of myotonic or myopathic changes. Muscle biopsy showed dystrophic changes, with both atrophic and hypertrophic fibers, central nuclei, endomysial fibrosis and fibroadipose replacement. Abdominal CT showed atrophy of paraspinal muscles, rectus abdominis muscles, and pancreatic atrophy but no fibrosis. Liver function tests showed elevated SGOT (26-100 U/L), SGPT (25-132 U/L), alkaline phosphatase (79-129 U/L) and gamma glutamyl transferase (106 U/L), however these resolved. Pulmonary forced vital capacity testing was 64% of predicted but CT scans show no evidence of pulmonary fibrosis. The patient was originally diagnosed in childhood with Rothmund Thompson syndrome, a close mimic of FAM111B. RECQL4 sequencing at that time failed to detect pathogenic variants. Whole exome sequencing on the proband and parents revealed a heterozygous de novo missense variant (c. 1883G>A; p. S628N) in the *FAM111B* gene. This mutation (p. S628N) was previously reported in 1 patient presenting with congenital poikiloderma of the face and extremities, diffuse alopecia, anhidrosis, lymphedema of the extremities and progressive muscle weakness (Am J Hum Genet. 93(6):1100-7, 2013). This individual consented to photography under our IRB-approved protocol AAAK2000.

2417T

Diagnostic difficulties in a newborn with rhizomelic dwarfism, multiple intrauterine fractures and dysmorphic features due to mucolipidosis type II (MLD II) due to compound heterozygote *GNPTAB* mutations and an unreported heterozygote change in the *SERPINF1* gene. *Y. Lacassie[1], K. Wood[2], R. M. Zambrano[1], B. J. Cheek[3], C. Arcement[4], M. Haymon[4], J. Steinkampf[1], S. Sampath[5], J. C. Hyland[6].* 1) Department of Pediatrics, LSUHSC and Children's Hospital, New Orleans, LA; 2) Department of Genetics, LSUHSC New Orleans, LA; 3) Department of Pathology, LSUHSC New Orleans, LA; 4) Department of Radiology, Children's Hospital New Orleans, LA; 5) Prevention Genetics, Marshfield, WI; 6) Connective Tissue Gene Tests, Allentown, PA.

We report a NB with a history of polyhydramnios, severe IUGR and rhizomelic dwarfism with bowed femurs and poorly calcified fibulas. Rhizomelic Chondrodysplasia Punctata Type I was suspected by MFM; however aCGH and mutation analysis for *PEX7* were normal. Hypotonia, weak cry, hypertelorism, rotated ears and rhizomelic limbs with camptodactyly were noticed at birth. Babygram showed possible fractures raising diagnosis of OI. Genetics ruled out any of the classical types of OI, but the possibility of OI type VII was mentioned. Family evaluation revealed that the 37yo AA father is tall, macrocephalic, has telecanthus and thumb subluxation. The mother is a non-consanguineous, G5P4L3, 31yo AA with a history of ovarian cysts, hypertelorism and long fingers with extra ffc. The proband was born via vaginal induction at 366/7 WGA. BW and BL at 3rd centile, HC at 10th centile. Telecanthus, anteverted nostrils, short columella, micrognathia, long fingers and toes with clinodactyly, camptodactyly and overriding fingers, extra crease R2-3F, prominent left interdigital areas and vertical main line A at left were observed. NGS panel for AR OI was requested. However, scannogram showed marked coarsening of trabecula with periosteal cloaking, metaphyseal cupping, flaring and irregularity with bending metaphyseal fractures, diaphyseal expansion in the bones of the hand with proximal and distal constriction, calcific stippling of both calcaneus and other skeletal abnormalities suggestive of dysostosis multiplex, specifically MLD type II (I-Cell Disease). Blood smear and MLD II/III plasma screen were consistent. *GNPTAB* gene sequencing showed compound heterozygosity for a pathogenic sequence variant (c. 1399delG; p. Asp467Ilefs*33) reported in MLD II, and also for an unreported but expected pathogenic sequence variant (c. 1905_1908delAAGG; p. Glu637Aspfs*4.). Both variants are predicted to result in a frameshift and premature termination of the protein N-acetylglucosamine-1-phosphotransferase subunits alpha/beta. MLD II is a rare AR disorder with a phenotype resembling Hurler syndrome and usually appearing after 6-12 months of age. Interestingly, the NGS panel for recessive OI revealed an extremely rare heterozygote variant (c. 1004A>G transition) in exon 8 of the *SERPINF1* gene. Parents are planning to be tested to determine the role of these variants. Complex interactions between different genes might explain the unusual and severe phenotype observed in this newborn.

2418F

Clinical heterogeneity of genetically confirmed six patients with Vici syndrome. *I. Hori[1], F. Miya[2], M. Nakashima[3], Y. Negishi[1], H. Shiraishi[4], Y. Nonoda[5], J. Tohyama[6], N. Okamoto[7], A. Hattori[1], N. Ando[1], M. Kato[8], T. Tsunoda[2], I. Nishino[9], H. Saitsu[3], Y. Kanemura[10], N. Matsumoto[3], K. Kosaki[11], S. Saitoh[1].* 1) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 2) Center for Integrative Medical Sciences, RIKEN, Yokohama, Japan; 3) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 4) Department of Pediatrics, Hokkaido University Graduate School of Medicine, Sapporo, Japan; 5) Department of Pediatrics, Kitasato University School of Medicine, Sagamihara, Japan; 6) Department of Pediatrics, Epilepsy Center, Nishi-Niigata Chuo National Hospital, Niigata Japan; 7) Department of Medical Genetics, Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan; 8) Department of Pediatrics Showa University School of Medicine, Tokyo, Japan; 9) Department of Neuromuscular Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo, Japan; 10) Division of Regenerative Medicine, Institute for Clinical Research, Osaka National Hospital, National Hospital Organization, Osaka, Japan; 11) Center for Medical Genetics, Keio University School of Medicine, Tokyo, Japan.

Vici syndrome (OMIM #242840) is a rare autosomal recessively inherited multisystem disorder characterized by agenesis of the corpus callosum, cataracts, cardiomyopathy, combined immunodeficiency, psychomotor delay, and hypopigmentation. Mutations in the gene *EPG5* gene in chromosome 18q12. 3 were recently shown to be causative for Vici syndrome. *EPG5* is the human homolog of the metazoan-specific autophagy gene epg-5, encoding a key autophagy regulator (ectopic P-granules autophagy protein 5) implicated in the formation of autolysosomes. To better delineate the clinicopathological features of Vici syndrome, we studied 6 patients from 4 families with genetically confirmed Vici syndrome. Three patients were male. All parents were non-consanguineous. Muscle biopsy was performed in 4 patients. We performed whole exome sequencing to four probands from respective families, and identified mutations were validated in all six patients as well as their parents by Sanger sequencing. Compound heterozygous truncating mutations were confirmed in all six patients, and parents were revealed to be carriers. Common features of six patients included developmental delay, hypotonia, seizures, recurrent infection, agenesis of the corpus callosum, and elevated liver enzyme. No patients had cataracts. Only two patients had cardiomyopathy based on the specific findings on cardiac ultrasonography. Although all patients had the suspicion of an associated immunodeficiency based on clinical findings of an increased number of infection, there were few immunological abnormalities, which are consistent with previous reports. Additional central nervous system abnormalities included polymicrogyria, cortical developmental malformation, paucity of white matter, irregularity of ventricular wall, ventricular dilation, and delayed myelination. Pathologic findings in muscle revealed no specific features. It is likely that the incidence of cataracts and cardiomyopathy would be low in early course of Vici syndrome. When patients have agenesis of the corpus callosum and severe developmental delay of unknown etiology, we need to consider the possibility of Vici syndrome and the molecular test should be applied even full-blown clinical features are not present.

2419W

Juvenile Xanthogranuloma in Noonan Syndrome. *M. M. Ali[1], A. E. Gilliam[2], A. C. Gilliam[3], K. A. Rauen[4].* 1) Pediatrics, UCSF, San Francisco, CA; 2) Dermatology, UCSF, San Francisco, CA; 3) Dermatology, Stanford University, Palo Alto, CA; 4) Pediatrics, UC Davis, Sacramento, CA.

Noonan syndrome (NS) is an autosomal dominant RASopathy that affects approximately 1 in 1,000-2,000 individuals. NS exhibits a widely variable clinical phenotype, however, it is characterized by distinctive craniofacial features, congenital cardiac defects, reduced growth, bleeding disorders and a variable degree of neurocognitive delay. One other phenotypic feature which assists in the clinical diagnosis of NS is its cutaneous manifestations whereby individuals may have scattered café au lait macules, junctional nevi, lentigines and keratosis pilaris. We describe a 10 month old female with the clinical diagnosis of NS who was confirmed molecularly by the identification of a heterozygous SHP2 p. Y62D missense mutation [*PTPN11* gene]. Physical exam was consistent with features of NS, but dermatologic evaluation revealed yellowish-red nodules consistent with juvenile xanthogranuloma (JXG). Biopsy with histopathologic evaluation confirmed the diagnosis of JXG. A RASopathy that is known to be associated with JXG is Neurofibromatosis Type 1 (NF1). NF1 individuals who are diagnosed with JXGs are thought to have an elevated risk of juvenile myelomonocytic leukemia (JMML). JMML is a myelodysplastic and myeloproliferative disorder that accounts for 1–2% of childhood leukemias in the United States. A JMML-like phenotype with spontaneous regression in infants with NS has been previously reported, with an estimate of the incidence being approximately 10% in patients with NS. The association of JXG, JMML and NF1 has been described before, however, this association is still controversial. Children with both NF1 and JXG have been estimated to have a 20–32 times higher risk of developing JMML than do those with NF1 alone. In understanding the common underlying pathogenetic dysregulation of the Ras/MAPK pathway in both NS and NF1, there may be a possible similar relationship between patients with NS and JXG. To our knowledge, this report represents the first case describing a NS individual with JXG, highlighting the importance of carefully examining all RASopathy individuals for these skin lesions.

2420T

GNE Myopathy in the Non-Jewish Population, a Genotype-Phenotype Correlation Study. *A. Haghighi[1], S. Nafissi[2], A. Qurashi[3], Z. Tan[3], H. Shamshiri[2], Y. Nilipour[4], A. Haghighi[5], R. J. Desnick[3], R. Kornreich[3].* 1) Harvard Medical School, Boston, MA; 2) Department of Neurology, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran; 3) 4 Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA; 4) Department of Pathology, Pediatric Pathology Research Center, Mofid Children Hospital, Shahid Beheshti Medical University, Tehran, Iran; 5) Toronto General Hospital, University of Toronto, Toronto, Canada.

GNE myopathy, also known as hereditary inclusion body myopathy, is characterized by progressive muscle atrophy and weakness, initially involving the distal muscles. The disease usually spares the quadriceps muscles. This autosomal recessive disease is caused by mutations in *GNE* gene. In this study, 18 non-Jewish patients from 11 Iranian families were comprehensively investigated. Genetic analysis identified three homozygous *GNE* mutations: c. 2228T>C (p. M743T) in 7, c. 830G>A (p. R277Q) in 2, and one novel variation (c. 804G>A) in 2 families. The latter is a novel mutation that results in a synonymous codon change (p. L268=) and is predicted to cause abnormal splicing. The youngest patient was an 18-year-old girl (with c. 830G>A, p. R277Q), whereas the oldest onset age (31 years) was seen in a male patient with c. 804G>A (p. L268=). Three of our p. M743T patients who were more than 6 years after the onset used wheelchairs, but none of them was wheelchair bound, even 17 years after the GNE myopathy onset. Our results confirmed that, however, p. M743T, in the kinase domain, is the most common disease causing variant in non-Jewish Iranian patients, other mutations, epimerase domain, can also cause the disease in this population. This study expands the knowledge on the phenotype and molecular genetic heterogeneity of the GNE myopathy.

2421F

A novel IRF6 mutation in female affected members of an Asian family with Van der Woude syndrome. *E. C. Tan[1], E. C. P. Lim[1], H. W. Lim[1], S. T. Lee[2].* 1) Research Laboratory, KK Women's & Children's Hospital, Singapore; 2) 3Department of Plastic, Reconstructive & Aesthetic Surgery, Singapore General Hospital, Singapore.

We describe a previously unreported IRF6 variant found in a pedigree that has five members with Van der Woude syndrome, the most common among syndromes which include cleft lip and/or cleft palate as one of the presentations. Mutations in the interferon regulatory factor 6 (IRF6) gene which encodes a transcription factor with a DNA-binding domain and a protein-binding domain have been identified in many VWS patients. The gene is thought to play a key role in regulating proliferation and differentiation during epidermal development. The proband is an 19-year old girl with bilateral cleft lip and palate and lip pits. She is the eldest child of non-consanguineous parents. Her mother is of Malay ancestry while her father was from Bangladesh in the Indian subcontinent. Her mother also presents with bilateral cleft lip and palate and lip pits. She has 2 affected sisters and an unaffected brother. The only other affected family member is a maternal aunt. None of the male members of the pedigree has the VWS phenotype. Peripheral blood was obtained from the proband, the parents and one of the affected sisters. Sequence analysis targeting coding exons of IRF6 uncovered a single nucleotide variant in Exon 3, which changes the amino acid encoded from an isoleucine to a threonine. This variant has not been reported and is not in any database. As the missense substitution occurs in the DNA-binding domain, it is likely to affect its binding with its target DNA site. Targeted testing performed on the family samples revealed the presence of the same variant in her affected mother and her affected sister but not in her unaffected father. Testing on saliva DNA of her other affected sister and her unaffected brother showed complete co-segregation of the mutation with the VWS phenotype. The evidence is strong that the identified mutation is causative of the phenotype in this pedigree given the variant's segregation with the VWS phenotype, and the substitution of a hydrophobic by a polar residue in the DNA-binding domain. One interesting feature is that none of the male members of the pedigree has the VWS phenotype.

2422W

Sengers syndrome: clinical investogation and identifying novel *AGK* mutations. *H. Haghighi-Kakhki[1], A. Haghighi[2], T. B. Haack[3], M. Atiq[4], H. Mottaghi[5], R. A. Bashir[6], U. Ahting[3], R. G. Feichtinger[7], J. A. Mayr[7], A. Rötig[8], A. Lebre[8], T. Klopstock[9], A. Dworschak[10], N. Pulido[11], M. A. Saeed[6], N. Saleh-Gohari[12], E. Holzerova[3], R. F. Chinnery[13], R. W. Taylor[14], H. Prokisch[3].* 1) School of Medicine, Mashhad Azad Unievrsity, Mashhad, Iran; 2) Department of Genetics, Harvard Medical School, Boston, MA, USA; 3) Institute of Human Genetics, Helmholtz Zentrum München, Neuherberg, München, Germany; 4) Department of Pediatrics, Aga Khan University, Karachi, Pakistan; 5) Department of Pediatrics, Imam Reza Hospital, Mashhad University of Medical Sciences, Mashhad, Iran; 6) Department of Paediatrics and Neonatology, Ahmadi Hospital, Kuwait Oil Company, Al Ahmadi, Kuwait; 7) Department of Paediatrics, Paracelsus Medical University, Salzburg, Austria; 8) Inserm UMR 1163, Imagine Institute, Paris Descartes University, Paris France; 9) Department of Neurology, Friedrich-Baur-Institute, Ludwig-Maximilians-Universitét München, Munich, Germany; 10) Department of Pediatric Cardiology, University Hospital, Aachen University of Technology, Aachen, Germany; 11) Pediatric Department, Hospital Dr. Gustavo Fricke, Viña del Mar, Chile; 12) Deptartment of Genetics, Medical School, Kerman University of Medical Sciences, Kerman, Iran; 13) Wellcome Trust Centre for Mitochondrial Research, Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK; 14) Wellcome Trust Centre for Mitochondrial Research, Institute of Ageing and Health, Newcastle University, Newcastle upon Tyne, UK.

Sengers syndrome is a rare autosomal recessive disease. The patients have characterized congenital cataract, cardiomyopathy, skeletal myopathy and lactic acidosis. Mutations in the acylglycerol kinase (*AGK*) gene can cause Sengers syndrome. We studied the clinical features of the disease and genetic variants of *AGK* in seven new families with Sengers syndrome from around the world. Congenital cataract and cardiomyopathy were found d in all patients. Lactic acidosis and tachydyspnoea were the other common findings in the patients. Mutational analysis identified disease-causing *AGK* mutations in all patients, including six novel variants: homozygous c. 523_524delAT (p. Ile175Tyrfs*2), c. 424-1G>A (splice site), c. 409C>T (p. Arg137*) and c. 877+3G>T (splice site), and compound heterozygous c. 871C>T (p. Gln291*) and c. 1035dup (p. Ile346Tyrfs*39). The findings of our study expand the knowledge on genotype-phenotype correlations in Sengers syndrome and will have important implications for genetic screening and diagnosis.

2423T

Primordial Dwarfism associated novel *POC1A* mutation reveals new insights for centriole biogenesis. *M. Ozen[1,2,3], A. Koparir[1], O. F. Karatas[4], A. Gezdirici[5], B. Yuceturk[6], B. Yuksel[7], A. O. Bayrak[6], O. F. Gerdan[6], M. S. Sagiroglu[6], K. Kirimtay[8], E. Selcuk[8], A. Karabay[8], C. J. Creighton[9], A. Yuksel[3].* 1) Department of Medical Genetics, Istanbul University, Cerrahpasa Medical School, Istanbul, Turkey; 2) Department of Pathology & Immunology, Baylor College of Medicine, Houston, TX, 77030, USA; 3) Biruni University, Department of Molecular Biology and Genetics, Topkapi, Istanbul, Turkey; 4) Molecular Biology and Genetics Department, Erzurum Technical University, Erzurum, Turkey; 5) Department of Medical Genetics, Kanuni Sultan Suleyman Training and Research Hospital, 34303, Istanbul, TURKEY; 6) Advanced Genomics and Bioinformatics Research Center (IGBAM), BÝLGEM, TUBITAK, Kocaeli, Turkey; 7) Genetic Engineering and Biotechnology Institute, TUBITAK Marmara Research Center, Kocaeli, Turkey; 8) Department of Molecular Biology and Genetics, Istanbul Technical University, Istanbul, Turkey; 9) Department of Medicine and Dan L Duncan Cancer Center, Division of Biostatistics Baylor College of Medicine, Houston, TX, 77030, USA.

POC1A encodes a WD repeat protein localizing to centrioles and spindle poles and is associated with short stature, onychodysplasia, facial dysmorphism and hypotrichosis (SOFT) syndrome. These main features are related with the defect in cell proliferation of chondrocytes in growth plate, which were observed in a mouse model in recent study. In the current study, we report two patients with primordial dwarfism (PD) in a single family. We utilized Whole Exome Sequencing (WES) in the patients. A novel homozygous p. T120A missense mutation was detected in *POC1A* in both patients, a known causative gene of SOFT syndrome, and confirmed using Sanger sequencing. To confirm the pathogenicity of the detected mutation, primary fibroblast cultures obtained from the patients and a control individual were used. To further evaluate the global gene expression profile of cells carrying p. T120A mutation in *POC1A*, we performed gene expression array analysis which showed that 4800 transcript probes were significantly deregulated in cells with p. T120A mutation in comparison to the control. GO term association results showed that deregulated genes are mostly involved in extracellular matrix and cytoskeleton. Furthermore, the p. T120A missense mutation in *POC1A* caused formation of abnormal mitotic spindle structure, including supernumerary centrosomes, and changes in POC1A was accompanied by alterations in another centrosome associated WD repeat protein p80-katanin. As mutations in p80-katanin gene have been previously reported, which also can cause mitotic spindle abnormalities similar to those we associate with POC1A p. T120A mutation, further studies into the functional relationships involving POC1A and p80-katanin are warranted.

2424F

Rare familial *TSC2* gene mutation associated with atypical phenotype of tuberous sclerosis complex. *S. Ben-Shachar[1, 2], J. Fox[2], S. Oliel[1,2], R. Svirski[1,2], H. Saitsu[3], N. Matsumoto[3], A. Fattal-Valevski[1,2].* 1) Tel-Aviv Medical Center, Tel-Aviv, Israel; 2) Sackler Faculty of Medicine, Tel-Aviv University, Tel Aviv Israel; 3) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan.

Tuberous sclerosis complex (TSC) is a neurocutaneous disorder that results from mutations of either the *TSC1* gene or the *TSC2* gene. Diagnosis is based on well-established clinical criteria or molecular analysis. We describe an 18-month-old boy with seizures and a single hypopigmented macule which did not meet consensus criteria for the diagnosis of TSC. Whole exome sequencing revealed a heterozygous *TSC2* mutation (c. 5138G>A (p. Arg1713His)) in the patient, his mother as well as several other family members on the maternal side. The mother and other family members with the mutation were asymptomatic except for the presence of hypomelanotic macules. The phenotypic characteristics of the individuals in this family were not indicative of a *TSC2* mutation since none satisfied the criteria for even suggesting a diagnosis of possible TSC. This case provides evidence of a unique *TSC2* mutation that resulted in an atypical clinical presentation and indicates potential shortcomings of the current diagnostic criteria for TSC. These findings may have implications for genetic counseling and screening.

2425W

X-Linked Congenital Disorder of Glycosylation in a female infant due to *SLC35A2* gene mutation - Expanding the clinical phenotype. *A. Pandya[1], B. Lincoln-Boyea[1], R. S. Greenwood[2].* 1) Pediatric Genetics and Metabolism, University of North Carolina, Chapel Hill, CHAPEL HILL, NC; 2) Child Neurology, Department of Neurology, UNC, Chapel Hill,.

Congenital Disorders of Glycosylation (CDG) are a heterogeneous group of neurometabolic diseases, caused mainly due to defects in N-linked glycosylation. Although the majority of these are inherited as autosomal recessive disorders, several X-linked forms of CDG have been identified recently. Since 2013, mutations in three genes, *ALG13, SLC35A2* and *SSR4* have been reported to cause X- Linked form of CDG. The *SLC35A2* gene encodes the only known Golgi localized UDP-galactose transporter, mutations in which decrease the UDP-galactose transport, causing galactose deficient glycoproteins. Interestingly, of the six patients identified with mutations in *SLC35A2,* four were female patients, and two males, the latter demonstrating somatic mosaicism. We report a female infant who presented with hypotonia, motor delay and distinct facial features. The MRI was notable for diffuse atrophy and decreased white matter. At age nine months she had an abnormal transferrin isoelectric study highly suggestive of CDG, based on unusual transferrin glycated variants, reflecting a loss of sialic and galactose moieties. Testing of *SLC35A2* gene was positive for a heterozygous c. 991 G > A (V331I) variant in exon 4, determined to be pathogenic. The child has distinct facial features, no seizures and unique skin pigmentary changes following the lines of Blaschko. This is the first report of pigmentary changes noted in a patient with *SLC35A4* mutation, which can be used as an indicator for the presence of genetic mosaicism. This observation provides an opportunity to expand the clinical phenotype of this recently identified X- linked form of CDG.

2426T

Evaluation of Growth Parameters in the LCRC Dataset Demonstrates Differences Based on Osteogenesis Imperfecta Subtype for All Age Groups. *A. Tam[1], M. Jain[1], D. Cuthbertson[2], J. R. Shapiro[3], R. D. Steiner[4, 5], P. A. Smith[6], M. B. Bober[7], T. Hart[8], J. Krischer[2], P. H. Byers[9], M. Pepin[9], M. Durigova[10], F. H. Glorieux[10], F. Rauch[10], B. Lee[1], V. R. Sutton[1], S. C. S. Nagamani[1].* 1) Baylor College of Medicine, Houston, TX, USA; 2) University of South Florida, Tampa, FL, USA; 3) Kennedy Krieger Institute, Baltimore, MD, USA; 4) Oregon Health & Science University, Portland, OR, USA; 5) University of Wisconsin, Marshfield, WI, USA; 6) Shriners Hospitals for Children, Chicago, IL, USA; 7) Alfred I. DuPont Hospital for Children, Wilmington, DE, USA; 8) OI foundation, Gaithersburg, MD, USA; 9) University of Washington, Seattle, WA, USA; 10) Shriners Hospital for Children and McGill University, Montreal, QC, Canada.

Osteogenesis imperfecta (OI) predisposes to recurrent fractures and bone deformities, factors that affect bone growth. There is a lack of large-scale studies that investigate growth parameters, including height and weight in patients with OI. The Linked Clinical Research Centers (LCRC) is a network of five centers established to foster clinical research and advance the care of patients with OI across North America. Using data from the LCRC, we analyzed height, weight, and arm-span to height ratio in 552 individuals with OI (type I n=244; Type III OI n=110; type IV OI n=150). Chi square analysis was used to analyze for differences in proportion of individuals below the third percentile of CDC growth curves and unpaired t-test was used to compare growth parameters as continuous variables. As expected, subjects with type III OI had an increased probability of being below the third percentile for height and weight as compared to those with type I, and IV OI ($p<0.0001$). Whereas the arm-span to height ratio was not different between the various subtypes prior to age 10, increased ratio was observed in subjects with OI type III compared to OI type I older than 10 years ($p<0.05$). The final adult height, as expected was in the order OI type I >OI type IV> OI type III. As many factors contribute to adult height including familial stature, we further examined individuals without a parental history of OI by calculating mid-parental height, and compared this to the actual adult height. Our results suggest that final adult height is affected across OI subtypes, with 48% of individuals with type I OI, 58% of type IV OI and 100% of type III OI being below the predicted third percentile for mid-parental height (OI type III vs type I, $p<0.0005$; OI type IV vs type I, $p<0.01$). Our data support expected growth patterns for OI subtypes and will be utilized to examine growth over time in a longitudinal fashion, with the hopes of defining age-expected distributions for height and weight for OI subtypes. Characterization of growth patterns in OI subtypes will help clinicians determine whether patients with OI are following expected growth based on their OI subtype.

2427F

Behavioral Phenotype of Muenke Syndrome (FGFR3-Related Craniosynostosis). *Y. Addissie, C. Yarnell, P. Kruszka, M. Guillen, R. Hart, N. Agochukwu, E. Wiggs, D. Hadley, M. Muenke.* Medical Genetics Branch, National Human Genome Research Institute, Bethesda, MD.

Background: Muenke syndrome is the most common craniosynostosis syndrome defined by the presence of the p. Pro250Arg mutation in the *FGFR3* gene and is clinically characterized by craniosynostosis, hearing loss, developmental delay, and carpal and tarsal fusions. Previous research suggests that children with craniosynostosis syndromes have an increased risk of developing behavioral problems, yet there has been no research to date focusing solely on individuals with Muenke syndrome. This study investigates executive function and adaptive behavior in persons with Muenke syndrome using validated instruments with a normative population and unaffected siblings as controls. **Study Design:** Participants in a cross sectional study included individuals with Muenke syndrome (P250R mutation in *FGFR3*) and their mutation negative siblings. Participants completed validated assessments of executive functioning (Behavior Rating Inventory of Executive Function; BRIEF) and adaptive behavior skills (Adaptive Behavior Assessment System; ABAS-II). **Results:** Forty-four *FGFR3* mutation positive individuals, median age 9 years, range 7 months to 52 years were enrolled. Additionally, 10 unaffected siblings were used as controls (5 males, 5 females, median age of 13 years, range 3 to 18 years). For the General Executive Composite scale of the BRIEF, 32. 1% of the cohort had scores greater than +1. 5 SD, signifying "Potential Clinical Significance. " For the General Adaptive Composite of the ABAS-II, 28. 2% of affected individuals scored in the 3rd -8th percentile of the normative population and 56. 4% were below the "Average" category (less than the 25th percentile). Multiple regression analysis showed that craniosynostosis was not a predictor of BRIEF (P = 0. 7) and ABAS-II scores (P = 0. 7). Other factors includingsurgery for craniosynostosis, hearing loss, and history of seizureswere also not predictive. In the sibling pair analysis, affected siblings performed significantly poorer in the BRIEF General Executive Composited and the ABAS-II General Adaptive Composite. **Conclusion:** Individuals with Muenke syndrome are at an increased risk for developing adaptive and executive function behavioral changes when compared to a normative population and unaffected siblings, regardless of whether craniosynostosis is present.

2428W

Böök syndrome (Premolar aplasia, Hyperhidrosis, and Canities prematura) is an Autosomal Dominant Ectodermal Dysplasia with Variable Expression. *M. Hajianpour, C. Evers, p. Trapane.* Pediatrics/ Medical Genetics, University of Iowa, Iowa City, IA.

Proband is a Three-year-old female with a history of **slow hair growth**, and severe **dental caries**. She has excessive sweating (**hyperhidrosis**). Her height and weight are just above97th percentile, and herhead size isat 70thpercentile for age (parents and sibling have large built). Her development is normal. **Proband's mother**has a history of **slow hair growth**with **premature hair graying** at age 15 years and **hyperhidrosis**. Maternal grandfather and his mother, and maternaluncle alsohad premature hair graying. Maternal grandfather and his motherhaddental caries. A maternal first cousin,and several of mother's second cousins have **missing teeth** (more than six). It is not clear which teeth are missing. Böök (1950) reported 25 persons in 4 generations of a Swedish family with **P**remolar aplasia, **H**yperhidrosis, and **C**anities prematura (premature graying of hair), anddesignated the disorder**"PHC"** syndrome. Inheritance of the disorder in this family was clearly **autosomal dominant** with high penetrance. Salinas et al. (1992) described "**a new ectodermal dysplasia syndrome**" in an 18-year-old Caucasian woman with congenitally missing premolars, narrow palate, severe functional hyperhidrosis of the hands and feet, small hands,hypoplastic nails, disorganized eyebrows, unilateral simian creases,and poorly formed dermatoglyphicsincludingdistal digital creases. The authors suggested that this may be a new case of **Böök** syndrome. However, the patient **lacked premature graying of hair. Ectodermal Dysplasia,**is a group of disorders characterized by abnormal development in two or more ectodermal structures(hair, teeth, nails and sweatglands),with or withoutother systemic findings. Therefore,based on this definition,**Böök syndrome,**is an **ectodermal dysplasia** by involving teeth, hair and sweat glands. Our patient shows severe dental caries, hyperhidrosis, no dental aplasia at age three years, and no canities prematura, which is often seen in teenage period. However, her mother had premature graying of hair and hyperhidrosis, and some other relatives had dental aplasia and canities prematura in three generations, presenting an **autosomal dominant** mode of inheritance with **variable expressivity**. We agree with Salinas et al. (1992) who suggested that "the lack of other reports of Böök syndrome may result from the clinical features being of relatively little consequence to most affected individuals or from symptoms being treated as separate entities by different specialists. ".

2429T

Clinical and molecular spectrum of Noonan syndrome and other RASopathies. *J. Kim[1], G-H. Kim[1], B. Lee[1,2], J-H. Choi[1,2], H-W. Yoo[1,2].* 1) Medical Genetics Center, Asan Medical Center, Seoul, South Korea; 2) Department of Pediatrics, Asan Medical Center Children's Hospital, University of Ulsan College of Medicine, Seoul, South Korea.

Background: Noonan syndrome (NS) and NS-related disorders (cardiofaciocutaneous [CFC], Costello, Legius syndromes, and Noonan syndrome with multiple lentigines [NSML]) share common clinical features characterized by unique facial features, postnatal growth failure, psychomotor retardation, ectodermal abnormalities, congenital heart diseases, chest and skeletal deformities, and delayed puberty. To date, this group of disorders has been known to be caused by germ-line mutations of at least 13 genes of the Ras-mitogen-activated protein kinase pathway (*PTPN11, SOS1, RAF1, SHOC2, BRAF, KRAS, NRAS, CBL, HRAS, MEK1, MEK2, SPRED1*, and *RIT1*), mostly with gain of function. **Purpose:** This study was performed to investigate mutation spectrum and phenotype of Noonan syndrome and its related disorders. **Methods:** Clinical characteristics and genotypes of 10 previously known (*PTPN11, SOS1, RAF1, SHOC2, BRAF, KRAS, NRAS, HRAS, MEK1, MEK2*) and two candidate genes, *SPRY1-4* and *SPRED1*, were investigated in 175 patients with NS, 25 with CFC syndrome, 12 with Costello syndrome (CS), and 7 with NSML. **Results:** *PTPN11* (41. 1%), *SOS1* (13. 1%), *RAF1* (6. 3%), *BRAF* (3. 4%),*KRAS* (2. 9%), *SHOC2* (1. 1%), and *MEK1* (0. 6%) mutations were identified in NS; *BRAF* (48%), *SHOC2* (24%), and *MEK1* (8%) mutations in CFC; and *HRAS* and *PTPN11* mutations in CS and NSML, respectively. Normal stature was frequently observed in individuals with *SOS1* mutations, hypertrophic cardiomyopathy in *RAF1*and *HRAS*, pulmonic stenosis in *SOS1* and developmental delay in *RAF1*, *BRAF* or *SHOC2* mutations. Cardiac hypertrophy should be carefully evaluated in patients with *RAF1* and *HRAS* mutations. Tumor surveillance is recommended for subjects with *HRAS* mutations. Mutation was not identified in 29. 7% of NS and 20% of CFC syndrome. Whole exome sequencing analysis and array comparative genomic hybridization were selectively performed in patients without mutations by Sanger sequencing of the candidate genes. Novel sequence variants were found in *RUVBL2* and *MYH11*, and haploinsufficiency of *ARID2* was identified. **Conclusions:** This study described the mutation spectrum and identified new candidate genes for RASopathies. Functional studies are needed to verify the impact of sequence variants of novel genes on patients' phenotypes. Molecular genetic testing have a role in the process of differential diagnosis of NS and NS-related disorders because of the constellation of overlapping clinical features of RASopathies.

2430F

Congenital nephrosis, cerebral ventriculomegaly and heterotopias – expanding the phenotype associated with *CRB2* mutations. *A. Slavotinek[1], K. Gray[2], L. Rodan[2], B. P. Chaudhari[3], A. Rajkovic[3], D. Steele[3], R. Vanderwall[3], K. Owens[4], M. Hannibal[4], W-H. Tan[2].* 1) University of California, San Francisco, San Francisco, CA; 2) Boston Children's Hospital, Boston, MA; 3) Center for Medical Genetics and Genomics, Magee-Womens Hospital of University of Pittsburgh Medical Center, Pittsburgh PA; 4) Pediatrics and Communicable Diseases, University of Michigan Medical School, Ann Arbor, MI.

We recently reported five fetuses and a child with a novel phenotype comprising cerebral ventriculomegaly, echogenic kidneys and greatly elevated maternal serum alpha-fetoprotein (MSAFP) and amniotic fluid alpha-fetoprotein (AFAFP; CRB2 syndrome). Renal findings included tubular cysts at the corticomedullary junction, diffuse effacement of the epithelial foot processes and microvillous transformation of the renal podocytes similar to congenital nephrotic syndrome, Finnish type. All cases had deleterious mutations in *CRB2*. Studies in *Danio rerio* have shown that reduced *crb2b* function results in loss of the slit diaphragms of the podocytes. We now provide a further 12 cases with similar findings and an additional child with a single *CRB2* mutation. A 5-day old female had nephrotic syndrome with renal microcysts, massive hydrocephalus, a large atrial septal defect, probable Scimitar syndrome and a Muellerian duct anomaly. A fetus with ventriculomegaly had elevated MSAFP and AFAFP. Testing for *CRB2* mutations in these two cases is pending. Clinical exome sequencing in a 2-year old male with hypotonia, developmental delays, an absent left kidney and a smaller right kidney had a single *CRB2* mutation, p. Arg633Trp that was paternally inherited; coverage of the coding region was high and array showed no deletions in the *CRB2* gene region. We found 10 reported cases consistent with CRB2 syndrome (Reuss et al. , 1989; Jolly et al. , 2003), with additional findings of pericardial effusion (2/10), postaxial polydactyly (1/10), choroid plexus hyperplasia (1/10), duplication of the right kidney/ureter (1/10) and absent lung fissure with incomplete intestinal rotation (1/10). Gray matter heterotopias were prominent in one of the first reported children with *CRB2* mutations. This feature overlaps with Galloway-Mowat syndrome (GMS), characterized by microcephaly, hiatal hernia and nephrotic syndrome, in which heterotopias were described. We found one reported child with periventricular nodular heterotopias, a raised AFAFP and renal echogenicity that was reported as GMS, but probably had CRB2 syndrome (Palm et al. , 1986). We conclude that the phenotype associated with CRB2 mutations overlaps with other conditions with early-onset nephrosis and cerebral malformations, such as GMS. The high AFP measurements and the renal histology may help clinical distinction. Jolly et al. Prenat Diag 2003;23:48; Palm et al. Arch Dis Child 1986;61:545; Reuss et al. , Am J Med Genet 1989;33:385.

2431W

Pseudoarthrosis of the tibia and retroperitoneal neurofibroma in NF1: case report and 4-year follow-up. *M. Aceves-Aceves[1], I. M. Salazar-Dávalos[1], S. A. Alonso-Barragán[2], A. Marín-Medina[2], D. García-Cruz[1,2], N. O. Dávalos[1,2], G. Chiapa[3], M. Puebla-Mora[4], M. Salazar-Páramo[5], I. P. Dávalos[1,2].* 1) Instituto de Genética Humana, DGH, CUCS, Universidad de Guadalajara, Guadalajara, México; 2) Doctorado Genética Humana (DGH), CUCS, Universidad de Guadalajara, CIBO, IMSS, Guadalajara, México; 3) Servicio de Ortopedia y Traumatología, UMAE, HP, CMNO, IMSS, Guadalajara, México; 4) Servicio de Anatomía Patológica, UMAE, HP, CMNO, IMSS, Guadalajara, México; 5) Depto. de Fisiología, CUCS, Universidad de Guadalajara, Div. Investigación, UMAE, HE, CMNO, IMSS, Guadalajara, México.

Introduction: Neurofibromatosis type 1 (NF1), MIM #162200, is an autosomal dominant disorder with an incidence of 1 in 3000 live births. NF1 is caused by mutation in the *NF1* gene on chromosome 17q11. 2, characterised by café-au-lait spots, Lisch nodules, fibromatous tumors of the skin and benign and malignant tumors. Pseudoarthrosis of the tibia (PT) is associated with NF1, 55% of the patients with PT presented NF1 and 6% of NF1 patients develop tibial dysplasia. The affected bone shows insufficient mechanical strength, excessive bone resorption, impaired vascularization and decreasing osteogenic capability. **Objective:**To present a case with NF1 with congenital pseudoarthrosis of the tibia and retroperitoneal neurofibroma. **Case Report:** The *propositus* a boy aged 9 year-old, father (52 years) and sister (17 years) had NF1 with no pseudoarthrosis. Mother (42 years) and a healthy sister (16 years). At physical examination: weight 24. 4 kg (pc 10), height 130 cm (pc 10), head circumference 54 cm (pc 10), thorax with café-au-lait spots (>30), axillary freckling, post surgical scars on right costal region (plexiform neurofibroma), retroperitoneal intraneural neurofibroma and tibial pseudoarthrosis. At 3 months he presented anterolateral deformity of the left tibia, at 4 year-old was resected a plexiform neurofibroma on right costal region, next year presented a fracture of the left tibia secondary to trauma and was surgically treated with transcalcaneal intramedullary nail fixation, with inadequate fracture consolidation. Two years later intramedullary nail was replaced and bone graft was added. At 8 year-old abdominal ultrasonography revealed retroperitoneal mass, confirmed by CT as a solid lobulated mass, from upper border of the pancreas extending to retroperitoneum, histopathological study reported intraneural neurofibroma. At 9 year-old remains refractory to surgical treatment. **Conclusions:** We present a patient with NF1 diagnosis (NIH Consensus Development Conference criteria); he presented pseudoarthrosis of left tibia and retroperitoneal intraneural neurofibroma. PT is an incapacitating condition with lack of normal bone formation. Pseudoarthrosis treatment is complex. NF1 disorder requires a periodical clinical evaluation, imaging studies and multidisciplinary approach to allow an opportune detection and treatment of complications.

2432T

A family with Novel *MED12* Mutation Substantiates the Existence of a Fourth Recognizable Clinical Phenotype. *V. Ottaviani[1], D. Rogaia[2], I. Isidori[2], A. Mencarelli[2], G. Merla[1], P. Prontera[2].* 1) Medical GenMedical Genetics Unit, IRCCS Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Foggia, Italy; 2) Medical Genetics Unit, University and Hospital of Perugia, Perugia, Italy.

Different missense mutations in the *MED12* gene have been described in patients with Optiz-Kaveggia Syndrome, Lujan-Fryns Syndrome, or X-linked Ohdo Syndrome. Recently, a different phenotype, characterized by dysmorphisms of the face, severe intellectual disability (ID), absent language and friendly personality, was reported in a family carrying a frameshift *MED12* mutation (c. 5898dupC). This phenotype is clinically distinguishable from the other *MED12*-related disorder, supporting the presence of a fourth clinically recognizable *MED12* syndrome. Here we report a family, including two males and one females, affected with moderate to profound ID, in which targeted next-generation sequencing disclosed a novel c. 2312T>C (p. Ile77Ther) *MED12* mutation. This highly conserved variant segregates with the disease and is predicted to be deleterious by different mutation prediction tools; moreover it was not found in healthy first-degree male relative as well as in the SNP-database. The phenotype of male patients clearly overlap to those described in the patients carrying the frameshift mutations thus indicating the existence of a fourth, clinically recognizable *MED12*- related disorder.

2433F

Intragenic *CFTR* micro-duplication and 5T/12TG track in a patient with non-classical cystic fibrosis. *P. B. S. Celestino-Soper[1], E. Simpson[1], S. Dlouhy[1], M. Vatta[1, 2], J. Yeley[3], C. Brown[3], S. Bai[1].* 1) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA; 2) Krannert Institute of Cardiology, Division of Cardiology, Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, USA; 3) Division of Pulmonary, Allergy, Critical Care, Occupational, and Sleep Medicine, Indiana University, Indianapolis, IN, USA.

Cystic fibrosis (CF [MIM 219700]) is an autosomal recessive disorder characterized by the accumulation of sticky and heavy mucus that can damage several organs. The disorder shows variable expressivity in different affected individuals, but it typically causes respiratory and digestive system complications as well as reproductive system problems in males. Individuals with the classical form of CF usually have two classic mutations in the cystic fibrosis transmembrane conductance regulator gene (*CFTR* [MIM 602421)] such as c. 1521-1523-delCTT (p. F508del). Individuals with other variants, including a CF disease-causing mutation in *cis* with a short 5T nucleotide track and extended 12TG or 13TG di-nucleotide track in intron 8 of *CFTR* may present with classic, non-classic or milder forms of CF due to possible resulting lower levels of functional CFTR protein. In this case report, a female patient presents with chronic daily cough and sputum production, chronic sinusitis, mild reversible airflow obstruction, minimal bronchiectasis, and intermediate sweat chloride level (51mmol/L), but no pancreatic insufficiency. Clinically, the patient was classified to have features of asthma and mild or non-classic CF. We performed *CFTR* sequence analysis and deletion/duplication analysis by multiplex ligation-dependent probe amplification (MLPA). BigDye terminator *CFTR* sequence analysis demonstrated that she is a carrier for the 5T/12TG variant. Deletion/duplication analysis showed the presence of an intragenic duplication involving exons 7-11 of the *CFTR* gene. This large duplication is predicted to result in the production of a truncated CFTR protein lacking the terminal part of NBD1 domain and thus can be considered a null allele. Therefore, the combination of this large intragenic duplication and 5T/12TG probably causes the mild or non-classic CF features in this individual.

2434W

Mutation in *NRAS* in familial Noonan syndrome: A case report. *M. Wilbe[1], S. Ekvall[1], J. Dahlgren[2], E. Legius[3], A. van Haeringen[4], O. Westphal[2], G. Annerén[1], M. L. Bondeson[1].* 1) Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala, Sweden; 2) Department of Paediatrics, the Sahlgrenska Academy, Gothenburg University, Sweden; 3) Department of Human Genetics, KU Leuven, Leuven, Belgium; 4) Department of Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands.

Noonan syndrome (NS) is a heterogeneous developmental disorder associated with variable clinical expression including short stature, congenital heart defect, unusual pectus deformity and typical facial features. It is caused by activating mutations in genes involved in the RAS-MAPK signaling pathway. We present a clinical and molecular characterization of a small family with Noonan syndrome. Comprehensive mutation analysis of *NF1, PTPN11, SOS1, CBL, BRAF, RAF1, SHOC2, MAP2K2, MAP2K1, SPRED1, NRAS, HRAS* and *KRAS* was performed using targeted next-generation sequencing. The result revealed a recurrent mutation in *NRAS*, c. 179G>A (p. G60E), in the index patient. This mutation was inherited from the index patient's father, which also showed signs of NS. We describe clinical features in this family and review the literature for genotype-phenotype correlations for NS patients with mutations in *NRAS*. Neither of the affected individuals in this family presented with juvenile myelomonocytic leukemia (JMML), which together with previously published results suggest that the risk for NS individuals with a germline *NRAS* developing JMML is not different from the proportion seen in other NS cases. Interestingly, 50% of NS individuals with an *NRAS* mutation (including our family) present with lentigines and/or Café-au-lait spots. This demonstrates a predisposition to hyperpigmented lesions in *NRAS*-positive NS individuals. In addition, the affected father in our family presented with a hearing deficit since birth, which together with lentigines are two characteristics of NS with multiple lentigines (previously LEOPARD syndrome), supporting the difficulties in diagnosing individuals with RASopathies correctly. The clinical and genetic heterogeneity observed in RASopathies is a challenge for genetic testing. However, using the advent of next-generation sequencing, which allows screening of a large number of genes simultaneously, will have an important impact on the correct diagnosis as well as prognosis and treatment of these patients in the future.

2435T

Complex Inheritance of *ABCA4* Disease: Four Mutations in a Family with Multiple Macular Phenotypes. *W. Lee[1], Y. Xie[1], J. Zernant[1], B. Yuan[2], S. Bearelly[1], S. H. Tsang[1, 3], J. R. Lupski[2], R. Allikmets[1, 3].* 1) Department of Ophthalmology, Columbia University, New York, NY; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 3) Department of Pathology and Cell Biology, Columbia University, New York, NY.

Over 800 mutations in the *ABCA4* gene cause autosomal recessive Stargardt disease. Due to extensive genetic heterogeneity, observed variant-associated phenotypes can manifest extensive variability of expression. In addition, high carrier frequency of pathogenic *ABCA4* alleles in the general population (~1:20) often results in a pseudo-dominant inheritance pattern. Here we describe a genotype/phenotype analysis of a two-generation family with multiple macular disease phenotypes segregating four *ABCA4* mutant alleles. Complete sequencing of the *ABCA4* gene discovered 2 known missense mutations, p. C54Y and p. G1961E. Array comparative genomic hybridization revealed a large novel deletion combined with a small insertion, c. 6148-698_c. 6670del/insTGTGCACCTCCCTAG, and complete sequencing of the entire *ABCA4* genomic locus uncovered a complex allele with two new deep intronic variants, c. 302+68C>T. Patients with the p. G1961E mutation had the mildest, confined maculopathy phenotype with peripheral flecks while those with all other combinations of mutant variant alleles exhibited a more advanced stage of generalized retinal and choriocapillaris atrophy. This family epitomizes the complexity of the analysis of *ABCA4*-associated diseases. It contained variants from all classes of mutations, in the coding region, deep intronic, both SNV and CNV, and suggested dominant inheritance. Unequivocally defining disease-associated alleles in the *ABCA4* locus requires extensive analyses by a combination of mutation detection approaches.

2436F

A BOY WITH GNAS MUTATION. *S. Tulgar Kinik[1], NM. Sahin[2], FB. Atac[3], H. Verdi[3], S. Thiele[4].* 1) Pediatric Endocrinology, Baskent University Faculty of Medicine, Ankara, Turkey; 2) Dr. Sami Ulus Obstetrics, Children Health and Diseases Training and Research Hospital, Ankara, Turkey; 3) Baskent University Faculty of Medicine, Department of Medical Biology, Ankara, Turkey; 4) Division of Experimental Paediatric Endocrinology and Diabetes, Department of Paediatrics, University of Lübeck Lübeck, Germany.

Pseudohypoparathyroidism (PHP) is a disorder of end-organ resistance primarily affecting the actions of parathyroid hormone (PTH). Parental imprinting and the type of the genetic alteration play a determinant role in the phenotype expression of GNAS locus associated with pseudohypoparathyroidism (PHP). GNAS locus gives rise to several different messenger RNA transcripts that are derived from the paternal allele, the maternal allele, or both and can be either coding or non-coding. As a consequence, GNAS mutations lead to a wide spectrum of phenotypes. Here we present a boy with GNAS mutation. Case:12. 2 years old boy admitted to our center with the history of congenital ptosis,osteoma cutis and primary hypothyroidism. His laboratory findings revealed mild hypocalcemia, hyperphospatemia, elevated PTH level with the normal vitamin D and magnesium level. He has Albrigth Hereditary Osteodistrophy (AHO) features, mild mental retardation with the normal height and appropriate pubertal stage (Tanner stage III). He was born term and low birth weight. The parents werenon-consanguineous. His mother, fatherand sister laboratory findings were normal and they have not AHO features. In the mother osteoma cutis was detected. The Gsα enzyme activity was low in proband (%53,8/55); while %85,7/82,1 and :%96,2 in mother and father respectively. The molecular data has revealed c. 936_946dup CTTTCCAGAAT in exon 11. No mutation was determined neither mother nor father. All the findings of the proband implies to PHP type1a. However the lack of mutation, normal Gsα enzyme activity and the existence of osteoma cutis in mother were the findings that differ from the previous reports. In addition the coexistance of PHP-congenitalptosis was defined as PHPtype 1b, that clinical presentation may points out mild methylation defects or other epigenetic variations.

2437W

Novel *SOX10* variant causing Waardenburg syndrome without cutaneous pigmentary changes. *C. J. Adams[1], K. Chao[1], L. A. Wolfe[1,2], B. Brooks[3], W. A. Gahl[1,2], C. Toro[1], D. R. Adams[1,2].* 1) Undiagnosed Diseases Program, National Human Genome Research Institute, Bethesda, MD; 2) Office of the Clinical Director, National Human Genome Research Institute, Bethesda, MD; 3) National Eye Institute, Bethesda, MD.

We present an 11 year, 10 month old Northern European and Ashkenazi Jewish male who was diagnosed with SOX10-associated Waardenburg syndrome during an evaluation by the NIH Undiagnosed Diseases Program. Extensive phenotyping highlights typical and atypical features of this rare syndrome. Early disease manifestations included sensori-neuro hearing loss, first identified at 8 weeks of age. Nystagmus with left eyelid ptosis and cortical visual impairment associated with hypotonia were diagnosed soon after the hearing loss. A head CT scan demonstrated vestibular hypoplasia and a cochlear implant was placed at 1 year of age. Other findings included short stature, growth hormone deficiency, attention deficit hyperactivity disorder, neurogenic bladder, and a combination of spastic diplegia and peripheral neuropathy requiring a wheelchair. His NIH evaluation confirmed a prior diagnosis of attention deficit disorder and determined his total IQ, off medications, to be 83. Ophthalmological evaluation identified nystagmus, asymmetric, segmental choroidal hyperpigmentation, mild optic pallor, and myopia. Dental evaluation found no evidence of ectodermal dysplasia. Dermatologic evaluation concluded that the hypo-pigmented areas were not related to an overall syndrome and were most consistent with healing scratches from his dry skin. Clinical sequencing identified a novel *SOX10* frameshift mutation c. 1107delC in exon 4. Similar mutations have been reported in both Type 2 & Type 4 Waardenburg syndrome. This syndrome has been reported as both an autosomal dominant disorder with variable penetrance as well as sporadic. Neither parent carried this variant. Up to 90% of those affected with Type 2 Waardenburg syndrome have sensori-neural hearing loss and many have inner ear malformations such as those identified in our patient. Demyelinating neuropathy, leukodystrophy with spasticity, various degrees of intellectual disability and seizures have all been reported in both Type 2 and 4 Waardenburg syndrome. Retinal and skin pigmentary changes are highly variable among Waardenburg patients, although our patient had much more significant retinal pigmentary changes than has previously been described.

2438T

Desmosomal diseases: the key role of dermatological features for the detection of severe subclinical cardiac involvement and for the phenotype to genotype correlation : Proposition of an algorithm. *L. Polivka[1, 2], C. Bodemer[1, 2], S. Hadj-Rabia[1, 2].* 1) Dermatology, Necker-Enfants Malades Hospital, AP-HP, Paris, France; 2) Inserm U1163, Institut Imagine, Université Paris Descartes.

Inherited desmosomal diseases are characterized by skin and/or cardiac features. The dermatological features might be a clue in the determination of the underlying life-threatening cardiac disease. After systematic review of published articles (1997-2015), we propose a clinical dermatological algorithm in order to identify as soon as possible the patients at high risk of cardiac involvement and to orientate gene sequencing. The inclusion criteria were: 1) At least one identified mutation in the following desmosomal genes: *JUP*; *PKP1, PKP2*, *DSP*, *DSG1* to *4*, *DSC1-3* and *CDSN*; 2) Description of the cutaneous phenotype. Articles focusing on extra-cutaneous manifestations and/or cardiac involvement only and redundant articles were excluded. Sequence variations were checked and actualized using ENSEMBL software. A total of 78 articles fulfilled the inclusion criteria. They reported 458 patients carrying at least one mutation in one of the eight genes encoding proteins of the epidermal desmosome. Palmoplantar keratoderma (PPK), hair shaft anomalies (HSA), and skin fragility were the major features of cutaneous desmosome diseases. Isolated PPK or isolated HSA are associated to a desmosomal disease limited to skin. The combination of PPK and HSA was recorded in 161 patients, and this association is at high risk of cardiac disease (129/161 patients, 80. 1%). In all the 129 reported patients, dermatological features occurred before cardiac manifestations i. e. HSA since birth, PPK around one year. However, they had led to cardiac monitoring in only 2. 3% of those patients. Therefore regular cardiac monitoring, including cardiac ultrasound and electrocardiogram, is mandatory in all patients presenting with both PPK and HSA, even in the absence of clinical signs of cardiac involvement. Taking PPK, HSA and skin fragility into account, we delineated three major phenotypes: 1) the PPK-HSA-non fragile skin subtype (77%, 124/161), always associated to cardiac involvement; 2) the PPK-HSA-skin fragility-normal cardiac function subtype (19. 9%, 32/161) frequently associated to *PKP1* anomalies (*PKP1* is not expressed in the heart); 3) the PPK-HSA-skin fragility-cardiac involvement subtype (3. 1%, 5/161) always due to *DSP* mutations. Three mutation hotspots in *DSP* and *JUP* account for 90. 8% of the patients with cardiac involvement. The combination of PPK and HSA justify long-term cardiac monitoring unless *PKP1* mutation is identified.

2439F

Two Siblings With Conjenital Hyperinsulinism. *N. Muratoglu Sahin[1], H. Verdi[2], FB. Atac[2], S. Tulgar Kinik[3], SE. Flanagan[4], H. Khalid[5].* 1) Division of Pediatric Endocrinology, Dr . Sami Ulus Obstetrics, Children Health and Diseases Training and Research, Ankara, Turkey; 2) Baskent University Faculty of Medicine, Department of Medical Biology, Ankara, Turkey; 3) Baskent University Faculty of Medicine, Division of Pediatric Endocrinology, Ankara, Turkey; 4) Institute of Biomedicaland Clinical Science, University of Exeter Medical School, UK; 5) Departments of Pediatric Endocrinology, Great Ormond Street Hospital for Children NHS Trust, London, UK.

Congenital hyperinsulinism (CHI) is a complex heterogeneous disorder characterizedby dysregulated insulin secretion from pancreatic β-cells causes. The molecular basis of CHI involves perturbations in the key genes involved in controlling insulin secretion. One of the known etiologies includes inactivating mutations of the KATP channel genes. Autosomal recessive and dominant mutations in ABCC8/KCNJ11 are the commonest cause of medically unresponsive CHI. In addition the focal form appears to be higher in Asian countries; therefore, the establishment of treatment systems should be considered. Here, we report two siblings diagnosed with CHI, due to a novel mutation (p. Q392H) in the *ABCC8* gene. Case1: A male newborn (no problem in pregnancy and delivery) was admitted to hospital suffering from hypoglycemic seizure started in the first day. The parents were first cousins. During hypoglycemia, blood ketone was negative, insulin was 400 uIU/mL (5–25), c-peptide was 26. 5 ng/mL (0. 9–7. 1), the infant was diagnosed as CHI. During follow-up, with the intensive medical therapy (diazoxide, somatostatin, nifedipine, and intensive feeding), his blood glucose levels were normal. Although hypoglycemia was seldom detected, his neurological status never improved. He was died in two months. The sequencing analysis revealed a novel p. Q392H mutation in ABCC8 gene. Both parents were found to be heterozygote carriers of the same mutation. Case 2: Corion villus sampling was performed for the second baby. The molecular analyses have revealed that the baby girl was heterozygote for the same mutation. She was born from uneven fulpregnancy and delivery. In the postnatal second hour, hypoglycemia was detected. She was diagnosed CHI as well. With the treatment of octreotide, normoglycemia was maintained and the treatment was ceased at the first age. With the further molecular analysis, the paternal inheritance of mutant allele was determined. We were unable to perform a DOPA-PET scan in our case. Finally, it maybe autosomal dominant diffuse CHI. The difference in the severity of the disease may be due to homo and heterozygosity of the mutation in siblings. Moreover, paternally inherited allele as in the second baby might cause focal hyperinsulinism.

2440W

Mutational analysis of an adult with an RNU4ATAC associated primordial dwarfism: expanding the phenotype. *S. M. Kirwin[1], I. L. Gonzalez[1], R. L Smith[3], M. B. Bober[2], E. T. Rush[3, 4], V. L. Funanage[1].* 1) Molecular Diagnostics Laboratory, Nemours/A. I. duPont Hospital for Children, Wilmington, DE; 2) Division of Orthopaedics, Nemours/A. I. duPont Hospital for Children, Wilmington, DE; 3) Munroe-Meyer Institute for Genetics and Rehabilitation, University of Nebraska Medical Center, Omaha, NE; 4) Departments of Pediatrics, Internal Medicine, and Orthopaedic Surgery, University of Nebraska Medical Center, Omaha, NE.

Microcephalic osteodysplastic primordial dysplasia (MOPD I; OMIM 210710) is a rare autosomal recessive disorder characterized by intrauterine as well as postnatal growth delay, severe microcephaly, distinct brain malformations (pachygyria or agyria); short bowed long bones, broad metaphyses, dry skin, and sparse hair and eyebrows. Facial characteristics may include a sloping forehead; protruding eyes; a prominent nose with a flat nasal bridge; and micrognathia. Mutations in the *RNU4ATAC* small nuclear RNA have been linked to this disorder. This snRNA is mapped to intron 2 of the *CLASP1* gene on chromosome 2, (2q14. 2), and is a component of the minor U12-dependent spliceosome. Eleven different variants in *RNU4ATAC* have been reported and lead to a spectrum of severity within the MOPD I phenotype. One recurrent variant, n. 51G>A has been reported across differing ethnic backgrounds. We report here on a now 22 year old patient with microcephaly (-8-10 SD), lissencephaly, agenesis of the corpus callosum and dwarfism with a height of 132cm. He presented to hospital with new onset of type II diabetes. He was dysmorphic with sloping hairline and prominent supraorbital ridges with lateral periorbital fullness. He has sparse eyebrows, and palpebral fissures are narrow with hooded eyelids and a left Horner syndrome. His nose has high and tubular nasal bridge with broad tip and depression between the nasal tip and the alae nasi. He has macrostomia and fleshy lips. Hands are small compared with the rest of the body, skin appears loose and wrinkled. Terminal phalanges and fingernails are hypoplastic. Feet are edematous which is non-pitting. The patient has continued to have a progressive clinical course with onset of anemia over the past 15 months. He also has developed stage V chronic kidney disease and at the time of most recent evaluation, had a glomerular filtration rate of 11 mL/min. Sanger sequencing of the *RNU4ATAC* region identified two novel variants: a n. 46G>C within the 5' stem loop region, and n. 118T>C, in the Sm protein binding domain. The pathogenicity of n. 118T>C is supported by a previous study that mutations in the 3' end of the stem loop disrupts the normal spliceosomal complex. Our patient, who is likely a compound heterozygote, may be the longest surviving patient with MOPD I that we are presently aware of. Our findings may provide insight into the molecular basis of an MOPD I patient who demonstrates a phenotype unlike those previously reported.

2441T

Genetic Analysis in Beta Thalassemia Minors in Certain Communities of Gujarati Population. *M. V. Rao[1], S. R. SHAH[1], A. P. PATEL[2].* 1) Department of Zoology, BMTC & Human Genetics, School of Sciences, Ahmedabad-380009, Gujarat, India; 2) 2, Saumrut Laboratory, 402, Narayan Complex, Navarangpura, Ahmedabad-380009.

Thalassemia is one of the hemoglobinopathies which belongs to a class of genetic disorders. Beta thalassemia is an inheritable genetic anomaly whose gene belongs to chromosome 11. The genetic mutation in that chromosome leads to the abnormal synthesis of β-globin chain of the hemoglobin molecule. More than 300 mutations are noted for this gene and this disease occupies one of the most important genetic disorders in global populatio. In India, patients are identified by characterization of common mutation in them, where IVS 1-5 is predominant. In Gujarati Population of our study, a total number of 108 patients' DNA was analyzed for five common mutations using molecular biological methods. These include IVS 1-5 (G>C), -619 bp deletion, IVS 1-1 (G>T), CD 8/9(+G) also known as FS 8/9 along with CD 41/42 respectively. The results showed that IVS 1-5 mutation was found to be dominant (70%) followed by 11% of -619 bp deletion. All patients in this study consisted of Prajapatis (39), Muslims (32), Sindhis (14), Patels (8) and others (15) (Rajput, Lohanas, etc.). All these ethenic groups had also the predominance of IVS 1-5 followed by -619 bp deletion. However in Muslim community, second highest mutation was of IVS 1-1 and in Sindhis, -619 bp deletion was seen high followed by IVS 1-5 and CD 8/9, though this community had less in number in this study. Similarly other community populations showed variations in these mutations analyzed. These mutational variances are due to many factors like migration, natural selection, ethical issues and other factors and thus signifies the importance of community based studies in relation to mutational variations of thalassemia patients.

2442F

Hemoglobin beta chain -619 base pair deletion screening of patients with high HbA2 Level. *R. J. Raval[1], N. Shah[2], P. S. Shah[2], M. V RAO[1].* 1) Gujarat Genetic Diagnostic Center (GenDiCe), Department of Zoology, BMTC and Human Genetics, Ahmedabad, India; 2) Supratech Laboratory and Research Institute, Off C G Road, Opposite Krupa Petrol Pump, Parimal Garden, Ahmedabad, Gujarat 380006, India.

Introduction: Beta thalassemia is a blood disorder that affects the production of abnormal hemoglobin. Low levels of hemoglobin leads to a lack of oxygen in the body. Affected individuals also have a shortage of red blood cells (anemia), which can cause pale skin, weakness, fatigue, and more serious complications. Patients with beta thalassemia are at an increased risk of developing abnormal blood clots. Beta thalassemia is classified into two types depending on the severity of symptoms: thalassemia major (Cooley's anemia) and thalassemia intermediate (Minor). Thalassemia major is more severe. **Methods:**Whole Blood samples were collected from patients (n=107) with complain of high HbA 2 level. HbA 2 levels were evaluated. Genomic DNA was isolated from peripheral blood and its quality and quantity were checked. With help of Polymerase Chain Reaction (PCR) amplification of the exon of genes HBB was carried out. PCR products were checked by 2. 5 % Agarose Gel Electrophoresis. **Results**:In Beta Thalassemia, out of 107 patients 13 revealed -619 base pair deletion with heterozygous condition, hence declared as Thalassemia minor. No further mutations were seen in this study. **Conclusion**:In Asian population, Beta Thalassemia patients exhibited alterations in 619 base pair deletion only. However, further studies are necessary.

2443W

Next-generation sequencing uncovered unexpectedly high prevalence of neonatal Dubin-Johnson syndrome. *T. Togawa[1], T. Sugiura[1], K. Ito[1], T. Endo[1], A. Kikuchi[2], N. Ichinoi[2], S. Kure[2], S. Saitoh[1].* 1) Department of Pediatrics and Neonatology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan; 2) Department of Pediatrics, Tohoku University School of Medicine, Sendai, Japan.

Neonatal cholestasis results from impaired bile formation by hepatocytes or from obstruction of bile flow. The incidence of neonatal cholestatic jaundice varies between 1:2,500 and 1:5,000 newborns. The etiology is highly variable. Thanks to advances in the understanding of the molecular basis of neonatal/infantile intrahepatic cholestasis (NIIC), diagnostic methods taking into account the genetic make-up have developed. Several disorders have been divided from idiopathic neonatal hepatitis in the 1970s to 2000s. The Dubin-Johnson syndrome (DJS, [MIM: 237500]) phenotype results from mutations in the protein encoded by *ABCC2* [MIM: 601107], which is a member of the large ABC transporter family. DJS was rarely diagnosed in the neonatal period and thus far only five neonatal DJS cases have undergone molecular analysis. We developed a diagnostic custom panel of causative 18 genes (NIIC gene panel), included *ABCC2*, using the AmpliSeq (Life Technologies). From June 2013 to May 2015, we sequenced 134 patients with NIIC using Ion PGM system (Life Technologies) and analyzed with bioinformatics. We diagnosed six neonatal DJS cases (including 2 siblings) with the molecular analysis. Two patients had clinical diagnosis with the black liver. Other four had no definitive diagnosis before analysis. Five of 6 patients were Japanese patients and the other one was Chinese. Eight heterozygous pathogenic mutations were identified, those included 1 nonsense mutation, 4 splice site mutations, 2 missense mutations, and 1 small deletion. All patients had compound heterozygous mutations, including predicted truncating mutations or exon skipping at least in one allele. We suspected the presence of hot spots in *ABCC2* for neonatal DJS in Japanese population, because three Japanese patients had p. R768W. We have calculated the cumulative carrier rate of our identified mutations in Japanese population thanks to Human Genetic Variation Browser (URL: http://www. genome. med. kyoto-u. ac. jp/SnpDB), and the rate was predicted approximately 1/54. Based on the carrier rate, estimated incidence of the affected patients was calculated approximately 1/12,000, and we speculated that neonatal DJS would be more frequent in Japanese. In conclusion, NGS and bioinformatics with NIIC gene panel are clinically useful to uncover neonatal DJS. Neonatal DJS would be more frequent in Japanese population.

2444T

An unusual presentation of Schimke Immunoosseous Dysplasia. *M. AlAmr[1], A. Torkamani[2], L. M. Bird[3].* 1) Health Science International, University of California, San Diego, San Diego, CA; 2) Scripps Translational Science Institute, La Jolla, CA, United States; 3) Department of Pediatrics, University of California, San Diego, CA, and Rady Children's Specialists of San Diego, Division of Dysmorphology/Genetics, San Diego, CA, United States.

A Middle Eastern male was born to second-cousin parents at 35 weeks of gestation following premature rupture of membranes at 25 weeks. He was delivered by Cesarean section weighing 1. 7 kg. Bilateral cataracts were removed at 2 years of age. He had delayed tooth eruption and mild gross motor delay. Short stature became concerning around two years of age, and he was treated with an unknown injectable medication. Esotropia was surgically corrected. Pectus carinatum developed around age 5 years and hyperpigmented macules appeared at 6 years. He exhibited flesh colored papules on his neck consistent with flat warts; biopsy showed features of epidermodysplasia verruciformis. He gradually developed disproportionate short stature and lordosis. No specific diagnosis was suggested upon review of radiographs (taken at age of 5 years) at the International Skeletal Dysplasia Registry. Past work up included karyotype (46XY), lysosomal enzyme panel (normal) and microarray (no copy number variations, 216. 5 Mb of homozygosity consistent with stated consanguinity). At the age of 7 years, his evaluation showed microcephaly and short stature (- 7. 9 SDs) with arm span 14% greater than height and normal upper/lower segment ratio. He had a short trunk, pectus carinatum and dramatic lordosis. There were numerous tan macules scattered on trunk. Radiographs from age 7 years showed platyspondyly, decreased bone mineralization, dysplastic acetabulae, small irregular dislocated femoral heads, osteopenia and lacy ossification of the humeral head, relatively normal epiphyses at the knees and elbows, and metacarpal shortening. Again no specific diagnosis was made. Whole genome sequencing revealed a homozygous nonsense mutation in SMARCAL1 (chr2:217288391 snp G>T), p. E378X). The same mutation was reported previously in a Middle Eastern girl with Schimke Immunoosseous Dysplasia (SIOD). SIOD is an autosomal recessive progressive multisystem disorder causing skeletal, renal and immunologic dysfunction. The most consistent presentation is decreased postnatal growth rate and nephropathy in addition to skin pigmentation beginning in childhood. Cataracts have been described but are uncommon. Radiographs were re-reviewed in light of the mutation results, and were felt to be atypical and more severe than usually seen in SIOD. The lack of clinical immunodeficiency, lack of clinical renal disease, presence of cataracts and atypically severe skeletal findings make this case noteworthy.

2445F

A novel, SF3B4 missense mutation, as a possible cause for Nager syndrome. *C. K. Dzolang[1], J. S. Bamforth[1], K. Atta[2], E. Garside[2], A. MacMillan[2], O. Caluseriu[1].* 1) Medical Genetics, University of Alberta, Edmonton, Canada; 2) Biochemistry, University of Alberta, Edmonton, Canada.

Nager syndrome (NS; OMIM 154400) is the prototype for a group of disorders called acrofacial dysostoses characterized by mandibulofacial and limb anomalies. NS is an autosomal dominant genetic disorder with a variable, multisystemic involvement but hallmarks of the condition are a recognizable facial gestalt including downslanted palpebral fissures, malar hypoplasia, and micrognathia, and most commonly preaxial, predominantly upper limb anomalies including aplasia/hypoplasia of the thumb and radioulnar synotosis. In 2012, Bernier et al, have hypothesized that haploinsufficiency of SF3B4 is the cause of 61% of the individuals clinically diagnosed with Nager syndrome in their cohort. Further studies support this initial finding (Czeschik et al, 2013; Petit et al, 2014). Published pathogenic mutations associated with Nager syndrome include a specific missense in the initiator Met, nonsense, frameshift, and splicing mutations. To our knowledge, no other pathogenic NS missense mutations have been published to date. SF3B4 encodes SAP49, a highly conserved spliceosomal protein, with two RNA recognition motifs (RRMs) followed by a proline-glycine rich domain. During assembly of the U2SNP prespliceosomal complex, SAP49 binds to the pre-mRNA just upstream of the branch point sequence and is believed to play a crucial role in tethering the U2 snRNP to the branch site. This report brings preliminary evidence for the first SF3B4 missense mutation outside the initiator Met in a patient with a clinical diagnosis of Nager Syndrome. Our patient, a 12 years old boy, presented with a constellation of features including relative macrocephaly, downslanted palpebral fissures, zygomatic hypoplasia, micrognathia, bilateral hypoplastic thumbs, conductive hearing loss, and speech delay. Sanger sequencing of SF3B4 in our patient revealed a novel variant (c. 251T>G) in exon 3, a replacement of an isoleucine (hydrophobic amino acid) by an arginine (basic amino acid). In silico analysis (SIFT, Mutation Taster and Polyphen) indicate that this change is deleterious/ disease-causing. This mutation is situated in the first RRM of SF3B4, and protein homology and conformation study indicate severe disruption of Sap49 at an essential site. Further functional studies are under way to confirm pathogenicity. We are reporting, to our knowledge, the first patient with a clinical diagnosis of NS and a rare missense mutation likely pathogenic in the functionally essential SF3B4 RRM1 domain.

2446W

A novel SOX9 mutation with normal male genitalia and otherwise classic campomelic dysplasia. *G. C. Gowans, M. J. Curtin, K. R. Kerr, E. K. Schmitt, L. A. Autullo, P. L. Brock, K. E. Jackson, K. C. Platky Warren, K. L. Stoate, K. M. Goodin, J. H. Hersh, A. Asamoah.* Pediatrics, University of Louisville, Louisville, KY.

A21 month old male has campomelic dysplasia (CD), normal male genitalia, and de novo SOX9 mutation. The child was born after prenatal diagnosis of skeletal abnormalities and hydronephrosis. Complications in the newborn period were related to congenital abnormalities and included respiratory distress and feeding difficulties. The patient has a variety of congenital abnormalities including tracheomalacia, cardiac anomalies, conductive hearing loss, bilateral hydronephrosis, and low lying conus. The skeletal anomalies included hip dislocation, 11 pairs of ribs, rhizomelic extremities, bowed lower extremities with pretibial dimples in the skin, malformed tibia/fibula, and talipes equinovarus with wide spacing between the hallux and 2nd toes. Facial features are typical of CD, including relative macrocephaly, flat face, hypertelorism, upturned nose, long philtrum, and micrognathia with cleft palate. External genitalia were phenotypically male. The family history is negative for consanguinity, genetic conditions, birth defects, recurrent pregnancy loss, or ID. CD particularly impacts the skeletal, respiratory, and reproductive systems. Skeletal findings typically include bowing of the long bones of the lower extremities and less frequently the upper extremities as well that causes characteristic skin dimples over the curved bone(s). Shortened limbs, dislocated hips, underdeveloped shoulder blades, 11 pairs of ribs, cervical spine abnormalities, and talipes equinovarus are common. Reproductive findings include various levels of both external and internal genital defects, most commonly manifesting as ambiguous genitalia or sex reversal. Almost 75% of affected males have a 46,XY genotype with either ambiguous genitalia or normal female external genitalia. Internal genitalia are variable. The SOX9 gene on 17q24. 3 is the only gene known to be associated with CD. Following gene sequencing of SOX9, also known as SRY (sex-determining region Y)-box 9 protein, a c. 349C>G (Gln117Glu) variant of uncertain significance was found. This had not been reported previously. This c. 349C>G transversion in exon 1 converts a codon from glutamine (CAG) to glutamic acid (GAG). Parental testing revealed neither parent has the mutation, consistent with a de novo mutation pattern. This particular SOX9 variant may explain the normal male phenotype of this patient. We will compare our patient to some relevant previously reported SOX9 genotypes with their resulting CD phenotypes.

2447T

A novel mutation of *OTX2* contributing to a case of otocephaly initially diagnosed by prenatal ultrasound in the first trimester. *M. K. Jones[1], J. Chung[1], V. Kimonis[1], JA. Gold[1, 2].* 1) Pediatrics, University of California, Irvine, Orange, CA; 2) Loma Linda University Health, Loma Linda, CA.

We report a prenatal case of an unspecified facial defect first identified early in pregnancy during a routine first trimester ultrasound. CVS was subsequently performed with a normal female karyotype 46,XX and SNP-microarray. The patient is a 37-year-old G3 P0 TAB1 Hispanic female. With a previous partner, the patient had infertility with two unsuccessful intrauterine insemination (IUI) attempts and one unsuccessful *in vitro* fertilization (IVF) cycle. With the current partner, the patient had a previous pregnancy with trisomy 21, which she terminated. Family history was otherwise unremarkable. Detailed ultrasound in the second trimester revealed: agnathia, low set ears (otocephaly), aglossia, and microstomia. A dandy-walker variant was also suspected, although a subsequent fetal MRI did not confirm this finding, with brain anatomy appearing within normal limits. The rare, lethal agnathia-otocephaly phenotype was suspected. The patient was counseled regarding her options and elected continuation of the pregnancy with nonintervention at the time of delivery. The pregnancy was complicated by severe polyhydramnios requiring multiple amnioreductions for symptomatic relief. Birth was at 38. 5 weeks gestation after spontaneous labor; she was a stillborn female with confirmed otocephaly on physical examination by geneticist and on autopsy. Tissue was sent for DNA analysis of *OTX2* and *PRRX1*. Recently mutations in two genes, orthodenticle homeobox 2 (*OTX2*) and paired-related homeobox 1 (*PRRX1*), have been found to be associated with some cases of otocephaly. DNA analysis of *OTX2* and *PRRX1* in our case revealed a novel heterozygous nonsense mutation in *OTX2*. Specifically, the mutation found was c. 534C>A p. CYS178* (C178X) in exon 3 of the *OTX2* gene. DNA analysis of *PRRX1* was negative for sequence alterations and deletions/duplications within the gene. Other cases presumed to be caused by mutations in *OTX2* have been reported in the literature; still, many cases have no identifiable genetic etiology. There is one report of autosomal dominant transmission and presumed gonadal mosaicism. Parental testing is pending for the familial mutation to determine ifit is*de novo.* These findings could suggest that mutations in *OTX2* contribute; but, may not be the only genetic factor related to the overall phenotype. Additional studies might include DNA analysis of *SOX9*previously implicated in Pierre-Robin sequence, a less severe mandibular phenotype.

2448F

A novel mutation of the *TWIST1* gene in a family with Baller-Gerold syndrome without poikiloderma. *T. Kaname[1,2], K. Sameshima[3], K. Yanagi[1], Y. Kuroki[1], G. Nishimura[4], K. Naritomi[2], Y. Matsubara[5].* 1) Dept Gen Med, Natl Ctr Child Hlth Dev (NCCHD), Tokyo, Tokyo, Japan; 2) Dept Med Gen, Univ Ryukyu Grad Schl Med, Nishihara, Okinawa, Japan; 3) Div Med Gen, Gunma Child Med Ctr, Gunma, Japan; 4) Dept Ped Imag, Tokyo Mtrpltn Child Med Ctr, Tokyo, Japan; 5) Natl Res Inst Child Hlth Dev, Tokyo, Japan.

Baller-Gerold syndrome (BGS) is characterized by craniosynostosis, paraxial upper limb malformations, and growth retardation, and sometimes by poikiloderma. Although many patients with BGS are caused by mutations in the *RECQL4* gene, BGS is known to be of heterogeneous cause. In fact, three BGS patients with mutations in the *TWIST1* gene were reported to date. We met a family with clinically diagnosed with mild BGS without poikilodermic skin changes. The boys were the second and third children born to nonconsanguineous parents. There was no prenatal exposure to medications. They had bilateral coronal craniosynostosis with prominent forehead, dysmorphic facial appearance with left ptosis, and mild radial ray hypoplasia with abnormal thumb. Chromosome analysis showed a normal karyotype in each patient. To elucidate the cause of mutation in the patients, we performed a targeted resequencing analysis by next-generation sequencer (NGS) with a custom-made panel for genes related to craniosynostosis, including *RECQL4, TWIST1, FGFR1, FGFR2*, and *FGFR3*. Then, whilst no pathogenic mutations were detected by NGS data analysis, coverage at a part of exon 1 in the *TWIST1* gene was reduced compared to control individuals. Thus, we confirmed the sequence in *TWIST1* and plus *RECQL4* in the patient by conventional Sanger sequencing. Finally, although pathogenic mutations were not found in the *RECQL4* gene including promoter region, a 21 bp insertion at exon 1 of the *TWIST1* gene was found in the patients. The insertion is located at the coding region of basic helix-loop-helix domain, which may affect the activity of TWIST protein. Our result and previous reports of *TWIST1* mutations in patients with BGS may suggest a subtype of BGS without poikiloderma.

2449W

TWIST2 is the cause of Ablepharon Macrostomia and Barber Say Syndromes: Reascertainment of two cases lost to follow-up as keys to solving these conditions. *T. Markelllo[1], S. Marchegiani[2], T. Davis[1], F. Tessadori[3], G. van Haaften[4], F. Brancati[5], B. Pusey[1], L. Wolfe[1], C. Tifft[1], V. Ferraz[6], N. Roche[7], F. Ramos[8], M. Malicdan[1], C. Boerkoel[1], W. A. Gahl[1], B. de Vries[9], M. van Haelst[4], M. Zenker[10], special acknowledgement by the authors to F. Greenberg posthumously.* 1) NHGRI/NIH, Bethesda, MD; 2) WalterReed National Millitary Medical Center, Bethesda, MD; 3) Hubrect Institute-KNAW, CT Utrecht, The Netherlands; 4) University Medical Center Utrecht, The Netherlands; 5) University of G. d' Annunzio Chieti and Pescara, Chieti, Italy; 6) Universidade de Sao Paulo, Brazil; 7) University Hospital of Ghent, Ghent, Belgium; 8) Universidad de Zaragoza, E-50009 Zaragoza, Spain; 9) Radboud University Medical Center, Nijmegen, Netherlands; 10) Institute of Human Genetics, Magdeburg, Germany.

We recently published TWIST2 as the genetic locus of ablepharon macrostomia (AMS) and Barber-Say Syndrome (BSS). In several independent cases they are caused by a single base, changed into one of two different alternative alleles (Marchegiani et al. "Recurrent mutations in the basic domain of TWIST2 cause Ablepharon macrostomia and Barber-Say syndromes", AJHG in press). During ascertainment we discovered that the index case's father underwent mosaicism karyotype analysis at birth and due to fortuitous circumstances we were able to obtain the birth records from copies maintained by the grandmother and from photographs saved by Dr. F. Greenberg then inherited by Ms. Ann Smith after his death. Records and photographs were rediscovered after the molecular biological work was begun. Additional information from a second historical case reported one decade previous was re-ascertained by our Italian coauthor. The participation of that person facilitated understanding of the mosaic presentation in these unique two cases of AMS. The re-ascertainment of another cryptic genetic condition, previously studied then lost to follow-up has been described: Case 11-2005 — A 32-Year-Old Pregnant Woman with an Abnormal Fetal Karyotype. Holmes et al. N Engl J Med 2005; 352:1579-1587. These examples of two different genetic conditions that were lost to followup for an entire generation illustrate the importance of preserving and archiving data and photographs from old genetic cases, both unsolved or solved. The first example shows the benefit of preserving old photographs that require digitizing. This opportunity only exists prior to the medical-legal dates for destruction of hospital records, and prior to the retirement or death of senior genetic investigators.

2450T

Hutchinson-Gilford progeria syndrome: clinical and molecular characterization. *H. Pachajoa[1, 2], AF. Ramirez[1], X. García[1], LJ. Posso[1], F. Ruiz[1].* 1) Universidad Icesi, Cali, Colombia; 2) Fundación Clínica Valle del Lili, Cali, colombia.

Hutchinson-Gilford progeria syndrome (HGPS, MIM 176670) is a rare congenital disease secondary to mutations in the LMNA gene. It is characterized by lipodystrophy, short stature, low body weight, scleroderma, decreased joint mobility, osteolysis and senile facial features. Patients have a life expectancy of 13. 4 years and cardiovascular compromise is usually the cause of death. **Case report:** This is the case of a 14 year old female patient with scarse hair, dysmorphic features, short stature, senile appearance and congestive heart failure. LMNA sequencing detected a mutation a heterozygous mutation c. 1824C>T (p. Gly608Gly) which confirms a diagnosis of HGPS. **Conclusions:** The "classical" phenotype of HGPS is secondary to the mutation in LMNA gene c. 1824C>T (p. Gly608Gly), which leads to the formation of an aberrant farnesylated protein, progerin. There is an ongoing clinical trial in phase II which combines the use of farnesyltransferase inhibitor (lonafarmib), statins (pravastatin) and bisphosphonate (zoledronate) that has demonstrated an increase in mean survival by 1. 6 years. This is the first evidence of treatments improving survival for this fatal disease.

2451F

A pathogenic variant in *BMPR1A* results in craniofacial dysmorphism, Robin sequence, and juvenile polyposis syndrome. *W. H. Tan[1], A. H. O'Donnell Luria[1], V. L. Fox[2], J. B. Mulliken[3].* 1) Division of Genetics, Boston Children's Hospital, Boston, MA; 2) Division of Gastroenterology, Boston Children's Hospital, Boston, MA; 3) Department of Plastic and Oral Surgery, Boston Children's Hospital, Boston, MA.

Loss-of-function mutations in *BMPR1A* are found in 20 – 25% of individuals with juvenile polyposis syndrome (JPS). Facial dysmorphism and cleft palate have not been described in patients with JPS, although cleft palate has been reported in a child with a chromosome 10q22. 3-q23. 2 microdeletion that includes *BMPR1A* along with other genes [Petrova E. *Mol Syndromol*. 2014; 5:19-24]. Murine studies have shown that *Bmpr1a* is expressed in the developing primary and secondary palate, and knockout of *Bmpr1a* in palatal mesenchyme resulted in clefting of the anterior palate [Baek JA. *Dev Biol*. 2011; 350:520-31]. A more recent study revealed that the mandibular condyle was smaller in cartilage-specific *Bmpr1a* knockout mice, putatively due to down-regulation of Sox9, which controls the differentiation of chondrocytes [Jing J. *Connect Tissue Res.* 2014; 55 Suppl 1:73-8]. Herein we present a boy with facial dysmorphism, Robin sequence, and JPS due to a heterozygous mutation in *BMPR1A*. A 13 year-old boy first presented with gastrointestinal bleeding at the age of 6 years, and colonoscopy revealed multiple juvenile polyps throughout the colon and rectum. He was born with a complete cleft of the secondary palate that was repaired in late infancy. He had respiratory and feeding difficulties as a neonate, requiring treatment with continuous positive airway pressure and a feeding tube for about 5 weeks. He had minor gross motor delay in infancy, which he has since overcome; his cognitive development has always been normal. Physical examination at age 12 was significant for macrocephaly, a long and narrow face, long ears, retrognathia, and severe dental crowding with class II malocclusion. He had hyperconvex finger and toe nails; his limbs were otherwise normal. He had epispadias. Brain MRI and echocardiogram were both normal. Sequencing of *SMAD4* and *BMPR1A* revealed a *de novo* heterozygous nonsense mutation in *BMPR1A* (c. 1480C>T (p. R494X)). This mutation has previously been reported in at least two individuals who both had juvenile and hamartomatous intestinal polyps. SNP-based chromosomal microarray identified a maternally-inherited 128 – 217 kb gain on chromosome 5p15. 33 that does not encompass any protein-coding genes, although there are 2 microRNAs of unknown function. We hypothesize that heterozygous loss-of-function mutations in *BMPR1A* can be associated with syndromic Robin sequence in addition to juvenile polyposis.

2452W

A new case of Kenny-Caffey syndrome type 2 with molecular confirmation. *A. Wadley, F. Ramji, E. Kalaitzoglou, S. Krishnan, B. Drabu, C. Kesserwan.* Department of Pediatrics, University of Oklahoma Health Science Center, Oklahoma City, OK.

Background: Kenny-Caffey syndrome (KCS) type 2 is a rare genetic condition characterized by impaired skeletal development, hypoparathyroidism, hypocalcemia and distinct facial features. A perinatal lethal form of the syndrome is known as Osteocraniostenosis (OCS). KCS type 2 and OCS are allelic disorders caused by germline mutations in *FAM111A.* We describe here phenotypic and molecular findings in an infant with clinical presentation consistent with KCS type 2. **Case report:** The patient was born at 37 weeks gestational age to healthy non-consanguineous parents. Prenatal ultrasound showed intrauterine growth retardation, cloverleaf skull and shortened long bones. At birth the patient had a large anterior fontanelle, small palpebral fissures, bitemporal narrowing and supra-auricular bulging. Growth parameters were greater than 2 standard deviations below the mean. The patient's skeletal survey showed hypomineralization of calvarium but normal density of the long bones. At one month of age he developed hypocalcemia-induced seizures and was diagnosed with primary hypoparathyroidism and anemia. He responded well to calcium replacement. KCS type 2 was clinically suspected and DNA was sent for *FAM111A* mutation testing. **Results:** Sanger sequencing in a CLIA-approved lab returned a c. 1012A>G transition in exon 6 of the *FAM111A* gene (NM_001142520). Parental testing showed wild type alleles for the *FAM111A* gene. This change has been previously reported as a *de novo* mutation in one individual with OCS who died at 2 months of age. The *de novo* occurrence of this change in an affected individual suggests that this change is likely pathogenic. The patient is currently 6 months and his developmental milestones are appropriate for his age. Since the hospitalization at one month old he has had one other occurrence of hypocalcemia and seizure activity in the setting of a parainfluenza viral infection. He had a full recovery with no sequelae. He continues to have poor growth parameters and remains on calcium and activated vitamin D replacements. **Conclusion:** The clinical and radiological features in this patient are consistent with KCS type 2 rather than with OCS. We conclude that studying additional patients with the condition is needed to better address genotype-phenotype correlation.

2453T

Subtle Facial Dysmorphism, a Variable Phenotype and Negative 4 Gene Testing in Young Adults with Holoprosencephaly. *K. Weiss, M. Guillen Sacoto, C. Hadsall, R. Hart, E. Roessler, P. Kruszka, M. Muenke.* Medical Genetics, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA.

Background Holoprosencephaly (HPE) is a structural anomaly of the brain in which there is failed or incomplete separation of the forebrain, caused by chromosomal anomalies, teratogenic exposure, part of a syndrome, or due to mutations in more than 12 known HPE-associated genes of which 4 genes (*SHH*,*ZIC2*,*SIX3*, and*TGIF*) are seen most frequently (Solomon, 2010) . The common facial features include hypotelorism, midface hypoplasia, cleft lip and/or palate, and a single maxillary central incisor. HPE occurs in up to 1 in 250 conceptions, but in only approximately 1 in 10,000 liveborn infants (Leoncini, 2008). It is extremely rare that individuals with HPE will survive into adulthood and little is known about the phenotypic characteristics of these older patients. The aim of this study is to evaluate the phenotype and genotype of adolescents and adults with HPE. MethodsWe evaluated 7 families of individuals with HPE, above the age of 16 years ascertained through the international registry of HPE. All patients were consented to our institutional review board approved protocols for clinical and genetic evaluation at the National Institutes of Health Clinical Center. All patients underwent physical exams, labs, neuropsychological evaluations, ophthalmic evaluations, brain MRI and genetic testing for the 4 most common HPE genes. Results8 individuals from 7 families were evaluated, 3 males, 5 females, aged 16-33 years (mean 21). All had negative 4 gene testing. The MRI scans showed semilobar HPE in 6 of the patients and lobar HPE in 2. All had only subtle dysmorphic features and none of the hallmark facial features of HPE. The neuropsychological development was variable: one individual walks independently and communicates verbally, one is non-ambulatory and communicates verbally and 6 are non-verbal and non-ambulatory. 4 patients had recurrent seizures, 3 patients are G-tube fed due to recurrent aspiration and all of the patients had scoliosis. Only one patient had overt diabetes insipidus. The most common ophthalmologic findings were optic nerve abnormalities and cortical visual impairment. ConclusionsWe evaluated a unique and rare group of young adult with HPE. The phenotype of adults with HPE supports the negative correlation between severe facial findings and survival but do not correlate with the HPE type on imaging. The patients we evaluated had negative 4 gene testing and we speculate that patients who survive into adulthood have a different etiology.

2454F

CASK mutation identified by whole exome sequencing in a patient that expands the clinical spectrum for MICPCH syndrome. *J. Zhao[1], S. Tang[2], D. Schuessler[3], N. Dosa[1], RR. Lebel[1].* 1) Center for Development, Behavior and Genetics, SUNY Upstate Medical University, Syracuse, NY; 2) Ambry Genetics, Aliso Viejo, CA; 3) Community Health Center, Gouverneur, NY.

A 23-year-old mixed-European female with no family history of intellectual disability presented with multiple malformations and developmental delays. She exhibits microcephaly, severe intellectual disability, dyspraxia, congenital quadriplegia, dystonia of the upper extremities, spasticity, and scoliosis. Brain imaging revealed pontine and cerebellar hypoplasia with intact corpus callosum. We noted down-slanting palpebral fissures, midface hypoplasia, high-arched palate, dental crowding, large tongue, and long narrow asymmetric face. Chromosome analysis, metabolic testing, and microarray all revealed no abnormalities. Whole exome sequencing revealed c. 2065A>T, a single nucleotide change in the *CASK* gene, which is located on chromosome Xp11. 4 and encodes for a calcium/calmodulin-dependent serine protein kinase. This protein is essential in synaptic function and brain development. The *de novo* nonsense mutation truncates the CASK protein, which is likely the etiology of the patient's adverse phenotype. The major features in our patient resemble those reported in MICPCH syndrome (microcephaly with pontine and cerebellar hypoplasia). Since MICPCH syndrome is a rare X-lined dominant disorder (OMIM #300749) associated with mutations in the *CASK* gene, we believe our patient expands the phenotypic profile of *CASK* mutations.

Presentations	London Dysmorphology Database	Moog et al. 2011	Patient
Microcephaly	+	+	+
Pontocerebellar hypoplasia	+	+	+
Intellectual disability	+	+	+
Seizures/abnormal EEG	+	+	+
Long philtrum	+	+	+
Epicanthic folds	+	+	+
Large Ear	+	+	+
Prominent nasal bridge	+	+	+
Hyoptonia	+	+	-
Spasticity	+	+	-
Scoliosis	+	-	+
Sensorineural deafness	-	+	+
Hypermetropia	-	-	+
Midface hypoplasia	-	-	+
Dental crowding	-	-	+
Hypersomnolence	-	-	+

2455W

***PNPLA6* mutations – a rare genetic cause of syndromic hypogonadotrophic hypogonadism.** *A. L. Frederiksen[1], J. H. Langdahl[1, 2], K. Brusgaard[1], N. Nguyen[3], C. B. Juhl[2].* 1) Dept. of Clinical Genetics, Odense University Hospital, Odense C, Denmark; 2) Dept. of Endocrinology, Hospital of South West, Esbjerg, Denmark; 3) Dept. of Radiology, Odense University Hospital, Odense C, Denmark.

Background: Congenital hypogonadotrophic hypogonadism (HH) includes clinically and genetically heterogeneous conditions including both syndromic and non-syndromic forms. A growing number of genes are recognized as responsible of HH though, the individual genetic abnormalities are infrequent. Boucher-Neuhäuser syndrome (BNS) which describes the clinical triade of early-onset ataxia, hypogonadotropic hypogonadism and chorioretinal dystrophy has just recently been linked to autosomal recessive *PLPNA6.* However, complicated spastic paraplegia, type 39 (SPG39) was until now considered to be the prominent phenotype of mutations in *PLPNA6.* **Case**: We describe two siblings a male and female clinically diagnosed with BNS at age 56 and 64 carrying a known*PNPLA6* mutation and a novelvariantThe index patient a male age 56 was diagnosed with osteoporosis and refereed to the department of endocrinology. The previous medical history exposed that he had been diagnosed with HH at age 17 years and treated with testosterone replacement therapy. On clinical examination he presented with ataxic gait, dysarthria and nystagmus. A detailed family history was recorded. The patient was number 4 of 7 siblings. The oldest age sister age 64 did not have any children and was reported with osteoporosis and ataxic gait while the remaining five siblings, who all had children, were without any symptoms. The oldest sister was then referred for clinical evaluation. She reported a tendency to fall since childhood and had been diagnosed with ataxia. Due to primary amenorrhea she was diagnosed with HH at age 25 and treated with hormone substitution however, she discontinued due to side effects. Age 58 she had been diagnosed with osteoporosis. A subsequent ophthalmological evaluation exposed both patients with chorioretinal dystrophy and thus both fulfilled the criteria for BNS. **Result:** A genetic screening identified two heterozygous variants *PNPLA6*, c. 2944_2947dup, (p. Arg983fs), and c. 3932G>A, (p. Arg1311Gln) in both patients. The c. 3932G>A has not previously been reported. To support a potential significance of the mutations their five asymptomatic siblings were subsequently tested. All five were either non-carriers or heterozygous carriers of only one of the variants. The results indicate the mutations to be pathogenic. **Conclusion:** This case illustrates *PLPN6* mutations as a rare genetic cause of HH syndrome and to the best of our knowledge presents a novel *PLPLN6* mutation.

2456T

Point mutation in the proline dehydrogenase gene (*PRODH)* plus a microdeletion in the proximal part of DiGeorge syndrome chromosomal region (DGCR) in juvenile onset epileptic encephalopathy – a case report. *C. Lau, G. Golas, B. Pusey, W. Bone, E. Valkanes, T. Markello, D. Adams, W. Gahl.* NIH Undiagnosed Disease Program, National Institutes of Health, Bethesda, MD.

Introduction: Epileptic encephalopathies are severe brain disorders involving epileptic electrical discharges that contribute to progressive psychomotor dysfunction. Clinical manifestation often includes multiform and intractable seizures, behavioral and neurological impairments and sometimes early death. **Case Presentation:** The proband is a 22-year-old (y/o) male who was born via induced vaginal delivery from non-consanguineous parents with normal early developmental milestones. His first presenting symptoms occurred at age 7 y/o with attention difficulties, anxiety, hyperactivity and oppositional-defiant behaviors. Subsequently tonic clonic seizures (1-2/month) started, followed by progressive neurocognitive decline, progressive motor apraxia and retinopathy. Poor appetite led to failure to thrive. Neuroimaging detected cerebral and cerebellar atrophy. Exome sequencing detected a p. L289M variant in *PRODH* and SNP microarray analysis uncovered an approximately 100 kb microdeletion in the proximal part of DGCR, harboring *PRODH*, *DGCR2*, *DGCR5* and *DGCR6*. Subsequently, heightened proline and glutamine levels in CSF were found. In addition, two deleterious appearing compound heterozygous alleles in *PDE6B* (c. 1401+5G>A and c. 2470_2478del9) were detected. **Discussion:** Homozygous or compound heterozygous variants in *PRODH* have been shown to cause hyperprolinemia type 1 (HP1). However, clinical relevance of HP1 has been unclear since the clinical manifestations reported ranges from increased proline level with no other physical findings to severe neurologic manifestations including epilepsy and mental retardation. This specific p. L289M variant that moderately decrease PRODH activity (<30%) was identified in 2 of 63 unrelated patients with schizophrenia. In our patient and his affected brother, it is in trans with the deletion of *PRODH*. We believe this resulted in the heightened proline and glutamine levels, the contributing cause of the epileptic encephalopathy observed, and additionally the variants in *PDE6B* led to the retinopathy finding. **Conclusion:** We propose that a combination of the previously-reported *PRODH* mutation (p. L289M) plus a hemizygous deletion of *PRODH* contribute to and/or cause the epileptic encephalopathy and separately the *PDE6B* deleterious variants contributed to the retinopathy observed in these brothers. Cases such as this may shed light on modifiers that contribute to the wide spectrum of epileptic encephalopathies.

2457F

Mutation in *TWIST2* gene and its relation to a new clinical finding - lipodystrophyc habitus - in Ablepharon Macrostomia Syndrome. *E. M. S. M. M. Souza[1, 3, 4], K. C. Peronni[3, 4], J. R. Plaça[3, 4], G. A. Molfetta[1, 3, 4], S. C. S Carvalho[1, 3, 4], A. A. Marques[3, 4], L. F. Araújo[1, 3, 4], W. A. Silva-Jr[1, 3, 4], V. E. F. Ferraz[1, 2].* 1) Department of Genetics, Ribeirao Preto Medical School, University of Sao Paulo, Ribeirao Preto, Sao Paulo, Brazil; 2) Department of Medical Genetics, Clinics Hospital of the Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil; 3) Center for Cell-Based Therapy (CEPID/FAPESP); National institute of Science and Technology in Stem Cell and Cell Therapy (INCTC/CNPq), Regional Blood Center of Ribeirão Preto, Ribeirão Preto, Sao Paulo, Brazil; 4) Center for Medical Genomics (HCFMRP/USP); Center for Integrative Systems Biology (CISBi - NAP/USP), Ribeirao Preto, Sao Paulo, Brazil.

Ablepharon Macrostomia Syndrome (AMS) is a rare condition mainly characterized by absent or hypoplastic eyelids, absent eyebrows and eyelashes, and macrostomia. Only 17 cases of the syndrome have been reported since its first description and all the cases are sporadic except for two reports of familial recurrence. The aim of this study was to clinically reevaluate one family, consisting of recurrence among sisters and a father with mild phenotype, and to identify the gene mutation responsible for the syndrome in these patients. To achieve this goal exome sequencing was performed in five members of this family (father, mother, two affected sisters, and one unaffected sister) and all mutations in the protein-coding portions of the genome were identified. The mutations were filtered and evaluated by inheritance, pathogenicity, relevance to the AMS phenotype and by gene function. The mutation c. 223G>A (p. E75K) on the *TWIST2* gene was the only one segregating among the affected members, which encoded a transcription factor with regulatory function on dermal development. This mutation was validated by Sanger Sequencing and is predicted as pathogenic, causing the change of glutamic acid, a small and negatively charged amino-acid, to lysine, a larger and positively charged amino-acid. This change happens in position 75 of the protein, inside the basic Helix-Loop-Helix domain, which could disrupt dimerization and DNA binding. Additionally, we report the presence of a new feature in the two sisters with AMS – the loss of adipose tissue and fat accumulation in the lower back, common features in lipodystrophy. *TWIST2* mutations have already been associated with Familial Partial Lipodystrophy in mouse knockout studies, suggesting that *TWIST2* may play a role in the lipodystrophic features observed in affected sisters. Finally, we suggest that the phenotype heterogeneity associated with mutations in *TWIST2* can occur due to the wide regulatory function of this transcription factor in several developmental genes.

2458W

Autosomal recessive truncating mutations of *KLHL7* cause a Bohring-Opitz like syndrome. *J. Thevenon[1,2], S. Bigoni[3], J. St-Onge[1,2], G. Parmeggiani[3], A. Ferlini[3], G. Garani[4], E. Ballardini[4], C. Gilissen[5], B. van Bon[5], R. Acuna-Hidalgo[5], C. Thauvin-Robinet[1,2], A. Bohring[6], J. B Rivière[1,2], H. G Brunner[5], L. Faivre[1,2], A. Hoischen[5].* 1) FHU-TRANSLAD, Université de Bourgogne/CHU Dijon, Dijon, France; 2) Équipe EA42271 GAD, Université de Bourgogne, Dijon, France; 3) UOL of Medical Genetics, Department of Reproduction and Growth and Department of Medical Science, University-Hospital S'Anna, Ferrara, Italy; 4) Neonatal Intensive Care Unit, Department of Reproduction and Growth, University-Hospital S'Anna, Ferrara, Italy; 5) Radboud University Medical Center, Department of Human Genetics, Nijmegen, The Netherlands; 6) Department of Pediatrics, Ostholstein Kliniken, Eutin, Germany.

Bohring-Opitz syndrome (BOS) is characterized by distinctive craniofacial features, distal joint contractures, pre- and post-natal growth retardation, failure to thrive, severe developmental delay, brain malformations. Most typical cases are caused by *ASXL1 de novo* mutations. BOS-like was also described in cases with heterozygous *ASXL3* truncating mutations, suggesting genetic heterogeneity. We identified a case born from non-consanguineous parents, with a personal history of prenatal growth retardation and hydramnions. At birth, global hypotonia with poor mobility, reducible distal and proximal joint contractures suggesting the BOS posture were noted. Main facial dysmorphism were broad forehead with nevus flammeus, hypertelorism, prominent eyes, anteverted nares, micro-retrognathia. Evolution was marked by severe feeding problems, seizures, perceptive deafness, recurrent pulmonary infections. Brain MRI showed dysmyelination and a thin corpus callosum. The patient was negative for *ASXL1* mutation screening. Whole-exome sequencing detected a candidate homozygous non-sense change of *KLHL7*. International data-sharing identified an additional family carrying a homozygous splice-site mutation of *KLHL7*. In this family, two siblings were affected by a previously undiagnosed condition. Phenotypic manifestations included pre-natal growth retardation, hydramnios, distal joint contractures, severe epileptic disease, failure to thrive and feeding problems, recurrent pulmonary infections, hypertrichosis. Facial dysmorphism was very similar to patient from family one. Brain MRI revealed agenesis of corpus callosum, fusion of thalami and frontal region in both children. Additionally, sibling one had a left temporal cortical dysplasia and sibling two had a cleft palate and an atrio-septal defect. Both children died from acute respiratory distress in their second year. *KLHL7* encodes a BTB-kelch protein implicated in cellular cycle and ubiquitin pathway. Rare families with heterozygous missense changes in a specific domain of KLHL7 were associated with an autosomal dominant retinis pigmentosa (RP), with a suspected dominant negative mechanism. Conversely, none of the reported patients with truncating variant had signs of RP. Truncating events in *KLHL7* are rare in ExAC databases (0. 02%), with no homozygous individual reported suggesting a dosage-sensitivity. In conclusion, we suggest that autosomal recessive truncating mutations of *KLHL7* can cause BOS-like.

2459T

Identification of Mutation in Kelch-Like Family Member 15 Gene in a Patient with Abnormal Cortical Development and Hypergonadotrophic Hypogonadism. *G. Yesil[1], E. KARACA[2], Y. BAYRAM[2], T. GURAN[3], D. PEHLIVAN[2], Z. COBAN AKDEMIR[2], T. GAMBIN[2], S. JHANGIANI[4], D. MUZNY[4], R. GIBBS[4], J. R. LUPSKI[2,4,5,6], The Baylor-Hopkins Center for Mendelian Genetics.* 1) Department of Medical Genetics, Bezmialem University, Istanbul, Turkey; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 3) Department of Pediatric Endocrinology, Marmara University, Istanbul, Turkey; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston TX, USA; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 6) Texas Children's Hospital, Houston, TX, USA.

Whole-exome sequencing (WES) is a new groundbreaking technology that has changed the course of clinical/human genetics. As a more efficient and cost effective method, WES is now widely used as a diagnostic tool for identifying challenging genetic syndromes. Here we report a female patient with clinical diagnosis of Aicardi syndrome (MIM#304050) whom we identified heterozygous missense mutation in Kelch-Like Family Member 15 (*KLHL15*) gene by using WES. Mutations in *KLHL15* were shown in patients with either isolated intellectual disability (ID) or ID together with structural brain malformations and epilepsy. The present case had some features of Aicardi syndrome such as microcephaly, optic atrophy, convulsions, diffuse cerebral atrophy, agenesis of corpus callosum and facial dysmorhism with an upturned nose and hypertelorism. Hypergonadotrophic hypogonadism and thyroid hypoplasia were additional features. *KLHL15* is located on chromosome Xp22. 11, within the candidate locus for Aicardi syndrome. In conclusion, our clinical and genomic findings, together with previous clinical data in the literature, suggest that mutations of *KLHL15* might be associated with Aicardi syndrome, or a new neuroendocrine syndrome. Additional reports of individuals with a similar phenotype and animal studies will be crucial to clarify the potential role of *KLHL15* gene.

2460F

Clinical characteristics of fragile X syndrome for genetic counseling: First reported in Thai patients. *C. Charalsawadi[1], S. Jaruratanasirikul[2], J. Worachotekamjorn[2], P. Limprasert[1].* 1) Department of Pathology, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla, Thailand; 2) Department of Pediatrics, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla, Thailand.

Fragile X syndrome (FXS) is the most common inherited mental retardation, caused by CGG trinucleotide repeat expansion in the *FMR1* gene on chromosome Xq27. 3. About 7% of Thai boys with developmental delay are affected with FXS. This is the first systematic study in Thailand. We reviewed a retrospective cohort of Thai patients with FXS between the years of 1993 through till the year 2015. A total of 128 patients (104 males and 24 females) were diagnosed with FXS, using standard molecular methods. To determine frequencies of common clinical characteristics, based on the 5-item clinical checklist developed by Limprasert et al (J Med Assoc Thai. 2000;83:1260-6), 56 prepubertal and 3 post pubertal boys with FXS, who were the first affected persons in their families, were evaluated. Age at diagnosis ranged between 18 and 166 months of age (mean 80, SD 40). The majority of patients were diagnosed between the ages of 5 and 10 (48%) and under 5 years old (32%). The most common referral reason was developmental delay. The average age of first understandable word and unassisted walking were 31 months (SD 11) and 21 months (SD 6), respectively. X-linked inheritance pattern was observed in 14% of the 57 families. Out of 26 patients, IQ ranged between 30 and 64 (mean 44, SD 9). About 69% of the patients had an IQ within a moderate range. For prepubertal boys, attention deficit and/or hyperactivity were the most common clinical characteristics (96%). Prominent ears as well as elongated face were observed in 78% and 48%, respectively. About 52% of prepubertal boys had macroorchidism. Other common characteristics e. g. hyperextensible MP joint (n=22), autistic features (n=15), social anxiety/ shyness (n=13), flat feet (n=13), aggressive/ self injury (n=12), and high-arched palate (n=11) were observed. To identify rare characteristics, we examined 70 males and 7 females with available medical records. Rare characteristics e. g. seizure (n=5), Prader-Willi-like phenotype (n=3), Sotos-like phenotype (n=2), strabismus (n=2), astigmatism and hypermetropia (n=1), bilateral ptosis (n=1), cleft palate (n=1), bilateral conductive hearing loss (n=1), tetralogy of Fallot (n=1), medulloblastoma (n=1), unilateral macroorchidism (n=1), precocious puberty (n=1), primary amenorrhea (n=1), and hypothyroidism (n=1) were recognized. In conclusion, our study provides insight into clinical characteristics of FXS in Thai patients, which is necessary for genetic counseling in Asian ethnicities.

2461W

Mutations in the PI4KA-TTC7A-EFR3B Signaling Pathway Result in a Progressive Lethal Syndrome. *R. C. Lombardo, L. Wung, S. A. Kocoshis, A. H. Filipovich, C. A. Valencia, R. L. McMasters, K. Zhang, T. A. Kalfa, K. Risma, R. J. Hopkin.* Cincinnati Children's Hospital and Medical Center, Cincinnati, OH.

The PI4KA-TTC7A-EFR3B pathway plays a key role in critical intracellular functions and has recently emerged as a cause of human disease. We describe a fatal clinical outcome believed to be secondary to mutation in a gene affecting this pathway. Patient was a female infant, prenatal testing identified a single umbilical artery, symmetric IUGR, and complex congenital cardiovascular abnormalities. Clinical course was marked by progressive multiple organ failure with development of a dyserythropoietic anemia and immunologic abnormalities. Gastrointestinal evaluation revealed eosinophilic enterocolitis with prominent epithelial apoptosis, limited clinical response to oral corticosteroids, and progression to intestinal failure. Microarray and metabolic studies were normal, whole exome sequencing identified heterozygous mutations in *PI4KA*. At 10 months of age, she succumbed to a culture negative sepsis- like inflammatory cascade and died of multiple organ failure. Prior studies have demonstrated that genes contributing to the PI4KA-TTC7A-EFR3B pathway are critical to intracellular function and highly conserved among vertebrate species. Cellular models suggest EFR3B tethers TTC7A to the plasma membrane which in turn, plays a primary role in proper localization of phosphatidylinositol 4-phosphate (PtdIns4P). Alteration of this complex results in loss of plasma membrane integrity, cell polarity, and interferes with normal vesicular trafficking, cellular adhesion, RhoA-mediated cytoskeletal organization, and phosphatidylinositol signaling. Previous reports have described a small number of patients with TTC7A mutations with severe gastrointestinal manifestations, variable immunodeficiency, and death in early infancy1. Animal models exhibit small head size, decreased cerebral volume, and cardiovascular developmental abnormalities2 and conditional knock out mice develop intestinal failure, diffuse systemic inflammation, and cardiovascular collapse following pathway inhibition3. We propose that alterations of PtdIns4P signaling results in a recognizable syndrome characterized by congenital heart defects, failure to thrive, very early onset intestinal failure, severe immunodeficiency, and progressive multiple organ failure with lethal outcome. 1. Gastroenterology. Avitzur et al. 146: 1028-10392. Journal of Cell Science. Hui et al. 122: 4303-43103. Journal of Biological Chemistry. Bojjireddy et al. 289(9): 6120-6132.

2462T

Mutations in SEC61A1 cause autosomal dominant interstitial kidney disorder associated with anemiaand growth retardation. *C. Golzio[1], N. Bolar[2], C. Van Hemelrijk[4], A. Hoischen[3], J. Huyghe[2], A. Raes[4], E. Matthys[5], E. Sys[6], M. Gubler[7], C. Antignac[7], M. Azou[8], G. Van Camp[2], S. Kmoch[9], A. Bleyer[10], J. Vande Walle[4], G. Mortier[2], H. Brunner[3], L. Van Laer[2], N. Katsanis[1], B. Loeys[2,3].* 1) Center For Human Disease Modeling Duke Univ Medical Center, Durham, NC; 2) University of Antwerp, Antwerp, Belgium; 3) Radboud University Medical Center, Nijmegen, Netherlands; 4) University of Ghent, Ghent, Belgium; 5) Sint-Jan Hospital, Brugge, Belgium; 6) Sint-Lucas Hospital, Brugge, Belgium; 7) Inserm U983, Paris, France; 8) Damian Hospital,Ostend, Belgium; 9) Charles University, Prague, Czech Republic; 10) Wake Forest School of Medicine, Winston-Salem, NC, United States.

Autosomal dominant tubulo-interstitial kidney disease refers to a group of disorders characterized by progressive loss of kidney function and the eventual need for dialysis and kidney transplantation. While mutations in the *UMOD*, *MUC1*, and *REN* genes have been identified as the primary causes of this disorder, there remain a number of families in whom the genetic cause has not been identified. Here were report a three-generation family with autosomal dominant progressive chronic kidney disease associated with congenital anemia and intrauterine growth retardation. Ultrasound examinations revealed small dysplastic kidneys without cysts, and a kidney biopsy revealed tubular atrophy with secondary glomerular sclerosis. We performed genome-wide linkage analysis and identified a candidate region with a maximum LOD-score of 2. 7 on chromosome 3q. Combining whole exome sequencing data with our critical interval, we identified a variant p. Thr185Ala change in *SEC61A1*. *SEC61A1* encodes the alpha-subunit of the SEC61 complex, responsible for the translocation of proteins across the endoplasmic reticulum membrane. Suppression of sec61a1 in zebrafish embryos led to an absence of convolution of the pronephric tubules, whereas the pronephric ducts were unaffected, a phenotype similar to the tubular atrophy seen in our patients. This phenotype could be rescued by wild type *SEC61A1* mRNA, but not with mRNA encoded from the p. Thr185Ala allele, suggesting that the variant is a loss of function. In addition, we identified another variant in*SEC61A1*(p. Val67Gly) in a second family with similar phenotypes; the p. Val67Gly allele also appeared to be pathogenic in ourin vivocomplementation assay. Taken together, our genetic findings and the functional studies support *SEC61A1* as a causal gene for a novel, dominant syndromic form of progressive chronic kidney disease with tubular atrophy.

2463F

Siblings with Harderophorpheria: Early signs mimic infantile osteopetrosis. Revisiting a clinical diagnosis 15 years later with the aid of whole exome sequencing. *L. Dupuis[1], E. Bareke[2], T. Hartley[3], P. Kannu[1], K. Boycott[3], M. Yasuda[4], R. J. Desnick[4], D. O. Doheny[4], R. Mendoza-Londono[1], Care for Rare Consortium.* 1) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children and University of Toronto, Toronto, Ontario, Canada; 2) Department of Human Genetics, McGill University and Génome Québec Innovation Centre, Montréal, Quebec, Canada; 3) Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada; 4) 4Icahn School of Medicine at Mount Sinai, Department of Genetics & Genomic Sciences, New York, NY, United States of America.

The advent of whole exome sequencing (WES) has revolutionized gene discovery and led to the identification of atypical phenotypes for well known syndromes. Our case illustrates the phenotypic overlap between very different syndromes, harderoporpheria and infantile osteopetrosis, which was only resolved by WES. We report two siblings, both born with severe hepatosplenomegaly, jaundice, thrombocytopenia and anemia. The eldest sibling had increased bone density on radiographs and was diagnosed with infantile osteopetrosis. Both children responded well to bone marrow transplant (BMT). Molecular testing was negative for genes involved in infantile osteopetrosis*: CLCN7, TCIRG1, OSTM1, TNFSF11A* and *PLEKHMI1*. Many years later, the family was offered whole exome sequencing to clarify the underlying molecular etiology. The siblings were found to harbor a homozygous mutation (p. K404E) in the *CPOX* gene confirming a diagnosis harderoporphyria. No mutations were found in any known genes causing infantile osteopetrosis. Defects of heme biosynthesis enzymes result in porphyrias. Harderoporphyria is caused by homozygous mutations in the *CPOX* gene which encodes coproporphyrinogen oxidase. Individuals with harderoporphyria exhibit neonatal hyperbilirubinemia, hemolytic anemia, hepatosplenomegaly and photosensitivity. Infantile osteopetrosis presents with increased bone density, reduction of bone marrow spaces leading to anemia, hepatosplenomegaly, cranial nerves compression and severe growth failure. It is a lethal condition, but responds well to BMT. Given that both children had undergone BMT, their own hematopoietic precursors would have been replaced by those of the donor. Thus, they would not be expected to manifest a harderoporphyria phenotype. To our knowledge, there have been no reports of increased bone mineral density in patients with harderoporphyria. An initial clinical presentation suggesting a diagnosis infantile osteopetrosis was only challenged after the WES data returned. Since the siblings received BMT in the first year of life, it remains unknown whether increased bone density was a persistent finding. This case report illustrates the clinical utility of WES and the importance of establishing a definitive molecular diagnosis for appropriate genetic counselling and medical management.

2464W

Improved detection of FLCN mutations in patients with BHDS. *J. R. Toro[1, 2], B. Friedman[3], S. Bale[3].* 1) National Cancer Institute, Maryland, MD; 2) Dermatology Department, Veterans Affairs Medical Center, Washington, DC, USA; 3) GeneDx, BioReference Laboratories, Inc. , Gaithersburg, MD Maryland, USA.

Birt-Hogg-Dubé syndrome (BHDS) (MIM 135150) is an autosomal dominant predisposition to the development of follicular hamartomas (fibrofolliculomas), lung cysts, spontaneous pneumothorax, and kidney neoplasms. Germline mutations in FLCN are associated with the susceptibility for BHDS. To date 153 FLCN germline mutations have been reported in the online Folliculin sequence variation database. Objective: To characterize methods for improved FLCN mutation detection and novel mutations. Methods: Initial screening was conducted with direct bidirectional DNA sequencing of the coding regions and splice sites of exons 4-14 of FLCN. If no mutation was identified by sequencing analysis, large intragenic insertion and deletion mutations were screened by RQ-PCR and targeted arrays comparative genomic hybridization with exon-level resolution. Results: The FLCN mutation detection rate by direct sequencing was 89 percent. We detected 56 unique novel FLCN germline mutations: 22 deletions, 11 insertions, 13 missense,7 nonsense, 2 splice site and 1 deletion/insertion. To date only eight large unique intragenic mutations have been reported. We identified one whole gene FLCN deletion and eleven unique large FLCN intragenic deletions: four involving exon 1, one in exon 6 and six encompassing: exons 1-6, exons 2-5, exons 2-13, exons 6-14, exons 7-8 and exons 10-14. Including this report, to date there are 214 unique FLCN mutations identified: 76 deletions, 32 insertions, 99 substitutions and 7 deletion/insertion. A comprehensive worldwide review of published FLCN mutations and current ongoing efforts to detect novel BHDS causing mutations will be discussed. Conclusion: A systematic approach combining accurate and sensitive methods to detect FLCN mutations provides evidence that most patients with BHDS have mutations in FLCN.

2465T

A compound-heterozygous mutations in patients with autosomal-dominant Marfan syndrome. *P. Cibulkova[1,2], R. Richterova[1,2], M. Dvorakova[1,2], R. Krenkova[1,2], E. Augste[1,2].* 1) Laboratories AGEL, Novy Jicin, Czech Republic; 2) Education and Research Institute AGEL, Prostejov, Czech Republic.

Objectives: Marfan syndrome (MFS) is known as an autosomal dominant genetic disorder caused by mutation in *FBN1* gene, resulting in defective glycoprotein fibrillin-1. It is a multisystem connective tissue disease, mainly involving the cardiovascular, ocular and skeletal systems. We describe four cases of rare compound-heterozygous MFS patients. Three probands were from families where only one of the parent was affected and the second mutation was *de novo*, in two cases we did not have enough information if mutations are in cis or trans position. The last reported case of two mutations in the *FBN1* gene is from family where both mutations are in one *FBN1* allele that was inherited across three generations. **Methods:** We performed molecular analysis of all exons of *FBN1* gene to identify pathogenic mutations in unrelated patients with clinical diagnosis of MFS. DNA was isolated from peripheral blood. 750 patients were analyzed by SSCP method and 350 by NGS (Junior 454, Roche). MLPA analysis was also performed. All detected sequence variants were tested by *in silico* prediction tools for its effect on protein function. **Results:** We identified 129 mutations by SSCP and 81 by NGS in *FBN1* gene, some of them have already been published. Reported four cases of compound-heterozygous MFS demonstrate incomplete dominance of detected *FBN1* mutations. The phenotypes of proband´s heterozygous family members were mild, did not cause the comlpete MFS and they do not meet the Ghent clinical criteria for MFS. They have not been previously tested or suspected for MFS. The phenotypes of the compound-heterozygous probands were much more severe and had required cardiovascular, surgical or ophthalmological solution early in childhood, but did not lead to lethally affected individual. **Conclusion:** Genetic counselling in affected families is complicated after discovery of compound-heterozygous MFS due to autosomal-dominant heredity of MFS and intrafamilial variable phenotype manifestation. The fact that compound-heterozygous MFS patients had not fatal clinical manifestations after birth signify, that some mutations behave semi-dominantly rather than purely dominantly. However they are more severely affected than their simple heterozygote parents or relatives.

2466F

GenIDA: an international registry of individuals affected by monogenic forms of intellectual disability or autism and a families and professionals social network, to collect medically relevant information and gene specific natural histories. *J. -L. Mandel[1,2,3], F. Colin[1], T. Mazzucotelli[4], P. Parrend[4,5], A. Deruyver[4], T. Kleefstra[6], D. Koolen[6].* 1) Translational medicine and neurogenetics dept. , Institut de Génétique et de biologie Moléculaire et cellulaire (IGBMC) - INSERM U964 - CNRS UMR7104 - University of Strasbourg, Illkirch, France. (jlmandel@igbmc. fr and colin@igbmc. fr); 2) Genetic diagnostic laboratory, CHU Strasbourg, France; 3) Collège de France, Chaire de Génétique humaine, Paris, France; 4) Engineering, informatics and imaging sciences laboratory (ICube) - CNRS UMR 7357 - ENGEES - INSA of Strasbourg, Illkirch, France; 5) ECAM, Strasbourg-Europe, Schiltigheim, France; 6) Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands.

Intellectual disability (ID) and autism spectrum disorders (ASD) are overlapping disorders that constitute a major public health, social and educational problem because of their cumulated frequency (about 2. 5%), and constitute a great burden for affected individuals and families. ID/ASD are characterized by a striking genetic heterogeneity that underlies a phenotypic heterogeneity in severity and in associated medical problems. Progress in genome analysis has allowed the identification of many recurrent CNVs and of more than 600 genes implicated in monogenic forms of ID/ASD. An increasing number of genetic diagnoses are made in individuals with ID/ASD but the genetic heterogeneity renders extremely difficult the determination of genotype-phenotype correlations and natural history. Symptomatic treatments for comorbidities are proposed with limited opportunities to assess their efficacies or potential adverse effects. We have initiated the development of an alternative database model for specific genetic causes of ID/ASD, organized in a social network format whereby most clinical information (HPO-compatible) is entered and updated by the family of the affected individual based on wide range of structured questionnaires translated in different languages. We have recently merged our questionnaire with the one of a similar project, Syndrome Monitor, initiated at Radboud U. Nijmegen. Contacts between families affected by the same genetic cause will be possible if they wish so, creating gene or CNV specific social networks to which interested professionals could be associated, akin to disease specific patients associations. Anonymized summary data will be accessible to families and to professionals. Interested professionals will be able to submit specific questionnaires or recruit patients for clinical studies. Such proposals will be screened by a steering committee (professionals and patients representatives) that will control medical interest, feasibility and respect of ethical rules. We will present the structure and features of the GenIDA project (genida. unistra. fr) that is currently in its beta testing phase. This strategy aim is to collect information on natural history and comorbidities of rare monogenic forms of ID/ASD, including potential mutated-gene specific adverse effects of symptomatic drugs. It will promote families' empowerment and hopefully impact patients' care. The setting up of patient e-cohorts should favor international clinical studies.

2467W

Case report of family with Christianson syndrome and apparently novel pathogenic variant in the *SLC9A6* gene. *J. Moeschler[1], J. Filiano[2], S. Upton[1].* 1) Section of Medical Genetics, Geisel School of Medicine, Dartmouth-Hitchcock Med Ctr, Lebanon, NH; 2) Section of Child Neurology, Dartmouth-Hitchcock Medical Center, Lebanon, NH.

Christianson syndrome (OMIM#300243) is a rare X-linked intellectual disability syndrome with features similar to those seen in Angelman syndrome. First described in 1999 by Christianson AL, Stevenson RE, et al in a large South African family, there are approximately 30 patients reported in the literature. Microcephaly, impaired ocular movements, severe global developmental delay, developmental regression, hypotonia, abnormal movements, and early-onset seizures of variable types characterize the syndrome. Gilfillan GD, Selmer KK et al (2008) identified *SLC9A6* gene mutations as the cause of this XLID syndrome. The *SLC9A6* gene encodes the Na(+)/H(+) exchanger NHE6. Pescosolido MF, Stein DM (2014) described the phenotype to include early-childhood onset in boys, nonverbal status, moderate to severe intellectual disability, epilepsy, truncal ataxia, postnatal microcephaly and/or attenuation in growth of head circumference, and hyperkinetic behavior. Fewer patients are reported to develop eye movement abnormalities, developmental regression, particularly loss of independent ambulation after 10 years of age, low weight for age, and cerebellar vermis atrophy, particularly after 10 years of age. We describe a family with three affected males and two unaffected carrier females with a novel pathogenic variant in the *SLC9A6* gene. The proband was a 16-year old boy who presented to our medical center with the clinical diagnosis of Angelman syndrome. He presented with microcephaly, intellectual disability, no speech, epilepsy (staring spells and tonic-clonic seizures) and self-injurious behavior. He has one older and one younger affected brother with similar clinical presentation. All three have severe cognitive disability, seizure disorder, lack of speech. Family history describes one sister and one brother who are both unaffected. Mother is unaffected and the family history is otherwise negative. A hemizygous insertion variant was detected in exon 4 of the *SLC9A6* gene: c. 630_631insT (p. Ile211Tyrfs*53). The variant is predicted to result in reading frame shift leading to an early termination of the encoded protein. This is interpreted as a pathogenic variant. To our knowledge this variant has not been reported in the literature. The mechanism for the clinical phenotype is not yet clear. Treatment remains symptom-based.

2468T

Abnormal primary dentition : a clue for the diagnosis of the SATB2-associated syndrome. *M. Rio[1], A. Elmorjani[1], B. Kverneland[2], G. Barcia[1], A. Munnich[1], JP. Bonnefont[1], S. Hanein[1].* 1) Département de Génétique, Université Paris Descartes, Sorbonne Paris Cité, Institut Imagine, Hôpital Necker-Enfants Malades, Paris, France; 2) Service de stomatologie, Hôpital Necker-Enfants Malades, Paris, France.

The *SATB2*-associated syndrome (SAS) has been recently proposed as a new clinically recognizable syndrome that results from deleterious alterations of the *SATB2* gene in humans. We report the clinical characterization of 3 patients in whom targeted high-throughput sequencing for the diagnosis of intellectual disability identified a *de novo* mutation in the *SATB2* gene for all three patients. Patient 1 is a 8-year-old boy with intellectual deficiency and hyperactivity. He walked at 2 years. He was toilet trained at 5 years. He had delayed primary dentition. He fractured his humerus at 7 years. At examination, speech was severely impaired: he was able to speak single words but did not put words together. Nevertheless he used more than 25 signs to communicate. He had a pre auricular tag, smooth philtrum, and hirsutism. He had macrodontia and irregular shape of the primary teeth, narrow palate and dental crowding. Patient 2 is a 5-year-old boy with psychomotor delay and hyperactivity. He walked at 22 months. At examination, he had no speech and was not toilet trained. Dysmorphic features included a class II malocclusion with smooth philtrum and thin upper lip. He had narrow palate, dental crowding and irregular shape of the primary teeth. Patient 3 is a 7-year-old girl. Cleft palate was diagnosed at birth and was surgically corrected at 4 months. She walked at 21 months of age. At 2 years, she fractured her tibia. She had two febrile seizures. At 7 years, she could speak few single words and was able to use 5 signs to communicate. She had macrodontia, no permanent teeth and smooth philtrum. The 3 patients had overlapping features that support the diagnosis of the SAS: intellectual disability with limited speech development, palate abnormalities, dysmorphic features and dental abnormalities. Interestingly, anomaly of the primary teeth is a consistent finding detected in the 3 patients and a good handle for early diagnosis of SAS.

2469F

Somatic Mosaic *PIK3CA* Mutations in Chinese Patients with *PIK-3CA*-Related Overgrowth Spectrum from Hong Kong. *B. Chung[1], K. S. Yeung[1], G. Leung[1], W. L. Wong[1], C. H. Li[2], K. Y. Choi[3], E. Kuong[4], W. Chow[4], M. To[4], J. Ip[5], P. Beh[6], G. C. F. Chan[1].* 1) Paediatrics & Adol Med, The University of Hong Kong, Hong Kong, N/A, Hong Kong; 2) Paediatrics & Adol Med, Tuen Mun Hospital, Hong Kong; 3) Orthopaedics & Traumatology, Yuen Mun Hospital, Hong Kong; 4) Orthopaedics & Traumatology, Queen Mary Hospital, Hong Kong; 5) Radiology, Queen Mary Hospital, Hong Kong; 6) Pathology, The University of Hong Kong, Hong Kong,.

Somatic mosaicism of phosphatidylinositol-4,5-bisphosphate 3-kinase (*PIK3CA*) mutation, one of the genes involved in the PI3K/AKT/mTOR pathway, is associated with a group of rare asymmetrical overgrowth syndrome that is collectively named as *PIK3CA*-Related Overgrowth Spectrum (PROS). We identified six suspected patients of PROS, including two suspected cases of CLOVES Syndrome [MIM 612918], one with Cystic Hygroma, one with Klippel-Trenaunay Syndrome [MIM 149000], one with isolated macrodactyly, and one with multiple lipomatosis. We had the fresh surgical tissues from 3 patients; while for the others we had their formalin-fixed, paraffin-embedded (FFPE) tissues only. We aim to identify the mutations of *PIK3CA* in each patient. Since there are five recurrent *PIK3CA* mutations that are known to account for about 80% of PROS, droplet digital PCR was used to target these recurrent mutations in each patient. We identified three patients with p. Glu542Lys (*PIK3CA*: NM_006218. 2: c. G1624A); while the remaining had p. His1047Arg (*PIK3CA*: NM_006218. 2: c. A3140G). The percentage of mutant ranged from 3. 3% to 31. 6%. Our findings suggested that droplet digital PCR is useful in identifying mutations in PROS. Surgical debulking is now the r treatment of choice for PROS, inhibitors of PI3K/AKT/mTOR pathway may be used in controlling the progressive overgrowth in patients with the molecular confirmation of *PIK3CA* mutations in the affected tissue.

2470W

A novel *PITX1* mutation segregating with familial tibial hemimelia. *D. Ortiz[1], L. Karger[1], C. Jalas[2], J. Neidich[3], J. Juusola[3], S. Bale[3], J. E. Herzenberg[4], B. D. Webb[1].* 1) Dept. of Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Bonei Olam, Brooklyn, NY; 3) GeneDx, Gaithersburg, MD; 4) Rubin Institute for Advanced Orthopedics, Sinai Hospital of Baltimore, Baltimore, MD.

Tibial hemimelia (MIM 275220) is a rare congenital limb abnormality that occurs as an isolated anomaly in an estimated 1 in 1 million live births. Rare cases of familial tibial hemimelia have been reported with autosomal recessive or autosomal dominant inheritance. *PITX1* is a transcription factor that is known to have a critical function in hindlimb morphology, and *PITX1* haploinsufficiency has been implicated in congenital talipes equinovarus, polydactyly and additional lower limb malformations, including two cases of tibial hemimelia. We present a family with a spectrum of lower limb anomalies and a novel *PITX1* pathogenic variant (c. 442_443delGAinsCACT, p. E148HfsX21) identified via whole exome sequencing. The proband has congenital bilateral tibial hemimelia, with the right lower limb more severely affected than the left, left talipes equinovarus, congenital hip dysplasia, and bilateral pre-axial polydactyly of the feet. He was treated with serial casting and Achilles tenotomy for the left clubfoot and right knee disarticulation. Family history was significant for a younger male sibling with a nearly identical phenotype, including high grade congenital bilateral tibial hemimelia and bilateral pre-axial polydactyly of the feet, father with bilateral femoral anteversion, and a paternal male first cousin with congenital talipes equinovarus. The heterozygous p. E148HfsX21 *PITX1* pathogenic variant was identified in the proband, father, and an affected sibling. A 2 month old sibling without talipes equinovarus or tibial hemimelia was also noted to have the variant. This finding highlights the variable clinical phenotypic spectrum and incomplete penetrance of *PITX1* within families, and emphasizes its crucial role in tibial hemimelia. This case study strengthens the evidence that *PITX1* haploinsufficiency has a critical role in a variety of lower limb malformations, especially tibial hemimelia, and suggests that additional contributory factors of limb formation and development remain to be identified.

2471T

Delivery by Cesarean Section is not Associated With Decreased at-Birth Fracture Rates in Osteogenesis Imperfecta. *M. Jain[1,12], S. Bellur[1], D. Cuthbertson[2], D. Krakow[3], J. R. Shapiro[4], R. D. Steiner[5,6], P. A. Smith[7], M. B. Bober[8], T. Hart[9], J. Krischer[2], M. Mullins[1], P. H. Byers[10], M. Pepin[10], M. Durigova[11], F. H. Glorieuz[11], F. Rauch[11], V. R. Sutton[1,12], B. Lee[1,12], S. C. Nagamani[1,12], Members of the BBD Consortium.* 1) Human & Molecular Gen, Baylor college Medicine/Texas Children's Hosp, Houston, TX; 2) College of Medicine, University of South Florida, Tampa, FL, USA; 3) Departments of Orthopedic Surgery, Human Genetics, and Obstetrics and Gynecology, University of California, Los Angeles, CA, USA; 4) Department of Bone and Osteogenesis Imperfecta, Kennedy Krieger Institute, Baltimore, MD, USA; 5) Departments of Pediatrics and Molecular and Medical Genetics, Oregon Health & Science University, Portland, OR, USA; 6) Marshfield Clinic Research Foundation and University of Wisconsin, Marshfield and Madison, WI, USA; 7) Shriners Hospitals for Children, Chicago, IL, USA; 8) Division of Medical Genetics, Alfred I. DuPont Hospital for Children, Wilmington, DE, USA; 9) Osteogenesis Imperfecta Foundation, Gaithersburg, MD, USA; 10) Departments of Medicine and Pathology, Division of Medical Genetics, University of Washington, Seattle, WA, USA; 11) Department of Orthopedic surgery, Shriners Hospital for Children and McGill University, Montreal, QC, Canada; 12) Texas Children's Hospital, Houston, Texas, USA.

Osteogenesis imperfecta (OI) predisposes to recurrent fractures. The moderate-to-severe forms of OI present with antenatal fractures and the mode of delivery that would be safest for the fetus is not known. We conducted systematic analyses on the largest cohort of individuals (n=540) with OI enrolled to-date in the OI Linked Clinical Research Centers. Self-reported at-birth fracture rates were compared in individuals with OI types I, III, and IV. Multivariate analyses utilizing generalized linear modeling and backward elimination model-building were performed to assess the effect of multiple covariates including method of delivery on fracture-related outcomes. When accounting for other covariates, at-birth fracture rates did not differ based on whether delivery was by vaginal route or by cesarean section (CS). Increased birth weight conferred higher risk for fractures irrespective of the delivery method. *In utero* fracture, maternal history of OI, and breech presentation were strong predictors for choosing CS for delivery. Our study, the largest to analyze the effect of various factors on at-birth fracture rates in OI shows that CS is not associated with decreased fracture rate. With the limitation that the fracture data were self-reported, these results suggest that CS should be performed only for other maternal or fetal indications, but not for the sole purpose of fracture prevention in OI.

2472F

Congenital central hypoventilation syndrome mimicking mitochondrial disease. *K. Rojnueangnit[1,2], M. Descartes[2,3].* 1) Department of Pediatrics, Faculty of Medicine, Thammasat University, Thailand; 2) Department of Genetics, University of Alabama at Birmingham, Alabama, USA; 3) Department of Pediatrics, University of Alabama at Birmingham, Alabama, USA.

Congenital central hypoventilation syndrome (CCHS: MIM 209880) is a rare unique respiratory center defect cause hypoventilation mainly during sleep but in neonate and/or severe type presents during sleep and waking. Late-onset has been reported but is uncommon. Mutation in *PHOX2B* (paired-like homeobox) is a causative gene. Our patient is a previously healthy 18 months old who one week after routine immunization developed generalized muscle weakness, cardiomegaly, hepatomegaly, pleural effusion, and ascites. She developed acute respiratory failure requiring intubation. Her clinical symptoms improved and extubation was done at day 5. She recovered completely within 2 weeks. At that time, viral infection was suspected without any confirmation tests. Three months later, she developed a second episode of acute respiratory and cardiac failure with pulmonary hypertension. Echocardiogram showed the right ventricular hypertrophy and atrium enlargement. Intubation was required for a week. As in the first episode, she recovered within a couple weeks. Variety of tests were performed, mitochondrial disease was suspected given she had multiple organ involvement, low oxidative phosphorylation in complex I (3%) from muscle biopsy and microdeletion 270 kb of the short arm of chromosome 6 (6p25. 1), involved in *FARS2*, a nuclear mitochondrial gene. Therefore, we intensively investigated for *FARS2* mutation which was negative by sequencing the entire gene and a functional enzyme test confirmed happloinsufficiency the FARS2 protein as predicted. As polysomnography showed hypoxia, hypoventilation with hypercapnia (pCO2 range from 44 to 62 mmHg), sleep apnea could not be ruled out, therefore she was treated by noninvasive respiratory support during sleep. Her clinical symptoms, and polysomnography improved, cardiac impairment was resolved. We monitored her as a mitochondrial disease patient with a doubtful diagnosis. After several further tests, *PHOX2B* was tested for CCHS, revealing heterozygous expansion of the polyalanine to 25 repeats (p. Ala256_Ala260dup). A late-onset CCHS presented with an episodic respiratory failure was the final diagnosis. Therefore, CCHS should be included in the differential diagnosis of mitochondrial disease.

2473W

Autosomal recessive Usher Syndrome and autosomal dominant retinitis pigmentosa in a Louisiana Acadian Family. *A. Umrigar[1], A. Musso[1,2], A. Hurley[2], F. Tsien[1].* 1) Department of Genetics, LSU Health Sciences Center, New Orleans, LA; 2) Department of Audiology, LSU Health Sciences Center, New Orleans, LA.

Usher Syndrome accounts for greater than 50% of patients with combined deafness and blindness. Usher Syndrome Type 1 is inherited in an autosomal recessive manner and is characterized by congenital deafness followed by retinitis pigmentosa (RP) later in life. The most common mutation observed in the Acadian population is a G to A mutation in the USH1C gene. Isolated RP can be caused by a number of gene mutations and can follow an autosomal dominant, autosomal recessive, or X-linked pattern of inheritance. We present a Louisiana Acadian family affected by both autosomal recessive Usher Syndrome and autosomal dominant isolated RP. The proband is clinically diagnosed with isolated RP and is a heterozygous carrier for the common USH1C mutation. The proband's brother is also affected with isolated RP with no common USH1C mutations detected. The proband's two maternal half-brothers are homozygous for the USH1C G-A mutation. An extensive family history spanning five generations revealed eight (8) Usher Syndrome 1C affected individuals on the maternal side of the family, while most members of the paternal side of the family were clinically diagnosed with isolated RP. We performed exome sequencing to determine the mutation(s) responsible for the isolated RP and aid in the diagnosis, prognosis and genetic counseling of family members.

2474T

Parental Mosaicism in Type II Collagenopathies. *N. Beck[1], K. Seifert[2], A. Kumar[2], J. Hoover-Fong[1].* 1) Greenberg Center for Skeletal Dysplasias, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD; 2) Department of Plastic and Reconstructive Surgery, Johns Hopkins Hospital, Baltimore, MD.

COL2A1 (OMIM #120140) encodes the pro-alpha-1 (II) chain of type II collagen. Mutations in *COL2A1* are associated with multiple syndromes such as spondyloepiphyseal dysplasia congenital, Kneist syndrome, Stickler syndrome and others. Stickler syndrome is due to mutations in *COL2A1/COL11A1/COL11A2* (dominant forms) or *COL9A1/COL9A2* (recessive forms). The type II collagenopathies have pleiotropic manifestations in skeletal, ocular, orofacial and auditory systems, necessitating multidisciplinary evaluation and care. Inter- and intrafamilial variability and phenotypic overlap among the type II collagenopathies makes clinical diagnoses challenging. We report a young male with Pierre Robin sequence (PRS) who transferred care to our multidisciplinary Cleft Lip & Palate Clinic at 21 months of age without any prior genetics work up. He had normal global development, malar flattening, mild joint hypermobility and normal hearing, but vision had never been formally assessed. His father had joint hypermobility and extreme deficiency of cartilaginous tissue in his nasal tip and pinnae and no vision or hearing concerns. We referred our patient to ophthalmology where myopia was detected and close follow-up was recommended due to the risk of retinal detachment. With our clinical suspicion of autosomal dominant Stickler syndrome, molecular analysis of the *COL2A1/COL11A1/ COL11A2* genes was pursued and revealed a previously reported pathogenic variant in *COL2A1* (c. 3597+1G>C), a transversion in IVS50 at an invariant position of the donor splice site, resulting in aberrant mRNA processing. Genetic counseling was provided to the patient's parents and the father was found to share the same COL2A1 mutation in 10% of his peripheral blood sample; mother was negative. *COL2A1* mosaicism has been reported in three unrelated families with both Kneist and Stickler phenotypes by Winterpacht et al (1993) and Spranger et al (1994) and in five unrelated families with varying type II collagenopathies by Nagendran et al (2012). The somatic mosaic parent in these publications was paternal in 5/8 (4/8 molecularly confirmed). Our case highlights the importance of a comprehensive genetic medicine work up for patients with features of any of the type II collagenopathies to direct diagnosis-specific care and allow for familial testing. For this case, a confirmed diagnosis has provided our family specific recurrence risks and has allowed for directed medical care for the mosaic parent.

2475F

Mitotic Intragenic Recombination: A Mechanism of Survival for Several Congenital Disorders of Glycosylation. *M. S. Kane[1,2], M. Davids[1,2], C. Adams[1,2], L. A. Wolfe[1,2], H. Cheung[1,2], A. Gropman[1,3], V. Konasani[4], B. Ng[5], C. Scaman[4], H. Freeze[5], D. R. Adams[1,2], W. A. Gahl[1,2], C. F. Boerkoel[4].* 1) NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, MD; 2) National Human Genome Research Institute, Bethesda, MD; 3) Department of Neurology, Children's National Medical Center, Washington, DC; 4) University of British Columbia, Vancouver, Canada; 5) Sanford-Burnham Medical Research Institute, La Jolla, CA.

Congenital disorders of glycosylation (CDG) are disorders of protein glycosylation that affect multiple organ systems. Since most CDGs occur in only a few individuals, we have limited understanding of the associated phenotypes and the mechanisms of patient survival. We investigated two siblings, aged 6 and 11 years, with MOGS-CDG (previously referred to as CDG-IIb) and biallelic *MOGS* mutations (*NM_006302. 2:c. [65C>A; 329G>A];[370C>T]*), who survived much longer than the only other reported MOGS-CDG patient, a child who died at 74 days of age. In these two siblings, we detected multiple *MOGS* genotypes including wild type alleles in the DNA of their peripheral blood and cultured fibroblasts. Analysis of DNA from cultured fibroblasts of six individuals with compound heterozygous mutations of PMM2-, MPI-, ALG3-, ALG12-, DPAGT1-, and ALG1-CDG also revealed multiple genotypes including wild type alleles in each case. Droplet digital PCR showed a ratio of nearly 1:1 wild type to mutant alleles for most, but not all, mutations. This suggests that selection for cells carrying wild type alleles arising from mitotic recombination contributes to the survival and the variable expressivity of individuals with compound heterozygous CDGs. Our findings also provide an explanation for prior observations of a reduced frequency of homozygous mutations and the higher than expected residual enzyme activity in cultured fibroblasts of patients with CDGs.

2476W

Tuberous sclerosis complex-causing mutations analyzed in a cohort of 41 Brazilian patients. *L. Almeida[1], J. P. G Almeida[2], L. Masulk[3], M. Richartz[3], N. Miranda[3], S. A Antoniuk[3], S. Rosemberg[2], L. A Rosemberg[1].* 1) Department of Genetic and Evolutionary Biology, Universitade de Sao Paulo, Sao Paulo, SP, Brazil; 2) Department of Pediatrics, Division of Child Neurology, Faculdade de Medicina da Santa Casa de São Paulo, Sao Paulo, Brazil; 3) Department of Pediatrics, Universidade Federal do Parana, Curitiba, Brazil.

Tuberous sclerosis complex (TSC) is an autosomal dominant hamartomatosis with variable expression, affecting one out of 10,000 newborns. Hamartomas may develop notably in the brain, kidneys, heart, skin, lungs, and retina. Mutations in either tumor suppressor gene *TSC1* or *TSC2* are responsible for TSC. The proteins expressed by those genes, respectively hamartin and tuberin, form a heterodimer that inhibits the mammalian target of rapamycin complex 1 (mTORC1), controlling cell growth and proliferation. Comprehensive screens for *TSC1* and *TSC2* mutations are well established, disclosing approximately 70% of mutations in the *TSC2* gene and 30% in the *TSC1* gene. Even with the most current techniques, nearly 10% of the patients have no mutation identified (NMI). We aimed to search for TSC-causing mutations by standardizing different techniques to evaluate the *TSC1* and *TSC2* sequences. Here we present the first thorough molecular analysis of the *TSC1* and *TSC2* genes conducted in a cohort of patients from Brazil having definite clinical diagnosis of TSC. We examined leukocyte DNA from the first 41 patients referred to our laboratory from two collaborating clinical services. Each sample was evaluated by *TSC1* then *TSC2* Sanger sequencing, followed by multiplex ligation-dependent probe amplification (MLPA) and quantitative PCR analysis for both genes for every negative result. The full protocol has been employed so far for 37 patients. Other four patients have been enrolled but the analysis is not yet concluded. We detected 36 pathogenic mutations (36/37, 97. 3%) of which 11 (29. 7%) and 25 (67. 6%) distributed respectively to the *TSC1* and *TSC2* genes. The NMI patient presented an exonic DNA variant in the *TSC2* gene that may potentially impair pre-mRNA splicing. For the four patients whose molecular analyses have not been concluded, two may have splicing mutations in the *TSC2* gene, and molecular confirmation is underway. One additional patient has a DNA variant within a predicted intronic splicing-regulating element. The NMI rate for this study is anticipated to be between 2. 4% (1/41) and 12. 2% (5/41), lying within the range observed by other studies. Preliminary results indicating two additional splicing mutations suggest that the NMI rate may be not more than 7. 3% (3/41). By the end of the analysis, DNA from all NMI patients will be submitted to next-generation sequencing targeting the *TSC1* and *TSC2* genes.

2477T

Gynecologic and Obstetric Issues in a Large Cohort of Patients with Hermansky-Pudlak Syndrome. *M. Merideth[1], K. O'Brien[2], T. Markello[2], W. Gahl[1, 2].* 1) Medical Genetics Branch, NHGRI, NIH, Bethesda, MD; 2) Office of the Clinical Director, NHGRI, NIH, Bethesda, MD.

Objective: To evaluate the obstetric/gynecologic health in women with Hermansky-Pudlak syndrome (HPS)**Background:** HPS is a rare autosomal recessive disorder characterized by defective biogenesis of lysosome-related organelles. There are 9 known subtypes of HPS caused by mutations in 9 genes involved with intracellular vesicle formation and trafficking. Clinical manifestations include oculocutaneous albinism, a bleeding disorder, and, in some patients, granulomatous colitis and/or a fatal pulmonary fibrosis. The bleeding disorder in HPS is related to an absence of platelet dense bodies, which are lysosome-related organelles. Women with HPS often have problems with heavy menstrual periods and excess bleeding during deliveries, yet very little information has been published on the gynecologic/obstetric issues in HPS. **Methods:** Ninety-eight females with HPS were surveyed under an NIH IRB-approved protocol from 2007 to 2015. **Results:** The patients ranged in age from 12-67y with a median of 35y, and were diagnosed with HPS at a median age of 21y; median age at menarche was 12y (range 8-17). Eighty-two of the 98 patients (84%) report a history of heavy menstrual periods; 56 were treated with oral contraceptive pills to manage heavy menses, and 47 of 56 reported improvement. Six of 8 patients had improvement of heavy menses after insertion of a Mirena progesterone IUD; 4 had improvement with Depo Provera. Seven patients underwent hysterectomy for heavy menses at a median age of 38y. Forty-nine patients have been pregnant with a median of 2 pregnancies (range 1-5). Fifteen of 49 patients had problems with bleeding during pregnancy. Bleeding problems were reported in 15 deliveries, requiring platelets, blood transfusion or ddAVP in 6. Prophylactic platelet transfusion or ddAVP was given prior to 19 deliveries, with no subsequent bleeding problems reported. Postpartum bleeding problems occurred in 28 deliveries: treatment by surgery, transfusion of platelets or medication was required in 14. **Conclusions:** Women with HPS can have significant problems with vaginal bleeding leading to multiple interventions. Obstetrician/Gynecologists have an opportunity to assist in the diagnosis of HPS patients, particularly since a majority of patients surveyed were diagnosed many years after the onset of menses. The history of a menstrual bleeding disorder, combined with some degree of hypopigmentation, should prompt investigation into the diagnosis of HPS.

2478F

Novel *RAF1* mutation in a newborn with Noonan syndrome phenotype. *R. Zambrano[1], J. Patrick[2], J. Surcouf[2].* 1) Pediatrics, Clinical Genetics, Louisiana State University Health Science Center, New Orleans, LA; 2) Pediatrics, Neonatology, Louisiana State University Health Science Center, New Orleans, LA.

Noonan Syndrome (NS) is a dominant disorder characterized by distinctive facial features, short neck, short stature, congenital heart defects, pectus deformities and variable developmental delays. NS is genetically heterogeneous as mutations in several genes involved in the Ras/MAPK (mitogen-activated protein kinase) pathway have been associated with a NS phenotype. 50 % of patients harbor mutations in *PTPN11*, 3-17 % of patients have mutations in *RAF1.* We present a 32 3/7 week gestational age male infant born vaginally to a 24 year old G2P1 mother. Prenatal ultrasounds revealed polyhydramnios, hydrops fetalis with pericardial effusion and cardiac arrhythmias. Dysmorphic features at delivery included: Short neck; dysplastic, low set and posteriorly rotated ears; down-slanting palpebral fissures; short nose with anteverted nares and minor micrognathia. His prolonged and complicated NICU course included severe obstructive hypertrophic cardiomyopathy (HCM) , dysplastic mitral valve, atrial septal defect, severe pulmonary hypertension and arrhythmias. Cardiac failure developed, and despite aggressive medical management, it resulted in death at DOL#73. NS was suspected and molecular testing revealed a novel missense mutation of a highly conserved residue in *RAF1* (p. Pro261Arg: c782 C>G). *RAF1* encodes a downstream effector of RAS signaling in the MAPK pathway. It has three domains: CR1 (residues 61–192) contains a Ras-binding domain, CR2 (residues 251–266) is a site of regulatory phosphorylation and association with the 14-3-3 protein and CR3 (residues 333–625) also interacts with the 14-3-3 protein. In the literature, *RAF1* mutations have been associated with HCM in NS; 4 published patients were found to have a single nucleotide change at nucleotide 781 resulting in Pro261Ser (3) and Pro261Ala (1) substitutions , they all had non lethal HCM and septal defects. Our novel mutation was reported in ClinVar with conflicting interpretations of pathogenicity. This case clearly illustrates this mutation is responsible for our patient's NS phenotype, further supporting its classification as a pathogenic variant. .

2479W

Unusual Frequency of MPS IIIC on central Colombia: probable genetic cluster? *H. M. Velasco[1], Y. Sanchez[2], A. Martin[1], J. Galvis[1], L. A. Umaña[1].* 1) Unidad de Genética Humana, Universidad Nacional de Colombia, Bogotá Colombia; 2) Universidad Pedagógica y Tecnológica de Colombia, Túnja, Colombia.

There are several reports of an increased burden of genetic disease in several regions of Colombia; likely due to multifounder effect as wel as for endogamy and consanguinity. Using a community health approach, a group of geneticists and a pediatric neurologist explored the Boyaca region in central Colombia. We identified in a village of 2500 inhabitants a cluster of 11 patients with mucopolysaccharidosis type IIIC; Pedigree revealed that most members of the hamlet where all of the patients were found had an origin in 4 ancestral families and there were several instances of consanguinity; molecular analysis of *HGSNAT*identified a novel c. 1360C>T mutation which leads to an early termination codon. The estimated incidence for this region is 1:228 which is significantly higher than reported in the literature and might account for the first population cluster for Mucopolysaccharidosis type IIIC.

2480T

Medical complications and functional issues in 122 pediatric patients with non-vascular Ehlers-Danlos syndrome. *A. D. Kline, M. Ferguson, A. Kimball, C. Haakonsen, C. Francomano.* Harvey Inst for Human Genetics, Greater Baltimore Med Ctr, Baltimore, MD.

The Ehlers-Danlos syndromes (EDS) form a group of connective tissue disorders, most of which are autosomal dominant and due to mutations in genes coding for collagen (e. g. types III and V). Although numerous reports exist describing some of the clinical course, there has not been a comprehensive review of patients in the pediatric population to date. We describe an active cohort of 122 patients with non-vascular EDS in terms of specific type, inheritance within family, and involvement of the skin, joint, GI, cardiac, neurologic and neurodevelopmental systems. The majority of the patients (73%) have hypermobile EDS and the rest have classical EDS (15%, clinically diagnosed) and unclear or overlapping type (8%). Nearly 80% are familial cases, mostly maternal, although some additional family members could be very mildly affected and unaware of the diagnosis. Medical complications involve the musculoskeletal, skin, GI, cardiovascular and neurologic systems. Hypotonia was noted in 23%. Previously unreported or single case reports in the pediatric population with EDS have described some rare accompanying findings. In our cohort, these include migraines and headaches in 19%, postural orthostatic tachycardia syndrome in 6%, functional GI complications in 18% (e. g. motility problems, gastroesophageal reflux, irritable bowel syndrome) and Chiari malformation and/or craniocervical junction abnormality in 3%. In addition, we noted accompanying intellectual disability or autism in 15%, learning disability in 11%, ADHD in 14%, and significant behavioral issues, including anxiety or depression, in 12%. Medical work-up and management issues will be discussed.

2481F

Novel compound heterozygous STRA6 mutations in a case of isolated anophthalmia. *H. Chapman[1], A. Schneider[2], T. Glaser[1].* 1) Department of Cell Biology and Human Anatomy, Congenital Eye Study Group, University of California, Davis School of Medicine, Davis, CA; 2) Division of Genetics, Einstein Medical Center, Philadelphia, PA.

Microphthalmia, anophthalmia, and coloboma (MAC) are a spectrum of developmental eye disorders, which often occur together with multisystem defects. Mutations in STRA6 (stimulated by retinoic acid), the polytopic transmembrane (TM) receptor for retinol (vitamin A) binding protein, cause Matthew-Woods syndrome, characterized by anophthalmia, lung hypoplasia, diaphragmatic hernia, congenital heart defects and mental retardation. Few STRA6 variants have been identified in cases with isolated ocular malformations. Here we report novel STRA6 mutations in a compound heterozygous child with bilateral anophthalmia and no other clinical abnormalities. One mutation disrupts a splice acceptor, which is likely to trigger nonsense-mediated decay of the STRA6 transcript, resulting in a null allele. The second mutation changes a conserved aromatic amino acid in the first intracellular loop to an aliphatic residue. Aromatic side groups, especially near the ends of TM segments, are known to mediate protein-protein interactions, ligand binding, and lateral contacts between TM domains. In contrast to previous clinical mutations, which impair surface expression and affect STRA6 globally, this unique hypomorphic allele may be particularly informative, by disrupting a subset of STRA6 functions. We are currently applying a panel of molecular assays to reveal the specific role of this intracellular domain in vitamin A transport.

2482W

***HS6ST2* is mutated in a family with Brooks-Wisniewski-Brown syndrome.** *M. Miozzo[1], L. Paganini[1], L. Pezzani[1], D. Rovina[2], E. Bonaparte[1], L. Fontana[1], C. Pesenti[1], N. Carlessi[1], A. Pirelli[1], R. Falcone[1], M. Chetta[3], S. Tabano[1], S. Sirchia[2], D. Milani[4].* 1) Division of Pathology, Fondazione IRCCS Ospedale Maggiore Policlinico; Université degli Studi di Milano, Milano, Italy; 2) Medical Genetics, Dept of Health Sciences, Université degli Studi di Milano, Milano, Italy; 3) GenomiX 4 Life, Université degli Studi di Salerno, Salerno, Italy; 4) Pediatric Highly Intensive Care Unit, Fondazione IRCCS Ospedale Maggiore Policlinico; Université degli Studi di Milano, Milano, Italy.

Brooks-Wisniewski-Brown syndrome (BWBS, OMIM: 300612) is a very rare genetic disease characterized by severe developmental delay, encephalopathy with atrophic hydrocephalus, microcephaly, progressive spastic diplegia, characteristic facial appearance, optic atrophy, growth retardation and high lactate levels. Based on the male preponderance of BWBS subjects in the two unrelated families studied so far, a recessive X-linked inheritance has been proposed. Furthermore, BWBS overlapping phenotype was described in a XLMR syndrome with severe optical atrophy and decreased vision, whose linkage analysis revealed the involvement of the Xq26 region. A family with clinical diagnosis of BWBS was counseled and studied by NGS-approach in our Institution. It was composed of non-consanguineous healthy parents and three male children: a healthy one and two with BWBS manifestations. We performed the chromosome X exome sequencing by Illumina HiSeq 1500 in the mother and in the three children. Variants annotation was carried out with “Enlis Genome Research” software that identified the single nucleotide unknown variant c. 916G>C within the exon 3 of the Heparan Sulfate 6-O-Sulfotransferase gene (*HS6ST2*). Base change was confirmed by Sanger sequencing and we did not find this variant in more than one-hundred subjects. *HS6ST2* maps in Xq26. In mice model HS6ST2 has a role in retinal neurogenesis, optic fissure closure and optical vesicle development. The 6-O-sulfation enzyme encoded by *HS6ST2* catalyzes the sulfate transfer from 3’-phosphoadenosine 5’-phosphosulfate (PAPS) to position 6 of the N-sulfoglucosamine residue (GlcNS) of the heparan sulfate (HS). HSs are ubiquitous components of the cell surface, extracellular matrix, basement membranes and interact with various ligands to influence cell growth, differentiation, adhesion and migration. The adult *HS6ST2* mutant animals develop optic nerve aplasia/hypoplasia and displayed retinal degeneration. From a preliminary bioinformatics analysis performed by PolyPhen-2, c. 916G>C variant resulted to be deleterious for the HS6ST2 protein structure and function, supporting its role in the onset of optic atrophy associated to BWBS. Our hypothesis requires to be confirmed by *in-vitro* and *in-vivo* functional studies.

2483T

Inherited loss of function mutation in *CASK* causing microcephaly-pontocerebellar hypoplasia: first report. *L. Pisani[1], H. Raynes[2,3], J. Forman[2], S. Desiraju[2], B. Webb[1].* 1) Dept. of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Dept. of Pediatrics, Icahn School of Medicine at Mount Sinai, New York, NY; 3) Dept. of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY.

Mental retardation and microcephaly with pontocerebellar hypoplasia (MICPCH, OMIM #300749) is an X-linked disorder caused by heterozygous loss of function mutations or deletions in*CASK*(calcium/calmodulin-dependent serine protein kinase), located on chromosome Xp11. 4. The phenotype includessevere intellectual disability, microcephaly, variable degrees of pontocerebellar hypoplasia (PCH), very poor psychomotor development, absent ambulation or speech, axial hypotonia with or without hypertonia, and seizure disorder. The phenotype is significantly more severe in males, with perinatal lethality, early infantile epileptic encephalopathy (EIEE), opisthotonus, severe failure to thrive, and/or optic nerve hypoplasia. Dysmorphic features include broad nasal bridge and tip, large ears, long philtrum, micrognathia, and hypertelorism. All the cases thus far reported have been caused by *de novo* mutations. We report a 4 month old boy, born to nonconsanguineous Hispanic parents, presenting with microcephaly, infantile spasms unresponsive to topiramate, bilateral optic nerve hypoplasia, “back arching”, and severe failure to thrive. Upon evaluation, opisthotonus, absent tracking, poor head control, and absent social smile were noted. Facial features were notable for oval face, bilateral epicanthal folds, hypertelorism, low-set ears, high-arched palate, and micrognathia. Family history was reportedly unremarkable, but short stature, delayed speech response, slow processing, and possible staring episodes were noted in the mother. Brain MRI revealed pontocerebellar hypoplasia. Testing revealed 46, XY karyotype, and chromosomal microarray was unremarkable for copy number variation. Mutation analysis of*CASK*done as part of a NGS panel revealed a hemizygous p. M1? pathogenic variant in the translation initiation codon, previously reported in one male patient and shown *in vitro* to cause absence of protein expression. Targeted testing of the mother revealed she was heterozygous for the p. M1? pathogenic variant. . Formal neurologic evaluation and brain MRI were recommended for the mother, but have been deferred. Our patient is now 15 months old; he has received ACTH therapy and has received trials of various anti-epileptic drugs with no change in seizure frequency (6-7/day). Recent renal US revealed diffuse bilateral nephrocalcinosis, likely secondary to topiramate therapy. This is the first report of an inherited loss of function mutation in *CASK* causing the MICPCH phenotype.

2484F

Mild and Moderate Osteogenesis Imperfecta: A Retrospective Chart Review. *E. Carter, L. Tabanfar, J. Davis, C. Raggio.* Ctr Skeletal Dysplasias, Hosp Special Surgery, New York City, NY.

Background:OI is a heterogeneous genetic disorder characterized by bone and tissue fragility and skeletal deformities. Dominant mutations of the type I procollagen genes cause most cases; less common forms are due to mutation of collagen-interactive proteins. In 1979, Sillence et al classified OI into four types based on phenotype: mild type 1, moderate type 4, severe/progressive type 3, and perinatally lethal type 2. When a severe OI phenotype is present, a qualified physician can diagnose the disease. Milder phenotypes are more difficult to diagnose. We reviewed charts from individuals with mild and moderate OI for major clinical findings and genotype information. Hypothesis:Definitive determination of a diagnosis of mild/moderate OI is by direct genomic sequencing. Methods:Retrospective review of charts from patients diagnosed with mild/moderate forms of OI. Clinical findings (fracture frequency, blue/grey sclera, dentinogenesis imperfecta (DI), hearing loss, short stature, scoliosis, osteopenia/porosis, bone deformity, and ligament laxity) and molecular genetic test results were collected. Results:84 individuals (44 female, 40 male; age range 1 months to 84 years) were identified as having classic OI types 1, 4, or a non-classical mild/moderate type. 35 patientshad genotype information available (41. 67%). 30 haddominant mutation in either *Col1a1/1a2* (85. 7%). 8 had non-classical types (*IFITM5, Wnt1, PPIB*). Fifty-one had type 1 OI; 19 (37. 2%) had documented *Col1a1/1a2* haploinsufficiency mutations. Type 1 OI clinical findings: 71% had blue sclerae, 24% had DI, 27% had hearing loss, 57% had osteopenia/porosis, and 14% had bone deformity. Patients reported 0 to 300 fractures. Twenty-five had type 4 OI; 11 (44%) had documented *Col1a1/1a2* mutation. Type 4 OI clinical findings: 64% with DI, 28% with hearing loss, 35% with bone deformity, 48% with osteopenia/porosis; 36% had blue/gray sclerae, 40% had white sclerae, and 24% reported having blue/gray sclerae during childhood. Patients reported 0 to 200 fractures. Conclusion:The majority of patients clinically diagnosed with mild or moderate OI did not have molecular genetic confirmation. Genotype did not correlate with fracture number. Clinical findings in our cohort were consistent with those reported in other patients with types 1 and 4 OI. Our sample size was too small to draw statistically significant conclusions. We will continue to review phenotype/genotype information from patients with mild/moderate OI.

2485W

Clinical and Molecular Assessment of 78 OculoAuriculoVertebral Spectrum (OAVS [MIM 164210]) patients. *S. Bragagnolo[1], M. E. Colovati[2], M. F. S. Soares[3], C. A. Kim[4], M. I. Melaragno[2], A. B. A. Perez[1],* *Financial support: FAPESP 2013/04623-2.* 1) Genetic Medical Center, UNIFESP, Sao Paulo, Sao Paulo, Brazil; 2) Mophology and Geneic Department, UNIFESP, Sao Paulo, Brazil; 3) Imaging Department ,UNIFESP, Sao Paulo,Brazil; 4) Children Institute, USP, Sao Paulo,Brazil.

Purpose: Clinical characterization and classification of OAVS patients following the minimum criteria from Tasse et al (2007): facial asymmetry with asymmetric and/or preauricular pits or tags. Methods: We assessed 78 patients, 55 male and 24 female, with the protocol based in the Tasse et al. (2007) and Golgoutas e cols. (2005) criteria. Results: 23(28%) cases are familial and the remaining is sporadic (72%). 61 patients performed a craniofacial X ray and 52 (85%) presented with abnormalities. 64 patients performed spinal X ray and 47 (73%) presented with abnormalities. 55 patients performed temporal bones CT scan and 41 (74%) presented with abnormalities. 62 patients performed audiology evaluation and 50 (81%) presented some hearing loss. 45 patients performed a nasofibroscopy and 28 (62%) presented with abnormalities. 54 patients performed ophthalmologic evaluation and 32 (59%) presented with some kind of abnormality: 12 patients (22%) were diagnosed with a epibulbar dermoid. 64 performed an Echocardiogram and 42 (65%) presented with abnormalities. 62 patients performed a renal ultrasound and 26 (41%) presented with abnormalities. 27 (35%) patients were detected with some abnormalities in other systems and organs. 67 patients performed KT analysis and 9 (13%) presented with chromosomal abnormalities. 67 patients performed CMA and 29 (43%) patients presented altered results 16 (19%) variants considered pathogenic 13 (19%) variants of unknown significance. In between the pathogenic 8 (50%) were duplication, 7 (44%) were deletions (44%) and one del/dup (6%). 7 patients (44%) presented with a chromosome 4 alteration, 5 (32%) in the short arm (2 deletions and 3 duplications) and 2 (12,5%) in the long arm (1 duplication and 1 deletion). Discussion: The clinical assessment allowed us to classify these patients within the OAV spectrum and the phenotypic association was compatible with the literature. CMA provided a much more accurate laboratorial diagnose than the KT analysis and we could observe recurrent chromosomal abnormalities especially on regions 4p15. 33, 4p16. 1 (two patients), 4p16. 3p15. 33, 4p16. 3, 8q13. 3 e 22q11. 21. These regions contain genes related to phenotypes associated to OAVS, like: *NKX3-2* (*BAPX1* [MIM 164210]); *HMX1* [MIM 612109]; *EYA1* [MIM 166780]; and *YPEL1* e *MAPK1* (*ERK2)*, in 22q11. 2.

2486T

Performance of Facial Dysmorphology Novel Analysis: comparison to molecular testing-based diagnostics. *L. Basel-Vanagaite[1,2,3], S. Lyonnet[4], V. Cormier-Daire[4], M. Rio[4], J. Amiel[4], L. Wolf[2,3].* 1) Rabin Medical Center, Petah Tikva, Israel; 2) Tel Aviv University, Tel Aviv, Israel; 3) FDNA Ltd. , Herzlyia, Israel; 4) Université Paris Descartes-Sorbonne, Imagine Institute, Hôpital Necker Enfants Malades, Paris, France.

Introduction: Facial Dysmorphology Novel Analysistechnology automatically identifies facial patterns associated with genetic syndromes by analyzing two dimensional facial photos. We evaluated the technology's ability to retrieve a gestalt-based list of syndromes that matches molecularly confirmed diagnoses of individuals. Methods: 130 anatomical points were automatically located on the face and multiple lengths, angles and ratios were computed for each face. Local image information was integrated in order to provide the gestalt description of the face. We analyzed 331 frontal images of Caucasian children with 43 different molecularly confirmed genetic syndromes including chromosomal and monogenic syndromes. The images were submitted by clinical geneticists. Each image received a list of syndromes ranked by gestalt-based similarity. Results: The molecularly confirmed diagnosis was listed in the first 20 matches for 89% of the cases, 10 for 75% of the cases, 5 for 69% of the cases, 2 for 54% of the cases and as the first result for 40% of the cases. Statistically, the chances of obtaining such results are close to zero ($p<1e-100$). However, since some syndromes are more common than others, we ran permutation tests. Mixing the images and the diagnosis randomly, the rank 1 retrieval was less than 5%. For microdeletion syndromes, the best performance was observed for Angelman syndrome (listed in the first 5 matches in 93% of the cases) as compared to Williams syndrome and velo-cardio-facial syndrome (86% and 67%, respectively). For monogenic syndromes, SHORT Syndrome, Kaufman oculocerebrofacial syndrome and auriculo-condylar syndrome were ranked as correct 1st match in all cases. Conclusions: We conclude that a system supported with Facial Dysmorphology Novel Analysistechnology can be useful for medical professionals in diagnosing genetic syndromes. Future applications include: 1) guiding targeted molecular testing; 2) complementing next generation sequencing-based molecular testing by inferring causative genetic variants from sequencing data; 3) assisting in diagnostics in a country where molecular testing resources are poor and clinical diagnosis is therefore important.

2487F

Molecular Pathogenesis of Congenital Diaphragmatic Hernia Revealed by Family Based Genetic Strategies, Next Generation Sequencing and Bioinformatics. *H. Shanmugam[1], N. Matsunami[2], L. Baird[2], J. Stevens[2], B. Otterud[2], C. Rau[1], S. Baker[1], N. Bowels[2], M. Leppert[2], L. Brunelli[1].* 1) Department of Pediatrics-Neonatology, University of Utah Health Sciences Center, Salt lake city, UT; 2) Department of Human Genetics, University of Utah Health Sciences Center, Salt Lake City, UT.

Background: Congenital diaphragmatic hernia (CDH) is a severe birth defect with high morbidity and mortality. Despite the clinical significance of CDH, the molecular causes underlying the defect are not completely understood. CDH is considered to be a multifactorial disease, with strong evidence implicating genetic factors. Although aneuploidies and copy number variants are important in CDH, few single genes have been definitively implicated in human CDH. **Aim**: We aim to identify the pathogenic variants leading to syndromic and isolated CDH using family-based genetic strategies and next generation sequencing. **Methods**: After IRB approval, we recruited a syndromic trio with unaffected parents and a daughter with duplication of pituitary gland syndrome and CDH. Additionally, we recruited a highly penetrant three-generation isolated CDH family with an affected grandfather, an affected father, and two affected daughters. Whole exome sequencing was carried out on each blood DNA. DNA sequence was analyzed by BWA/GATK and Novoalign/GATK. Variant filtering, classification and prediction were performed using Golden Helix SNP & Variation Suite. Genetic analysis was carried out using recessive, compound heterozygous, and *de novo* models in the syndromic trio. Dominant model and variant sharing analysis among affected individuals were used for the three-generation isolated CDH family. Finally, we used pVAAST (Hu et. al. , *Nat Biotechnol* 2014) to prioritize variants. Candidate variants were validated by Sanger DNA sequencing. **Results:** In the syndromic CDH trio, we identified two *de novo*missense mutations in *H1ST1HE* and *FILIP1* genes. In the three-generation isolated CDH family, we identified about 50 rare functional variants inherited from the affected grandfather. The best candidates scored by pVAAST analysis include two missense variants in *PPARD* and *CWF19L1* genes, a nonsense variant in *STOX1* gene, and a frame-shift variant in *MYO1C* gene. We are currently testing some of these variants in mice. **Conclusion**: The combination of family based genetic strategy with next generation sequencing and bioinformatics reveal two *de novo* mutations in a syndromic CDH trio and the role of heritability in a three-generation isolated CDH family. By defining the spectrum of genes associated with CDH, we will be able to define common molecular causes of CDH and define the clinical syndromes associated with these genes to inform prognosis and hopefully guide treatment.

2488W

Fragile X syndrome and periventricular heterotopia: one or two syndromes? *A. Lavillaureix[1,8], M. Gerard[2], M. Baulac[3], S. Drunat[1,4,5], C. Beldjord[6,7], A. Verloes[1,4,5], S. Passemard[1,4,5].* 1) Département de Génétique, Hôpital Robert Debré, Paris, France; 2) Département de Génétique, Hôpital de Caen, Caen, France; 3) Department de Neurologie, Hôpital Pitié-Salpêtrière, Paris, France; 4) INSERM U676, Hôpital Robert Debré, Paris, France; 5) Université Paris Diderot, Paris 7, Paris, France; 6) Département de Biologie Moléculaire, Hôpital Cochin, Paris, France; 7) Département de Génétique et Développement, Inserm U1016, Institut Cochin, Paris, France; 8))Université Paris Descartes, Paris 5, Paris, France.

Children carrying Fragile X full mutation exhibit mild to severe intellectual disability, behavioral disorders and a higher risk of seizures. Epilepsy occurs in 12% of Fragile X syndrome patients, beginning between 4 and 10, with a higher prevalence in males than in females. Although evidences suggest that migration disorders may be observed in Fragile X syndrome, the origin of seizures is not systematically sought and brain MRI is rarely performed in these patients. Here we describe periventricular heterotopia (PVH) in two siblings, a male and a female, with Fragile X full mutation. The boy, with intellectual disability, because of abnormal movements during sleep, had investigations that brought out unilateral PVH at age 10. His sister presented late onset, drug resistant epilepsy beginning at age 18, and extensive bilateral laminar and nodular periventricular heterotopia. PVH has been only described in four males with Fragile X syndrome and has never been reported in females. We discuss here the interest of investigating these patients with brain MRI for the developmental and epilepsy follow-up and confirm that PVH may be caused by Fragile X mutation in males and should suggest other etiologies in females, such as FLNA mutation, even in a context of Fragile X full mutation.

2489T

-238 Tumor necrosis factor-alpha gene polymorphism in Periodontitis patients. *R. P. Mariaud-Schmidt[1], M. G. Fuentes-Lerma[2], J. R. Gomez-Sandoval[1], V. C. Martinez-Rodriguez[3], I. A. Gutierrez-Hurtado[4], M. P. Gallegos-Arreola[4].* 1) Instituto de Investigación en Odontología, Universidad de Guadalajara, Guadalajara , Mexico; 2) Departamento de Clinicas, Centro Universitario de los Altos, Universidad de Guadalajara, Mexico; 3) Posgrado en Periodoncia, DECOI, Universidad de Guadalajara, Mexico; 4) Laboratorio de Genetica Molecular, CIBO-IMSS, Guadalajara, Mexico.

Abstract*Background.* Periodontal disease is triggered by bacteria, predominantly Gram negative anaerobes, which produce inflammatory mediators leading to the destruction of connective gingival tissue and dental alveolar bone. Cytokines as tumor necrosis factor alpha (TNF-α) induces periodontal disease. TNF-α is a key mediator of the inflammatory process and the same cytokine gene polymorphism has been associated with periodontal disease, although such results remain controversial. The aim of the present study was to evaluate the frequency of –238 TNF-α gene polymorphism in periodontal disease patients. *Material and Methods*. The protocol was approved by the local clinical ethics committee and a written explanation of the purpose of the study was provided for each subject and signed consent according to the Helsinki Declaration was obtained for the genetic analysis. 77 periodontitis patients and 49 control subjects were enrolled in the study. Blood sample were taken from participants and DNA extraction was made by the Miller and Gustincich method and stored frozen at -20°C until use. TNF-alpha SNPs was genotyped by polymerase chain reaction-restriction fragment length polymorphism for all subjects. *Results*. All 77 patients enrolled in the study had clinical and biological features of chronic periodontitis (periodontal disease). 69 (90%) patients had GG genotype, 7 (9%) patients had GA genotype and only one patient (1%) had AA genotype. The control subjects 42 (86%) had GG genotype, 6(12%) had GC genotype and 1 (2%) had AA genotype. Statistical analysis. We used *X2* test for frequencies determination. Non significative difference was observed. *Conclusion.* Genetic polymorphism in the TNF-alpha gene at position-238 could not be identified as susceptibility factor in chronic periodontitis.

2490F

Complex chromosome 6 aberration in two sisters with MR and MCA derived from a maternal paracentric inversion. *F. A. T. de Vries[1], G. M. S. Mancini[1], F. Bal-van Breda[1], W. Y. Cheung[1], G. C. B. Bindels-de Heus[2], C. H. Wouters[1].* 1) Department of Clinical Genetics, ErasmusMC, Rotterdam, Netherlands; 2) Department of Pediatrics, ErasmusMC, Sophia Children's Hospital, Rotterdam, Netherlands.

Two sisters aged 5 and 3 with developmental delay, behaviour problems, excessive drooling, short stature, respiratory distress and stridor caused by narrow trachea, strabismus and very similar facial featureswere investigated with SNP array analysis. Besides two losses in the X-chromosome (p22. 33) and in chromosome 7 (p21. 2p21. 1), both inherited from the healthy mother, two duplications were detected in chromosome 6 (q14. 1 and q16. 1) together with a deletion (q23. 3q24. 1). Parental array analysis showed a normal pattern for the chromosomes 6. FISH performed on parental chromosomes showed a paracentric inversion of one of the chromosomes 6 (q14. 1q24. 1) from the mother. A meiotic recombination of the paracentric inversion could not result in the exact same derivative chromosome 6 found in both daughters. Moreover the resolution of the dicentric chromosome after meiotic recombination would result in a derivative chromosome with a single interstitial duplication and a terminal deletion. The only conclusion has to be that the mother is a mosaic, the inv(6) will be present in the mature part of her cells, and the similar der(6) (resulting from a mitotic recombination), as found in her daughters, will be (at least) present in her germ line. This is an unusual case of high recurrence risk for a pathogenic rearrangement in a healthy woman who is carrier of a paracentric inversion, generally considered a low risk chromosome aberration. In the prenatal counseling, this is an apparent exception to the general rule that paracentric inversions should not be offered invasive prenatal diagnosis.

2491F

Formation of triplications by a template switching mechanism can contribute to disease phenotype in humans beyond formation of simple copy-number gain. *C. M. B. C. Fonseca[1], N. A. Hanchard[1], B. Yuan[1], P. Stankiewicz[1], J. W. Belmont[1,2], S. W. Cheung[1], V. R. Sutton[1,3], P. I. Bader[4], J. R. Lupski[1,2,3].* 1) Dept Molecular Human Genetics, Baylor Col Medicine, Houston, TX; 2) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 3) Texas Children's Hospital, Houston, TX; 4) Northeast Indiana Genetics, Fort Wayne, IN.

Segmental triplications (copy number N = 4 in autosomes and N = 3 when on male X-chromosome) can contribute to disease etiology via gene dosage effects, gene disruption, position effects or fusion gene formation. We have recently published five cases for which the triplications were formed in association with a distal absence of heterozygosity (AOH) extending to the telomere on the same chromosome (PMID: 25799105). These regions of AOH ranged in size from 6 Mb to 50. 3 Mb in length. The genomic structure of those post-zygotic complex genomic rearrangements (CGRs) shows a consistent pattern of an inverted triplication flanked by duplications (DUP-TRP/INV-DUP) and is hypothesized to be formed by replication-based mechanisms such as microhomology-mediated break induced replication (MMBIR) and fork stalling and template switching (FoSTeS). We now report a patient who carries a 2. 1 Mb DUP-TRP/INV-DUP followed by 42. 2 Mb of AOH involving chromosome 14 identified by array comparative genomic hybridization and breakpoint junction sequencing. Using SNP array OmniExpress (Illumina) we determined that the segmental AOH occurred on the maternally inherited chromosome and therefore this patient has a high risk to present with clinical features of maternal uniparental isodisomy for chromosome 14 (UPD(14)mat) [MIM#616222] as well as for expression of recessive traits. Infinium Human Methylation450 Beadchip (Illumina) in samples extracted from this patient, unaffected family members and previously known UPD14 patients (both maternal and paternal) revealed that the genomic DNA of this patient shows methylation patterns typical of UPD(14)mat at both known (*MEG3* promoter) and novel (intron 2 of *MEG8*) differentially methylated regions along the *DLK1-DIO3* domain. This data provide experimental evidence that, in humans, triplication generated post-zygotically can lead to segmental UPD. Furthermore, the impact on the disease phenotype resulting from the underlying replication-based mutational mechanism in humans can extend beyond that due to the formation of copy-number variants.

2492F

A case of partial trisomy 21q resulting from meiotic recombination of a maternal pericentric inversion. *S. Jung[1], K. Lee[2], M. Jang[1], C. Chi[1], K. No[1], K. Lee[1].* 1) Molecular Diagnostics Testing Center, Seegene Medical Foundation, Seoul, South Korea; 2) Department of Physics, Kangwon National University, Chuncheon, South Korea.

Down syndrome (DS) is a genetic disorder caused by an extra copy of chromosome 21. In rare cases, DS is associated with a partial trisomy 21 and less than 3% of DS patients have duplication as a result of parental pericentric inversion of chromosome 21. Here we report the case of a two-month-old girl with features of DS but with an atypical karyotype. Initial chromosome analysis of peripheral blood lymphocytes was 46,XX-,21p+. To identify the extra chromosomal material on 46,XX,21p+, parental chromosome study was performed. The results showed a normal paternal karyotype (46, XY) and an abnormal maternal karyotype [46,XX-,inv(21)(p12q22. 1)]. The fluorescence in situ hybridization (FISH) study of the patient was performed. The interphase FISH analysis with Vysis LSI ETV6(TEL)/RUNX1(AML1) ES Dual Color Translocation Probe showed three signals of the probe for 21q22. The metaphase FISH showed two hybridization signals of 21q22 within one chromosome 21 homologue. According to FISH and parental karyotype results, her karyotype was designated as: 46,XX,rec(21)dup(21q)inv(21)(p12q22. 1)mat. ish rec(21)dup(21q)inv(21)(p12q22. 1)(RUNX1++). nuc ish(RUNX1x3). To our knowledge, this is the 1st report in Korea of a DS patient with rec(21)dup(21p).

2493F

Clinical, cytogenetic and molecular evaluation of a phenotipically male patient with a karyotype of 46, XX but bearing an intact SRY loci: a case report. *H. Bagis[1], M. O. CEVIK[1], A. CIFT[2], I. GUNEY[1].* 1) Department of Medical Genetics, Adiyaman University School of Medicine , Adiyaman, Turkey; 2) Adiyaman University School of Medicine Department of Urology, Adiyaman, Turkey.

OBJECTIVE:The objective of this study was to analyze of clinical, cytogenetic and molecular features of a 46,XX phenotypically male SRY positive patient in terms of infertility. METHODS:The clinical features of a phenotypically male patient with 46, XX (SRY positive) were reported. Chromosome analysis showed a 46, XX karyotype and translocation of SRY (testis-determining factor) was detected by fluorescence in situ hybridization (FISH) and molecular analysis of AZF microdeletions on the Y chromosome were performed by multiplex PCRs amplification. The set of PCR primers for the diagnosis of microdeletion of the AZFa, AZFb and AZFc region included: sY81a, sY84, sY86, sY87, sY88, s128a, sY121b, sY124c, sY127, sY128, sY130, sY133, sY134, sY143b, sY145, sY152, sY157, sY158, sY160, sY254, sY255c, SRY and ZFX/ZFY. RESULTS:A 24 year old married man was referred to urology clinic with a 4 year history of infertility due to male partner. No gynaecomastia was noted and his facial, axillary and pubic hair was all of normal density and distribution. Physical examination revealed that his testes were small in volume and soft in texture and has a bilateral varicocele, but his penis phenotype was normal. Semen analyses showed complete azoospermia. All were karyotyped as 46, XX. Molecular analyses display the presence of the SRY gene and absence of AZFa, b and c of the Y chromosome. FISH analysis showed that SRY genes were translocated to X. CONCLUSION:Classical, molecular cytogenetic and molecular genetics tests on 46,XX male syndrome may help to explain its genotype-phenotype relationships. Our laboratory test results suggest that genetic counseling useful information in these infertile men before starting assisted reproductive treatments.

2494F

High incidence of CNVs in patients with oculo-auriculo-vertebral spectrum disorders. *S. Oliveira[1 2], P. Costa dos Santos[2], C. Pereira de Oliveira[3], A. Bastos de Oliveira[2], H. Safatle[4], M. Santos Cordoba[4], M. T. Silva Rosa[4], C. Rosenberg[5], E. Freitas[5], E. Ribeiro[6], R. Pogue[7], R. Pereira[7], I. Ferrari[1], H. Mestrinho[1], A. Pic-Taylor[1 2], J. Mazzeu[2 3 8].* 1) Genetica e Morfologia, Universidade de Brasilia, Brasilia, Brasilia (Distrito Federal, Brazil; 2) Programa de Pós-graduação em Ciências da Saúde, Universidade de Brasília, DF, Brazil; 3) Programa de Pós-graduação em Ciências da Médicas, Universidade de Brasília, DF, Brazil; 4) Ambulatório de Genética, Hospital Universitério de Brasília, Universidade de Brasília, DF, Brazil; 5) Departamento de Genética e Biologia Evolutiva, Instituto de BioCiências, Universidade de São Paulo, SP, Brazil; 6) Hospital Infantil Albert Sabin, Genética Médica, Fortaleza, Ceará, Brazil; 7) Programa de Pós-graduação em Ciências Genómicas e Biotecnologia, Universidade Católica de Brasília, DF, Brazil; 8) Área de Clínica Médica, Faculdade de Medicina, Universidade de Brasília.

Oculoauriculovertebral spectrum (OMIM 164210) is characterized as a developmental defect involving the first and second branchial arch derivatives. This spectrum is clinically broad, ranging from isolated microtia to multiple visceral malformations and it includes the Goldenhar Syndrome. Main clinical features include unilateral or bilateral ear anomalies, hemifacial microsomia, ocular defects and vertebral malformations. Other features are cleft lip and/or palate; cardiac, renal, cerebral malformations; and mental retardation. Most cases are sporadic, but there are familial cases that exhibit autosomal inheritance. Genetic, environmental and multifactorial etiology, especially events that occur during the seventh week of pregnancy, were described. Therefore, OAVS is a clinic and genetically heterogeneous disorder that demands a better understanding. This work performed chromosomal microarray analysis in 26 patients with OAVS in order to identify new genomic loci that are potentially involved in disease pathogenesis. This study was approved by the institutional ethics committee and families provided written informed consent. As expected, results showed a large heterogeneous genetic etiology. No chromosomal abnormalities other than already described polymorphic CNVs were found for 53. 8% of the cases. Diagnosis of normal chromosomal microarray was determined as no rearrangements or chromosomal rearrangements that had been described as non-pathogenic CNVs in DGV database. New chromosomal rearrangements were found, as follows: del22q11. 2, del12q24. 21, LOH 3p26. 3-14. 3, del2p25. 3 and dup2p25. 3, dup4q21. 1, dup20p12. 3-12. 2, dup1q31. 1 and dup10q23. 31, dup11p15. 5, LOH 12q15-q22, dup12p12. 1 and dup6q21. Analysis of inheritance showed that three of these rearrangements were *de novo*, one maternally inherited and one paternally inherited. In seven patients the origin of the rearrangements were not determined. The chromosome abnormalities identified are non-recurrent. This is the first time that these CNVs are described, except for the deletion of 22q11. 2 and 20p12. 2 duplication. These results confirm the genetic heterogeneity of the syndrome and describe new genomic regions and genes involved in its etiology. Acknowledgement: FAP-DF; CAPES; CNPQ;PPSuS.

2495F

Genome-wide association study of copy number variation in a Thai autism spectrum disorder cohort. *N. Jinawath[1,2], A. Hnoonual[3], W. Thammachote[1], T. Pongrujikorn[1], MS. Shiao[1], P. Phokuanoy[4], T. Tim-Aroon[5], K. Rojnueangnit[6], T. Hansakunachai[6], P. Limprasert[3], N. Ruangdaraganon[7], D. Wattanasirichaigoon[5], B. Suktitipat[8,2].* 1) Research Center, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 2) Integrative Computational Bioscience Center, Mahidol University Salaya Campus, Nakhon Pathom, Thailand; 3) Division of Human Genetics, Department of Pathology, Faculty of Medicine, Prince of Songkla University, Songkla, Thailand; 4) Department of Pathology, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 5) Division of Medical Genetics, Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 6) Department of Pediatrics, Faculty of Medicine, Thammasart University, Pathumthani, Thailand; 7) Division of Developmental-Behavioral Pediatrics, Department of Pediatrics, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand; 8) Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok, Thailand.

Autism spectrum disorder (ASD) is an array of neurodevelopmental disorders characterized by difficulties in social interaction, verbal and non-verbal communication, and stereotyped behaviors and interests. The cause of ASD is still largely unknown, but there is a strong genetic link. Since 2010, the use of cytogenomics microarray (CMA) for detection of copy number variation (CNV) has been recommended as a first-tier genetic test for individuals with ASD in the US. In Thailand, the incidence of ASD is currently on the rise, with an incidence rate of 9. 9:10,000 in children aged 1-5 years. It is estimated that the average health care costs for individuals with ASD are 4-6 times more than those of general population, emphasizing the socioeconomic importance of ASD in Thailand. In this study, we employed Illumina CytoSNP-850K to detect CNVs in the 55 Thai patients who met the DSM-IV criteria for non-syndromic ASD, had normal karyotype, and a negative Fragile X and *MECP2* testing. CNV locations were identified using BlueFuse Multi, an HMM-based algorithm, to detect CNVs with at least 10 probes coverage. After excluding CNVs that overlapped with pericentromeric regions, we identified a total of 371 CNVs in all patients. The median number of CNVs found was 6 per patient, ranging from 2 to 14. The CNV size ranged from 1,632bp to 4. 5Mb, with a median of 90kb. The overall diagnostic yield of CMA in the Thai ASD cohort was 15%, which is comparable to the reported yield in other studies. ASD-associated pathogenic CNVs that cause 15q13. 3 microdeletion syndrome and 16p13. 11 microduplication syndrome were among the pathogenic CNVs identified in these patients. Furthermore, the genes overlapping with CNVs in ASD cohort were then compared with those of CNVs from the 3,017 general Thai subjects previously reported in the Thai CNV database. Using 50,000 permutation-based association tests, implemented in PLINK, CNVs overlapping with *NOTCH2* (1p12), *HBM* (16p13. 3), and *CYP2A7* (19q13. 2), and *PCDH11X* (Xq21. 31) were significantly associated with ASD (genome-wide corrected p-value < 0. 05). Interestingly, both *NOTCH2* and *PCDH11X* are expressed in the brain, and have been previously linked to developmental delay, as well as dyslexia. In summary, this is the first report of genome-wide association study of CNV in Thai ASD subjects. Our findings may contribute to the identification of novel autism-associated genes and elucidation of the underlying genetic basis of ASD.

2496F

***De Novo* Assembly of the Genome-in-a-Bottle Reference Ashkenazi Trio, Structural Variation Discovery and Comparison with Other Individuals by Genome Mapping.** *A. Hastie, E. T. Lam, T. Liang, A. Pang, S. Chan, Z. Džakula, H. Cao.* BioNano Genomics, Inc. , San Diego, CA.

Structural variation analysis (SVA) of human genomes is usually a reference-based process and therefore biased and incomplete. In order to have a comprehensive analysis of structural variation, a *de novo* approach is needed. As a result of the remaining limitations of DNA sequencing and analysis technologies, it is not feasible to create high quality *de novo* assemblies of individuals for detecting and interpreting the many types of structural variation that are refractory to high-throughput or short-read technologies. Using a single-molecule genome analysis system, the Irys® System, we produced high resolution genome maps that were assembled *de novo*. These maps preserve long-range structural information necessary for structural variation detection. The Genome in a Bottle reference trio of Ashkenazi Jewish decent (NA24385, NA24149, NA24143) has been *de novo* assembled by the Irys System. Structural variation analysis reveals insertions, deletions, inversions, including large deletions in the UGT2B17 gene (involved in graft versus host disease, osteopathic health and testosterone and estradiol levels) in the mother and son. We have also investigated the amylase locus in this trio as well as ~ 20 other individuals and have found at least 15 different structural variants. Human amylase genes have variable copy number and this variation is believed to have been evolved to adapt to increase starch intake. We were able to identify multiple copy neutral variants, such as inversions, in these individuals.

2497F

Danish Central Cytogenetic Register: Population-based survival analysis of mental disorders among carriers of chromosomal abnormalities. *L. Høffding[1], BB. Trabjerg[2], A. Vangkilde[1], L. Olsen[1], C. Bøcker Pedersen[2], T. Werge[1], W. Mazin[1].* 1) INSTITUTE OF BIOLOGICAL PSYCHIATRY, Mental Health Services of Copenhagen , Copenhagen, Denmark; 2) Centre for Integrated Register-based Research National Centre for Register-based Research, Aarhus University, Business and Social Sciences, Aarhus V, Denmark.

Aim: To examine the risk of psychiatric disease in individuals recorded in the Danish CentralCytogeneticRegister **(**DCCR) as carriers of chromosomal abnormalities**Background:** Chromosomal abnormalities conferring moderate to high risk of psychiatric disorders have in the later years been identified (1). However, there is still little knowledge on the clinical and epidemiological consequences at the population level of these genomic variants, including the clinical manifestation, disease trajectories, treatment response and social outcome. The Danish Cytogenetic Central Register (DCCR) was established in 1960 and prospectively collects all pre- and postnatal results from individuals clinically referred to genetic testing throughout Denmark. **Methods:** As an example, we identified and extracted all individuals in DCCR carrying the 22q11. 2 microdeletion in Denmark (N=270). Each individual was linked by their unique personal identification number to the Danish nationwide registers and survival analysis using Poisson regression was conducted and adjusted for the following covariates: sex, age, calendar time, place of birth, maternal and paternal age, parental origin and family history of psychiatric disorders. **Results:** The basic model showed an incidence rate ratio (IRR) in carriers of the 22q11. 2 microdeletion for schizophrenia of 8. 13 (95% CI: 3. 65-18. 09). The adjusted IRR of schizophrenia was 6. 55 (95% CI: 2. 94-14. 59). Thus, 22q11. 2 microdeletion carriers have a significantly increased risk of developing schizophrenia compared to the general Danish population qualifying previous association findings (1,2). **Perspectives:** We will report data from similar analysis applied to other chromosomal abnormalities identified in the DCCR including the 22q11. 2 microduplication, Turners and Kleinfelter syndrome. Survival analysis of clinically identified individuals carrying chromosomal abnormalities provide risk estimates for disease useful in genetic counseling and guidance of symptomatic monitoring and early clinical intervention. **References:**1. Stefansson et al. Large recurrent microdeletions associated with schizophrenia. Nature. 2008 Sep 11;455(7210):232–6. 2. Murphy et al. High rates of schizophrenia in adults with velo-cardio-facial syndrome. Arch Gen Psychiatry. 1999 Oct;56(10):940–5. .

2498F

Concurrent loss of heterozygosity and mosaic deletion 12p. *J. F. Mazzeu[1,2,4], R. S. Faria[2], C. P Oliveira[2,3], M. T. A. S. Rosa[2,3], M. S. Cordoba[4], A. P. Taylor[5], S. F. Oliveira[5], I. Ferrari[1,4].* 1) Faculdade de Medicina, Universidade de Brasilia, Brasilia, Brazil; 2) Programa de Pós-graduação em Ciências Médicas, Universidade de Brasília, Brasília, Brazil; 3) Hospital de Apoio de Brasília, Brasília, Brazil; 4) Hospital Universitério, Universidade de Brasília, Brasília, Brazil; 5) Departamento de Genética e Morfologia, Universidade de Brasília, Brasília, Brazil.

Deletions of the short arm of chromosome 12p are the rarest subtelomeric imbalances. Less than 20 patients have been reported to date, of which microdeletions were identified either by FISH or array-CGH without SNP data. Here we describe a patient with a mosaic 12p microdeletion and loss of heterozygosity (LOH) extending to the distal adjacent segment. Clinical examination at age nine revealed intellectual disability, short stature, facial dysmorphims including a small face, small downslanted palpebral fissures, protruded ears with overfolded helices, high nasal root, overbite, high narrow palate, retrognathia, irregular dental implantation and clinodactyly. Chromosome microarray analysis using Cytoscan HD platform (Affymetrix) disclosed a deletion in mosaic on the terminal portion of the short arm of chromosome 12 affecting bands p. 13. 33-p. 13. 32 within genomic positions chr12:1-4, 576,632 (hg19). SNP data revealed loss of heterozygosity at the deleted segment extending 4 Mb in the distal adjacent segment. To the best of our knowledge this is the first report of LOH in constitutional 12p rearrangements. Our patient presents a rearrangement appearing in mosaic, which could the result of a post-zygotic mitotic event, consistent with an error in DNA replication during mitosis, as proposed by the MMBIR model. Since previously reported cases were not investigated by SNP arrays we cannot speculate as to whether this mechanism is recurrent in the formation of 12p microdeletions. This study was approved by the institutional ethics committee and the family provided written informed consent.

2499F

Two sisters with similar phenotype, a shared unbalanced translocation, and discordance for a microdeletion. *A. Nastro[1], A. Shrimpton[2], N. Dosa[1], R. R. Lebel[1].* 1) Center for Development, Behavior, and Genetics, SUNY Upstate Medical University, Syracuse, NY; 2) Department of Pathology, SUNY Upstate Medical University, Syracuse, NY.

Two sisters presented with dysmorphism and global developmental delays, the elder more severely impaired (aggressive and destructive behavior) than the younger. They were examined at 27 and 23 years of age. They had emigrated from Poland as children; chromosome analysis there had documented 8p dup and a balanced 8p;11p translocation in their father. Microarray analyses at our clinic revealed an identical unbalanced translocation involving a loss of 11p15. 5 and a 26. 6 MB duplication of 8p21. 2-p23. 3 (the segment known to be translocated onto 11p). The 8p duplication involves 107 genes, while the 11p15. 5 deletion involves none. The younger sister also has a 1. 33 MB deletion at 10p11. 22 (variant of unknown clinical significance), that is not found in the elder sister not seen in the younger. Both sisters had underdeveloped breasts, abnormal EEG, a wide mouth, multiple nevi on the torso, a vocabulary consisting of only a few words that only the parents understand. The younger sister has a prominent forehead, square face, and decrescendo systolic murmur (all absent in the elder). The 10p11. 2 VUS includes only four genes, and hemizygosity for these is not known to be pathogenic. Thus 8p duplication probably is the sole cause of their adverse phenotypes. The small literature on 8p duplication generally involves inverted duplication. We found no report of 8p dup translocated to 11. Literature about 8p23. 1 duplication syndrome (which is implicated in marked developmental and speech delays), suggests that only four genes appear to clearly be dosage sensitive: *GATA4*, *TNKS1*, *SOX7*, and *XKR6*. Our patients have the first three of these genes duplicated. *TNKS1* is implicated in behavioral problems, *SOX7* in developmental delay, *GATA4* and *SOX7* both may cause heart defects, and the role of *XKR6* is not currently known. The family history is significant for multiple spontaneous miscarriages in the paternal grandparents, and in the father's brother's union, raising the possibility of a translocation segregating in the family. There was also a paternal half-brother who died at one hour of age due to "a spinal abnormality".

2500F

Mosaic supernumerary marker with Y chromosome material detected by NIPT in a newborn female. *V. Potluri[1], R. Burnside[2], G. S. Sekhon[3], H. Xie[3], L. Kline[2], P. Papenhausen[2].* 1) Div Lab Corp America, Dynagene, 3701 Kirby Dr, Suite 528, Houston, TX 77098; 2) Laboratory Corporation of America (R) Holdings, Research Triangle Park, NC 27709; 3) Integrated Genetics, 2000 Vivigen Way, Santa Fe, NM 87505.

Gender discrepant results from NIPT testing have been reported in the literature, the etiology of which is not always clear. Additionally, NIPT results may indicate a structural abnormality for which further invasive prenatal or postnatal testing is required to clarify. A newborn phenotypic female was referred for chromosome analysis and SNP microarray analysis due to NIPT results in which Y chromosome material was detected. Specifically, the NIPT report described the presence of Yp material and the absence of detectable Yq material. Invasive prenatal testing was declined, but placental villi, and peripheral blood from the neonate were sent for further cytogenetic study. Chromosome analysis revealed a mosaic supernumerary marker chromosome in 15 out of a total of 20 cells. Further FISH studies showed that the marker was negative for SRY, explaining the female phenotype. Using DYZ1 and DYZ3, the marker was positive for only DYZ3. Concurrent microarray analysis from the peripheral blood specimen detected a 979 Kb duplication at 1q43 [hg19] (241,108,507-242,087,176), 596 Kb triplication of 5q14. 3 [hg19] (83,523,414-84,119,787), two gains at Yp11. 2 (260 kb and 2. 51 Mb) and two gains within Yq11. 222->q11. 223 (557 kb and 437 kb). *TSPY1*, thought to increase gonadoblastoma risk in females with Y material, was present in the Yp duplication. The autosomal copy number gains do not include genes known to result in abnormal phenotypes, and may be a part of the marker chromosome. Interestingly, maternal chromosome analysis also demonstrated a non-mosaic marker chromosome similar in appearance to that of the proband. Additional FISH studies are in progress to investigate the nature of the maternal marker and whether the autosomal copy number gains are also part of the marker.

2501F

Molecular mechanisms of formation of insertional translocations. *P. Stankiewicz[1,2], M. A. Magrina[3], C. A. Bacino[1,2,4], S. R. Lalani[1,2,4], A. M. Breman[1,2], J. L. Smith[1,2], A. Patel[1,2], S. W. Cheung[1,2], W. Bi[1,2].* 1) Dept of Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Baylor Miraca Genetics Laboratories, Houston, TX; 3) Medical Specialties Unit From City Hall São José dos Campos, SP, Brazil; 4) Texas Children's Hospital, Houston, TX.

Insertional translocations (ITs) are genomic rearrangements with a chromosome segment inserted into a non-homologous chromosome or a different region on the same chromosome. ITs constitute ~ 2% of non-recurrent copy-number gains and are usually revealed through FISH or chromosome analyses. Very little is known about the molecular mechanism(s) of IT formation. Recently, microhomologies at the insertion sites were found in a few IT cases, suggesting a replicative mechanism of formation. We present a family with a mother with mild intellectual disability, carrier of an apparently balanced insertion ins(19)(p13q13. 3q13. 3), and her two affected children who inherited the reciprocal IT products: the 19q13. 3 deletion or duplication. Surprisingly, we found that the duplication in one child (~3. 461 Mb) is slightly larger than the apparently reciprocal deletion in the other sib (~3. 334 Mb). Supporting this finding, chromosomal microarray analysis in the mother revealed two small copy-number gains, 108 kb and 41 kb in size in 19q13. 3, at both ends of the duplication transmitted to one child, but absent in the deletion in the other offspring, indicating DNA replication error. To determine the prevalence of this phenomenon, we are analyzing 51 other ITs from our CMA database. We have identified 11 families with an apparently balanced IT in one parent detected by FISH analysis, 31 families with a gain associated with a simple IT event in the affected patients, and 9 families with an unbalanced IT accompanied by additional copy-number changes in the vicinity of the insertion sites. To search for CNVs in the parents with the apparently balanced ITs and characterize their structure, we have designed customized high resolution array CGH. In conclusion, we propose that a significant fraction of apparently balanced ITs may harbor small flanking copy-number gains and is caused by DNA replication errors.

2502F

Molecular and clinical delineation of the 2p15-16. 1 microdeletion syndrome. *J. Levy[1], A. Coussement[2], C. Dupont[1], F. Guimiot[1], Y. Capri[1], G. Viot[2], C. Baumann[1], S. Passemard[1], A. Verloes[1], C. Leroy[1], B. Benzacken[1], JM. Dupont[2], AC. Tabet[1].* 1) Genetic Department, Robert Debre Hospital, APHP, Paris, France; 2) Cytogenetic unit, Cochin hospital, APHP, Paris, France.

Interstitial 2p15-p16. 1 microdeletion is a rare microdeletion syndrome previously reported in 14 patients. It is characterized by moderate to severe intellectual disability, autism spectrum disorder, microcephaly, short stature, dysmorphic features, structural brain anomalies and multiple congenital organs defects. It is considered a contiguous gene syndrome involving deletion of several genes. Deletions previously reported are variable in size (from 203 kb to 6. 9 Mb) and have non recurrent breakpoints. We report here three additional patients carried a 2p15-p16. 1 microdeletion shown by SNP-array analysis (Cyto12-SNP, Illumina) : a prenatal case and two postnatal cases. The prenatal and one of the postnatal cases share an overlapping 2p15p16. 1 deletion of 105 kb including only *XPO1* and the distal end of *USP34*. Both patients presented features overlapping the clinical spectrum of the 2p15p16 microdeletion syndrome including dysmorphic facial features and brain structural abnormalities. We confirm that one or both genes are probably involved in facial dysmorphic features, cognitive impairment and brain structural abnormalities observed in 2p15p16. 1-deletion syndrome. The third patient we reported on is a 4 years-old male with an heterozygous de novo a 427 kb deletion at 2p16. 1 (chr2:60624940-61051867) containing *BCL11A* and *PAPOLG* and a phenotype characterized by speech delay, autistic traits and stereotyped behavior but without microcephaly. Considering previous deletions in the 2p15p16. 1 region and our three new cases, we precise the genotype-phenotype correlation of the microdeletion syndrome. Moreover we suggest that the *REL* gene could be considered as a candidate gene for microcephaly.

2503F

Is there significant evidence to associate intronic deletions in the *AUTS2* gene with the typical features seen in *AUTS2* microdeletion syndrome? *R. E. Pyatt[1,2], A. McKinney[1], L. Erdman[1], C. Weber[1], M. Haughn[1], E. de Los Reyes[3], S. Thomas[4], S. E. Hickey[5,6], J. Indyk[5], L. Lehwald[7], S. Santoro[6], D. Lamb Thrush[1,5], S. Hashimoto[1], C. Astbury[1,2].* 1) Pathology and Laboratory Medicine, Nationwide Children's Hospital, Columbus, OH; 2) Department of Pathology, The Ohio State University, Columbus, OH; 3) Department of Neurology, Nationwide Children's Hospital, Columbus, OH; 4) The Center for Child Development, Wood County Hospital, Bowling Green, OH; 5) Department of Pediatrics, The Ohio State University, Columbus, OH; 6) Molecular and Human Genetics, Nationwide Children's Hospital, Columbus, OH; 7) Pediatric Neurology, Nationwide Children's Hospital, Columbus, OH.

Exonic deletions in *AUTS2* have been associated with syndromic features including low birth weight, intellectual disability, autism, short stature, microcephaly, cerebral palsy, and mild facial dysmorphisms. Exonic deletions have also been linked with ADHD, heroin dependence, schizophrenia, and speech/language disorder. As part of routine clinical microarray testing, we have occasionally seen deletions in *AUTS2* which are limited to a single intron and interpretation of these findings can be challenging. Oksenberg et al (2013) demonstrated functional enhancers in zebrafish and mice in introns 1 through 4 and 6. A review of the Database of Genomic Variants showed no reports of copy number variation limited to intron 1 and rare reports of individuals with intron 4 losses. Copy number variation was more frequently observed within introns 5 and 6. The Clinical Genome Research database contains a single clinical case of an intron 1 deletion which was *de novo* and interpreted as pathogenic in an individual with autism. We have assembled a series of clinical cases to investigate whether intronic *AUTS2* deletions are associated with a common phenotype. This series included two cases with deletions limited to intron 1, two deletions within intron 4, one deletion within intron 5, and one deletion within intron 6, which were all interpreted as variants of unknown significance. Inheritance information was not available for any of these cases. The intron 5 deletion was co-identified with a pathogenic duplication at 2q13 in a 2 month old with cleft lip/ palate and subtle dysmorphic features. One of the intron 1 losses was co-identified with a pathogenic 14q32. 3q32. 31 loss in an individual with learning difficulties, precocious puberty, low birth weight, growth retardation, short stature, and hypotonia. The second intron 1 loss was reported in an individual with ADHD, but autism was excluded using the Autism Diagnostic Observation Schedule. One of the intron 4 deletions was identified in an individual with autism, but the second was reported in an individual with dysmorphic features and gross motor delay, with no concern for autism. The individual with the single intron 6 deletion presented with ataxia, cyclic vomiting, and tonic paroxysmal upgaze but no significant neuropsychological issues. While neurodevelopmental features are present in some cases, this could be the result of ascertainment bias as no common phenotype is evident across this case series.

2504F

Two Cases with De Novo 3q26. 31 Microdeletion Reveal the Role of *FNDC3B* in Human Craniofacial Development. *Y. Cao, E. Mitchell, N. Hoppman.* Cytogenetics Laboratory, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.

Here we report the first two cases with de novo 3q26. 31 microdeletions. An 18-month-old girl with dysmorphic facial features was referred to the Mayo Clinic Cytogenetics laboratory for testing. Chromosomal microarray analysis (CMA) on a peripheral blood specimen identified a novel 513 kilobase deletion at 3q26. 31 (171,677,940-172,191,359, genome build hg19), containing two known genes (*FNDC3B*, OMIM #611909; *GHSR*, OMIM #601898). Subsequently, a second patient with a similar 3q26. 31 deletion was ascertained. In this case, CMA on a peripheral blood specimen identified a 345 kilobase deletion at 3q26. 31 (171,622,716-171,968,102, genome build hg19), containing only the *FNDC3B* gene in a 9-day-old boy with dysmorphic facial features. Subsequent metaphase FISH studies for each case confirmed 3q26. 31 deletion, and parental FISH studies demonstrated that both deletions were de novo; therefore, these two 3q26. 31 microdeletions likely contribute to each patient's dysmorphic features. The 9-day-old male patient has a large anterior fontanelle, broad forehead, mild plagiocephaly, hypertelorism, proptosis, short anteverted nose, clinodactyly, and 5th finger hypoplasia. No additional information is available on the 18-month old patient as she was lost to follow-up. Interestingly, both de novo 3q26. 31 microdeletions share the *FNDC3B* gene in common. FNDC3B (fibronectin domain III-containing protein 3B), also known as FAD104 (factor for adipocyte differentiation-104), was first identified as a positive regulator of adipogenesis in a mouse model. Excitingly, further studies in a mouse model have recently demonstrated that *FNDC3B* is required for normal calvarial bone formation and negatively regulated calvarial cell differentiation through inhibition of BMP/Smad signaling. *fndc3b*-deficient mice have multiple cranial and skeletal malformations, such as craniosynostosis-like premature calvarial ossification, and skeletal deformities in the anterior fontanel and femurs. In summary, we report the first two patients with de novo 3q26. 31microdeletions. Both have dysmorphic features, consistent with the phenotypes seen in *fndc3b*-deficient mice in animal studies, which implies a critical role of *FNDC3B* in human craniofacial development.

2505F

A recurrent 47. 5 kb deletion downstream of the *SHOX* gene: is it pathogenic? *L. Matyakhina, D. Pineda, Z. Xu, G. Li, T. Brandt, G. Richard, J. Meck.* GeneDx Inc, Gaithersburg, MD.

SHOX, located in the pseudoautosomal region 1 (PAR1) of the sex chromosomes, is implicated in human growth. *SHOX*-related haploinsufficiency causes Leri-Weill dyschondrosteosis (LWD), a skeletal dysplasia characterized by short stature and/or the Madelung deformity, and isolated idiopathic short stature (ISS). Deletions of the downstream regulatory domain (DR) that encompasses three active enhancer sequences (AESs) have been reported to result in a phenotype similar to that of patients with mutations in the *SHOX* coding region. Reported deletions varied in size but generally included a common 200 kb region starting 48 kb downstream of *SHOX*. Smaller deletions containing one or two AESs have been also observed. A recurrent 47. 5 kb deletion, possibly involving another potential AES and located 160 kb downstream of the *SHOX* gene, has been recently suggested to cause LWD and ISS. We retrospectively reviewed 6672 cases referred to our diagnostic laboratory for chromosome microarray (CMA) to determine if any had a deletion limited to the DR and if the reported phenotype was consistent with LWD or ISS. Using a custom designed Agilent 180K array, we detected the heterozygous 47. 5 kb deletion in 30 patients with various phenotypes, primarily including developmental disorders. Of these, none were referred due to ISS and only 3 patients had some features overlapping with LWS: one patient had short stature, the other had radio-ulnar dysostosis, and the third had Madelung deformity. However, each of these cases had additional clinical features, such as cardiac defects, ADHD and café-au lait spots, and intellectual disability, respectively. Fourteen of the 30 cases (46%) showed additional array findings: 9 cases had pathogenic aberrations, 4 cases had copy number variants of unknown significance, and one patient had multiple blocks of homozygosity. The age of our patients varied from newborn to 34 years old, with 10 patients older than 9 years. Though deletion of AES is thought to have reduced penetrance as compared to individuals carrying mutations within the *SHOX* gene or larger downstream PAR1 deletions, our data do not support pathogenicity of this 47. 5 kb deletion for LWD or ISS and indicate that more unbiased data are needed to clarify its clinical significance.

2506F

Clinically Relevant Copy Number Variants Detected In Cerebral Palsy. *M. J. Gazzellone[1,2], M. Oskoui[3], B. Thiruvahindrapuram[1,2], M. Zarrei[1,2], J. Andersen[4], J. Wei[1,2], Z. Wang[1,2], R. F. Wintle[1,2], C. R. Marshall[1,2,5], R. D. Cohn[2,6,7,8], R. Weksberg[2,6,9], D. J. Stavropoulos[10], D. Fehlings[11], M. I. Shevell[3], S. W. Scherer[1,2,12].* 1) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 3) Departments of Pediatrics and Neurology/Neurosurgery, McGill University, Montreal, Quebec, Canada; 4) Department of Pediatrics, University of Alberta, Edmonton, Alberta, Canada. Glenrose Rehabilitation Hospital, Edmonton, Alberta, Canada; 5) Genome Diagnostics, Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 6) Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada; 7) Centre for Genetic Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 8) Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, Ontario, Canada; 9) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada; 10) Cytogenetics Laboratory, Department of Pediatric Laboratory Medicine, Hospital for Sick Children, Toronto, Ontario, Canada. Department of Laboratory Medicine and Pathobiology Hospital for Sick Children, Toronto, Ontario, Canada; 11) Holland Bloorview Kids Rehabilitation Hospital, Department of Paediatrics, University of Toronto, Toronto, Ontario, Canada; 12) Department of Molecular Genetics and McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada.

Cerebral palsy (CP) represents a group of non-progressive, clinically heterogeneous disorders that are characterized by motor impairment and early age of onset, frequently accompanied by co-morbidities. The cause of CP has historically been attributed to environmental stressors resulting in injury to the developing brain although interplay with genetic vulnerability is suspected. Guidelines for diagnostic assessment of children with CP recommend against routine genetic testing. However, with emerging evidence for a genetic contribution to CP etiology, and the increasing number of reports of etiologic copy number variants (CNVs) in other neurodevelopmental disorders, we hypothesized that CNVs might be contributory. Using microarrays, we genotyped a population-based prospective cohort of children with CP from the Canadian CP Registry and their parents. *De novo* CNVs were detected in 8/115 (6. 9%) of CP patients (~1% rate in controls). In four children, large chromosomal abnormalities deemed pathogenic were found, and they were significantly more likely to have severe motor impairments than those CP subjects without such alterations (p=0. 04). Rare-inherited CNVs affecting loci of established clinical relevance (e. g. 1q21. 1, 16p13. 11, *DMD*) were also found in other unrelated families. In this cohort, the CNV data would have impacted our diagnosis or classification of CP in 11/115 (9. 6%) of families.

2507F

Novel genotype-phenotype correlations for exonic and intronic *NRXN1* deletions. *C. Lowther[1], M. Speevak[2], M. Gazzellone[3], A. C. Lionel[3], C. R. Marshall[3], S. W. Scherer[3], E. McCready[4], D. J. Stravopoulos[5], A. S. Bassett[1].* 1) Institute of Medical Science, University of Toronto, Toronto, Ontario, Canada; 2) Trillium Health Partners Credit Valley Site, Toronto, Ontario, Canada; 3) The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada; 5) Cytogenetics Laboratory, Department of Pediatric Laboratory Medicine, The Hospital for Sick Children, Toronto, Ontario, Canada.

Background: Rare pathogenic copy number variations (CNVs) overlapping *NRXN1* exons have been identified in individuals with neurodevelopmental and/or neuropsychiatric conditions. However, consistent genotype-phenotype correlations have not been established. Also, few intronic *NRXN1* deletions have been reported and their pathogenicity remains unclear. Objectives: To 1) further characterize the extended phenotype of exonic *NRXN1* deletions and 2) systematically evaluate the pathogenicity of intronic *NRXN1* deletions in order to help inform clinical diagnostic practice. Methods: We examined clinically ascertained cases from three Canadian cytogenetic laboratories for novel exonic *NRXN1* deletions. Referring physicians were contacted and invited to complete a clinical checklist including a comprehensive medical, psychiatric, and physical assessment to identify major lifetime features and illnesses. One lab's clinically referred cases were available to investigate the prevalence of intronic *NRXN1* deletions in comparison to 15,254 controls. For all *NRXN1* deletion cases, genome-wide CNV data from high resolution microarrays were investigated for the presence of additional rare variants. Results: We identified 41 (0. 21%) novel exonic *NRXN1* deletions among 19,263 clinically referred cases, an over eight-fold increase compared to controls (p<0. 0001). Novel phenotypes identified in two or more exonic *NRXN1* deletion cases included movement disorders, automutilation, sleep disorders and obsessions and preoccupations. All nine of the adults with exonic *NRXN1* deletions had a dual diagnosis of intellectual disability and a psychiatric illness. In contrast, the prevalence of congenital anomalies in exonic *NRXN1* deletion cases was low. There was no difference in the prevalence of intronic *NRXN1* deletions between clinical cases (19/6,022; 0. 32%) and controls (55/15,524; 0. 35%). Additional pathogenic variants located elsewhere in the genome were four times more common in intronic (n=6/19; 31. 6%) compared to exonic (n=3/41; 7. 3%) *NRXN1* deletion cases. Conclusions: These data suggest that the expression of exonic *NRXN1* deletions is primarily neuropsychiatric and may associated with comorbid intellectual disability and a psychiatric condition in adulthood. To our knowledge this is the first study to demonstrate that the majority of intronic *NRXN1* deletions by themselves are unlikely to cause clinical phenotypes, however further study on their potential functional impact is needed.

2508F

The complexity of non-recurrent duplications detected by microarray: a review of 167 cases. *J. Nicholl[1], S. Brown[1], C. Barnett[2], E. Haan[2], E. Thompson[2], J. Liebelt[2], L. McGregor[2], L. Hinton[1], S. Smith[1], A. Attwood[1], Y. Hull[1], S. Yu[1].* 1) Cytogenetics, SA Pathology, North Adelaide, SA, Australia; 2) South Australian Clinical Genetic Service, North Adelaide, SA, Australia.

Between May 2010 And December 2014 our laboratory has carried out microarray on ~7,700 patients with ID/DD/dysmorphism/malformation. Our laboratory uses reporting criteria whereby duplications < 400kb are not reported unless they contain one or more OMIM disease related genes. We have identified ~550 duplications involving autosomes using these criteria. Excluding unbalanced translocations, marker chromosome and recurrent microduplications, ~167 duplications were classified as non-recurrent duplications. We have reviewed the oligo/SNP array profiles, FISH results and family follow-up information for these non-recurrent duplication cases and summarised the outcomes and complexity. TRIPLICATIONS: Five changes had Log2 ratios which suggested 3 or more copies; including a 3Mb triplication at 15q21. 1 which segregates with developmental delay in a 3 generational family and a *de novo* 1. 3Mb region at chromosome 6p24. 1-p25. 1 involving 6 OMIM genes. Inspection of the array profile revealed that these triplications were derived from normal two copies, not from a duplicated region. INSERTIONAL TRANSLOCATIONS: Five duplications were the result of insertional translocations; including a duplication at chromosome 14q32 inserted into Xq27. 3, adjacent the FMR1 gene, and a paternally inherited 7q11. 22 intra-chromosome insertion. FISH is the only way to detect these cases. MORE THAN ONE CNV: Five patients were found to have a 2nd change (all deletions). In all five patients, the2nd change was considered likely to be the cause of the patient's medical condition. PATHOGENICITY: 29 (17%) were considered likely pathogenic based on gene content, size or inheritance. Sixty eight duplications (41%) were classified as "unknown clinical significance", with forty four involving a single inherited duplication as the only copy number variant (cnv). The remaining twenty four cases had a more complex picture with either more than one inherited cnv, or involving a disease gene not related to the patient's phenotype. Twenty patients (12%) had inherited duplications classified as non-pathogenic based on available evidence (DECIPHER, ISCA, DGV literature, segregation in the family). We concluded that the non-current duplications can be very complex and routine FISH and parental /family testing are necessary to reveal the complexity and to determine the pathogenicity of the duplication.

2509F

Comparison of CNVs results in the group of the patients with autistic spectrum disorders by MLPA and aCGH – one center experience. *J. Srovnal[1,2], P. Capkova[2], H. Slavik[1], V. Curtisova[2], K. Adamova[2], K. Staffova[1], R. Vrtel[2], M. Prochazka[2], M. Hajduch[1].* 1) Institute of Molecular and Translational Medicine, Faculty of Medicine and Dentistry, Palacky University, Olomouc, Czech Republic; 2) Department of Medical Genetics, Faculty of Medicine and Dentistry, Palacky University and University Hospital Olomouc, Czech Republic.

Abstract: Autistic spectrum disorders (ASD) are characterized by simultaneous deficits in three domains of behavior: reciprocal social interaction, communication and stereotyped and restricted behaviors. Submicroscopic CNVs can have causal or susceptibility-related role in heritability ASD. The aim of our one-center study was to compare MLPA and aCGH results in ASD patients. **Method:** In total, 42 patients with ASD characteristics were referred for genetic testing using MLPA and aCGH. Genomic DNA purified from peripheral blood was obtained after informed consent. MLPA kits (SALSA MLPA P343 Autism 1, P297 and P245 Microdeletions, P070 and P036 Subtelomers, MRX P106, MRC-Holland, Amsterdam, Netherlands) and Cytoscan HD (Affymetrix, Santa Clara, CA, USA) were used for CNV detection. **Results:** In the group of the children with ASD three likely pathogenic CNVs were identified. Microduplication 16p11. 2 (576kb) inherited from father, microduplication 1q. 21. 1q21. 2 (1. 3Mb) de novo, unbalanced translocation with microdeletion 20q13. 33 (1. 58Mb) and microduplication 22q13. 33 (app. 170-190kb) covering SHANK3 de novo. MLPA results were confirmed in all cases using aCGH. In another four cases, the inherited variants of unknown significance involving genes NRXN1, SNRPN, MAPK3 and CD160 were identified using aCGH not MLPA. The microdeletion 16p11. 2 (611kb) was incidentally found in the female patients with Mayer-Rokitansky-KØster-Hauser Syndrome (MRKHS) without autistic features. The loss of 1. 3Mb in 15q13. 2q13. 3 was also discovered by MLPA P343 Autism 1 in the patient suffered from intellectual disability but without ASD. **Conclusion:** SALSA MLPA probemix P343 Autism 1 has proved to be efficient for detection CNVs in the group of children with ASD. However, the simultaneous tests with another probemixes (P297, Human Telomere 3 and 5 etc.) or with aCGH are recommended especially when intellectual insufficiency in the patient is present. The high variability in the CNVs expression suggests polygenic etiology in ASD. Massive parallel sequencing is a promising platform for further research in this area. **Disclosure:** Nothing to disclose. **Acknowledgment:** This work was financially supported by IGA UP LF 2015_010, NPU LO1304, CZ. 1. 05/2. 1. 00/01. 0030 and CZ. 1. 07/2. 3. 00/30. 0004.

2510F

Deciphering isolated 4q deletion syndrome cases with a focus on an inherited case of 4q35. 1q35. 2 deletion. *N. Inamdar[1], R. W. MARION[2], R. NAEEM[3].* 1) MOLECULAR PATHOLOGY, MONTEFIORE MEDICAL CENTER, 1635 POPLAR ST. , BRONX, NEW YORK, NY 10461; 2) THE CHILDREN'S HOSPITAL AT MONTEFIORE, 3415 BAINBRIDGE AVE, BRONX, NY 10467; 3) MOLECULAR PATHOLOGY, MONTEFIORE MEDICAL CENTER, 111 EAST 210TH ST. , BRONX, NEW YORK, NY 10467.

4q deletion syndrome is characterized by a proximal, interstitial, or distal regional loss of the long arm of chromosome 4. Phenotypic presentation varies even among individuals with overlapping deleted intervals and the syndrome is rare with an incidence of about 1 in 100,000. Making phenotype-genotype correlations is therefore a challenge and there is hence, a need to document all cases of 4q deletion syndrome. To contribute to the knowledge base of this syndrome we reviewed our aCGH cases for the last seven years and identified 13 individuals with isolated 4q deletion. Six of them, including three members of one family, harbor a distal deletion. Deletion intervals were identified to be 4q32. 2q35. 2, 4q32. 3, 4q32. 3q35. 2, and 4q35. 1q35. 2. We used 44K Agilent microarray. We present phenotype-genotype correlations for all available cases and further studies of one family of two siblings and their father with heterozygous ~4Mb deletion in 4q35. 1q35. 2 region. The deletion was inherited by the siblings from their apparently phenotypically normal father. The proband is a four year old (at the time of diagnosis) male with developmental delay, mildly affected sleep pattern, and autism spectrum-like disorder. Of the several known genes localized to the deleted region, two are of particular relevance: MTNR1A and FAT1. MTNR1A, one of the melatonin receptors, is expressed in the hypothalamic suprachiasmatic nucleus that is associated with circadian rhythm. It is interesting to note that variants of the protocadherin gene FAT1 have not only been associated with autism spectrum and bipolar affective disorders but more recently have been linked to facioscapulohumeral muscular dystrophy (FSHD)-like phenotype. Neuromuscular deficits are a common finding in 4q deletion syndrome. While the exact gene/s responsible for FSHD is/are not known, chromatin remodeling arising from D4Z4 repeat truncation and/or hypomethylation of 4q35. 1q35. 2 region is thought to deregulate expression of genes located in the vicinity leading to FSHD. It would be interesting to decipher if similar change in methylation status/chromatin remodeling occurs in 4q cases leading to deregulation of genes flanking the deleted regions. Such a mechanism could potentially be responsible for the phenotypic variations/incomplete penetrance seen in individuals harboring the same deletions as in our inherited case where the father is apparently unaffected.

2511F

Array-CGH analysis in fetuses with multiple malformations reveals 50% of cases with pathogenic deletions/duplications or variants of unknown significance. *E. Di Gregorio[1,2], G. Gai[1], G. Botta[3], A. Calcia[2], P. Pappi[1], F. Talarico[1], E. Savin[1], M. Ribotta[3], S. Cavalieri[1], E. Giorgio[2], C. Mancini[2], M. Ferrero[2], E. Pozzi[2], E. Riberi[2], G. Restagno[4], G. B. Ferrero[5], E. Grosso[1], A. Brusco[1,2], A. Brussino[2].* 1) Città della Salute e della Scienza University Hospital, Torino, Italy; 2) University of Torino, Department of Medical Sciences, Turin, 10126, Italy; 3) Città della Salute e della Scienza University Hospital, Department of Pathology, Turin, Italy; 4) Città della Salute e della Scienza University Hospital, Laboratory of Molecular Genetics, Turin, Italy; 5) University of Torino, Department of Public Health and Pediatrics, Turin, Italy.

Major congenital malformations (MCMs) are anatomic anomalies affecting life expectancy or organ normal function occurring in about 2-3% of live born infants. Fetuses with abnormal ultrasound findings show chromosomal aberrations detected by standard karyotype in 9–35% of the cases, depending on the presence of single or multiple anomalies. The etiology of MCMs remains uncovered in the majority of cases. In the last few years, several studies evaluated the diagnostic role of array-CGH analysis (a-CGH) on fetuses with malformations detected in prenatal diagnostic setting or after interrupted pregnancy. While a-CGH diagnostic yield in the postnatal population has been established at 10-12%, studies investigating fetuses reported pathogenic variants in 0-12. 5% of cases. The aim of our study was to evaluate the diagnostic yield of array-CGH analysis in a selected group of fetuses with MCMs. We used whole-genome a-CGH (60K Agilent) to investigate 28 fetuses presenting at least two MCMs detected on autopsy with or without ultrasound. Standard karyotype was performed in 18 out the 28 cases and failed to provide a diagnosis. Array-CGH disclosed Copy Number Variants (CNVs) in 50% of the analyzed fetuses (14/28). We identified a clinical significance in 32% of the cases (9/28) and variants of unknown pathogenic significance (VOUS) in 18% (5/28). Three of the nine positive cases carried a known microdeletion: a 22q11. 21 deletion, a central chromosome 22q11. 21 deletion and a TAR syndrome). In six the rearrangement was always >11 Mb and the gene content or the region involved were correlated with the phenotype. VOUS were 130-900 kb in size (four duplications and one deletion) and contained 1-7 genes. In conclusion, our study showed a-CGH is a powerful technique to unravel the genetic etiology of fetal major congenital malformations. Differences in the proportion of significant results among various cohorts may be related to the design of the array used and the different inclusion criteria based on the severity and number of fetal malformations.

2512F

Unusual translocations t(2;11)(q23;q13) y t(9;14)(q12;q11. 2) in acute promyelocytic leukemia. *A. Corona-Rivera[1,2], C. Almodóvar-Cuevas[2], E. Torres-Anguiano[1], C. Ortega de la-Torre[2], H. Pimentel-Gutiérrez[2], C. Barba-Barba[2], G. Serafín-Saucedo[2], G. Macías-Salcedo[2], M. Velázquez-Rivera[2], A. Márquez-Mora[2], S. Gallegos-Castorena[3], F. Sánchez-Zubieta[3], I. Bobadilla-Morales[1,2].* 1) Universidad de Guadalajara,Centro Universitario de Ciencias de la Salud, Instituto de Genética Humana "Dr. Enrique Corona Rivera", Laboratorio de CitoGenética Genotoxicidad y Biomonitoreo, Guadalajara, Jalisco, México; 2) Unidad de CitoGenética, Nuevo Hospital Civil "Dr. Juan I. Menchaca", División de Pediatría, Servicio de Hematología Oncología Pediátrica , Guadalajara, Jalisco, México; 3) Servicio de Hematología Oncología Pediátrica , Nuevo Hospital Civil "Dr. Juan I. Menchaca", Guadalajara, Jalisco, México.

Acute promyelocytic leukemia (AML-M3) is a biologically and clinically distinct variant of AML, approximately 92 percent of patients with AML-M3 carry the balanced translocation t(15;17). It involves the retinoic acid receptor alpha (*RARA*) gene on chromosome 17 and the*PML*gene on chromosome 15. Although several variant translocations involving *RARA* have been identified, including t(11;17) and t(5;17), a small number of AML-M3 cases lacking t(15;17) have been described**.** We present the cytogenetic alterations of a four-year-old patient with a diagnosis of AML-M3, without the usual translocation t(15;17). A four-year-old female diagnosed with AML-M3, with two years of disease evolution. Laboratory findings showedleukocytosis (20,000/µl), lymphocytes of 35. 4%, neutrophils of 58. 9%, Hb 7. 92 g/dl, hematocrit of 24. 7%, MCV 86. 3 fl, MCH 27. 7 pg, platelet count of183,000/dl. The treatment began with the national protocol and two subsequent cycles continued PETHEMA 99 protocol. The patient achieved complete remission until the second cycle, after seven months in the maintenance phase she presented a bone marrow relapse. Mitoxantrone was administered as a rescue protocol and reaches the second complete remission up to the second cycle. Six months after relapse, she died by therapeutic failure. At diagnosis, cytogenetic studies demonstrated the absence of translocation. By RT-PCR analysis, the*PML/RARA*transcript was not detected. On the other hand, *PML/RARA*fusion signal was not detected by FISH analysis. Cytogenetic study was carried out on cells from bone marrow aspiration in relapse stage. The karyotype was interpreted as 45~46,XX,t(2;11)(q23;q13),t(9;14)(q12;q11. 2),18[11]/46XX[8]. This case report reveals different alterations to the t (15; 17) that produce the same phenotype AML-M3. There are no reports of translocations with cleavage sites observed in this type of disease. In the present case, treatment with ATRA and chemotherapy showed a poor response. These findings may indicate that these translocations are indicators of poor prognosis.

2513F

A t(17;19)(q21;p13. 3) involving *TCF3* , a t(1;9)(p13;p13) and a 5'*IGH@* abnormality in a case of adult B-cell Acute Lymphoblastic Leukemia. *R. Chow[1], D. Shabsovich[1], G. Schiller[2], M. Kallen[1], S. T. Pullarkat[1], N. Rao[1], C. A. Tirado[1].* 1) Pathology & Laboratory Medicine, UCLA, Los Angeles, CA; 2) Medicine, Hematology & Oncology, UCLA, Los Angeles, CA.

TCF3(19p13. 3) abnormalities are relatively common in B-cell acute lymphoblastic leukemia (B-ALL). Although the t(1;19)(q23;p13) involving*PBX1*is the most common of these rearrangements, rare variants involving different partner genes have been observed. Among these, the t(17;19)(q21;p13. 3), which results in a chimeric gene involving*TCF3*and*HLF*(17q21), is associated with an extremely poor prognosis in B-ALL. Herein, we present the case of a 25-year-old male diagnosed with B-ALL who underwent a hematopoietic stem cell transplant (HSCT) and experienced multiple relapses. Conventional cytogenetic analysis prior to the patient's first relapse revealed a t(17;19)(q21;p13. 3). Fluorescence in situ hybridization (FISH) analysis confirmed that the t(17;19) involved*TCF3*and also revealed a loss of the 5'*IGH@*(14q32) region, suggesting a deletion or an unbalanced rearrangement involving 14q32. Conventional and molecular cytogenetic analysis after the first relapse prior to HSCT revealed a t(3;10)(q27;q24) as an additional abnormality. After undergoing a HSCT, the patient experienced another relapse, at which point a bone marrow biopsy revealed a hypocellular marrow with lymphoblasts consisting of 85% of total marrow elements. Conventional cytogenetic studies revealed an abnormal karyotype bearing the (17;19) translocation, a t(1;9)(p13;p13), and complex structural anomalies involving chromosomes 5, 7, and 14. FISH studies confirmed the involvement of*TCF3*and structural abnormalities involving chromosomes 5 and 7, but did not show any*IGH@*abnormalities. As of now, the patient is undergoing re-induction therapy with Blinatumomab. The t(1;9)(p13;p13) has been reported in a single case of B-ALL involving*PAX5*(9p13) and*HIPK1*(1p13) but to the best of our knowledge, this is the second reported case of t(17;19) bearing a concomitant 9p13 rearrangement as a secondary abnormality in the context of a complex karyotype and the first reported case of t(17;19)-ALL with a concomitant*IGH@*(14q32) abnormality. This case provides insight into the clonal evolution of t(17;19)-ALL and the possible involvement of*PAX5*and*IGH@*in this subset of malignancies.

2514F

Unusual phagocytosis in a woman with a non-specific B-cell chronic lymphoproliferative disorder. *M. Dominguez[1], C. Borjas[1], M. Magaña[2], J. González[2].* 1) Universidad de Guadalajara, Guadalajara, Jalisco, México; 2) División de Genética, CIBO, Instituto Mexicano del Seguro Social, México.

Introduction. Phagocytosis is the mechanism by which a dead or dying cell is internalized and degraded by another living cell. The normal process of phagocytosis comprises several steps: invagination, engulfment, internalization, formation of the phagocytic vacuole, fusing of lysosomes to the phagocytic vacuole, and release of digested products. We describe an unusual phagocytic activity in a woman with a non-specific B-cell chronic lymphoproliferative disorder. **Case.** A 78 year-old woman had abdominal pain, fever, hepatomegaly, left axillary and mediastinum lymph node. Values of laboratory test were: Hb 11. 4 g/dL; 34,500 leucocytes/mm3 (with 76% of lymphocytes); 293,000/mm3 platelets; HIV, hepatitis B and Hepatitis C were negative. Immunophenotype of peripheral white blood cells was concordant with a non-specific B-cell chronic lymphoproliferative disorder. Nearly after the cytogenetic study, the patient passed away. **Materials and methods.** Chromosomes were obtained from a peripheral blood culture stimulated with a mixture of phorbol myristate acetate and Pokeweed mitogen. The karyotype analysis was done by GTG-banding. Posteriorly, we performed a fluorescent in situ hybridization with the AML1/ETO probe (Cytocell). **Results and discussion.** We observed the karyotype 46,XX,inv(14)(q11q32)[2]/46,XX[20]. Additionally, at the GTG-banding analysis, we observed several strange dyad-cell-like structures (or an hourglass-cell conformation) that looked like cell-in-cell structures in ~3% of the total cells. In several of these dyad-cell-like structures seemed as if they were suffering a nuclear fusion process. FISH analysis with the dual color AML1/ETO probes disclosed two green and two red expected signals on only one of the two cells of each dyad; whereas, the another cell was negative for the expected FISH signals. In addition, we found that 57 of 348 (16%) individual cells in our patient showed any AML1/ETO signals, which was statistically significant ($p<0.001$) when compared with values observed in a control sample where 12 of 349 (3. 5%) individual cells had no FISH signals. We discarded the occurrence in vitro of such dyad-cell structures, because they were present in the patient's peripheral blood and bone marrow smears taken the same day we did the cell culture. Conclusion. A strange form of phagocytosis was present in this patient. The main difference between the patient's phagocytosis with the normal process is the absence of an internalized phagocytic vacuole.

2515F

An adult male presenting with concurrent plasma cell myeloma involving an *IGH@/CCND1* translocation and chronic myelogenous leukemia with a variant (9;22) translocation. *C. Tirado, D. Shabsovich, Y. Naeini, N. Rao.* Pathology & Laboratory Medicine, UCLA, Los Angeles, CA 90024.

The t(11;14)(q13;q32) involving *IGH@* and *CCND1* as well as the t(9;22)(q34;q11. 2) involving *BCR* and *ABL1* are common abnormalities in plasma cell myeloma (PCM) and chronic myelogenous leukemia (CML), respectively. However, the concurrence of the two malignancies is extremely rare and reported infrequently in the literature. Herein, we present the case of an 87 year-old male who presented with anemia and monocytosis, hemoglobin of 10. 2 g/dL with mean corpuscular volume of 95. 7 fL and white blood cells count of 11. 9 K/uL, with a differential of 50. 9% granulocytes, 41. 8% lymphocytes, and 7. 3% monocytes. A bone marrow biopsy showed a mildly hypercellular marrow (50% cellularity) with mild myeloid hyperplasia. Plasma cells were increased in number and they occurred scattered in the marrow. Identified were 8-10% monoclonal kappa-restricted plasma cells, consistent with a plasma cell dyscrasia. Flow cytometry performed on the bone marrow aspirate demonstrated B-cells with normal antigen expression and kappa:lambda ratio, with no evidence of a monotypic B-cell population. Granulocytes were left shifted but they showed normal antigen expression. Monocytes also showed normal antigen expression. CD34-positive blasts were within the normal range and comprised 1% of the total cells. An abnormal plasma cell population was identified, comprising approximately 2. 8% of the total cells. This population showed kappa light chain restriction, and was positive for CD38, CD138, partial CD45, CD20, dim CD27, and CD200, and negative for CD19, CD56, CD117, CD28, and CD81, consistent with a plasma cell dyscrasia. FISH studies on a sample enriched for plasma cells detected a t(11;14) involving *IGH@* and *CCND1* in 92% of nuclei, consistent with a plasma cell neoplasm. However, conventional cytogenetic analysis on a sample not enriched for plasma cells revealed an apparent t(9;22)(q34;q11. 2) in 40% of metaphases. Interphase and metaphase FISH studies on the same sample confirmed the presence of the *BCR-ABL1* fusion in 88% of nuclei, but did not show any signals corresponding to the derivative 9, suggesting a variant t(9;22) with a deletion or additional material of unknown origin at the 9q34 band of the derivative 9 and a derivative 22 bearing the *BCR-ABL1* fusion gene. This case provides insight into the coexistence of PCM and CML. .

2516F

Myelodysplasia transforming to mast cell leukemia with t(9;22): A rare occurrence. *G. Velagaleti[1], C. Mendiola[1], M. Shupe[2], V. Ortega[1].* 1) Dept Pathology, Univ Texas Hlth Sci Ctr, San Antonio, TX; 2) Dept Hematology/Oncology, San Antonio Military Medical Center, San Antonio, TX.

Mast cell leukemia (MCL) is a rare form of systemic mastocytosis in which an abnormal increase in mast cells (MCs) proliferates in tissues. MCs are distinct hematopoietic, multifunctional cells originating from uncommitted myeloid progenitor cells expressing a unique profile of antigens, involved in vascular cell regulation and allergic disease states. MCs involvement in myelodysplastic or leukemic disease states have rarely been described and mast cell lineage involvement in leukemic disease states is an uncommon event. There are limited reports of AML/MDS with transformation to MCL. Although there are clonal abnormalities reported in both MCL and mastocytosis, there are no characteristic or consistent chromosome abnormalities identified. We report a unique case of high risk MDS with progression to MCL with t(9;22). An 80-year-old female presented with new onset pancytopenia in 2012. Chromosome analysis on the bone marrow showed 46,XX,del(20)(q11. 2q13. 1)[16]/46,XX[4] and FISH analysis with AML/MDS panel showed 20q deletion and monosomy 7/del 7q. Following palliative chemotherapy with azacitidine, she did not show remission with follow-up studies in 2013, showing similar cytogeneticsas before but without the monosomy 7/7q deletion. In 2014, she presented to the ER with complaints of weakness, nausea, vomiting, as well as abdominal pain. Bone marrow biopsy showed features consistent with MCL transforming from her MDS. Bone marrow chromosome analysis showed clonal evolutionwith the karyotype 46,XX,del(20)(q11. 2q13. 1)[10]/46,idem,t(9;22)(q34;q11. 2)[7]/46,XX[3]. Additional FISH studies with BCR/ABL probes confirmed BCR/ABL rearrangement. FISH analysis withBCR/ABLon her two previous stored bone marrow samples showedno BCR/ABL rearrangement. Although MDS transforming to AML or ALL with t(9;22) has been reported, presence of BCR/ABL fusion in MCL transformed from MDShas never been reported before. Overall MCL carries a poor prognosis given the aggressiveness of the disease and ineffective treatments. Due to inability to tolerate aggressive chemotherapy and the poor prognosis associated with this rare case of MCL with t(9;22), our patient was counseled and provided palliative care after which she passed away. Further investigation may provide insight to the molecular basis of this rare disease and the clinical significance of the t(9;22), if any, in our patient, and allow guidencefor proper management and targeted therapies for these patients.

2517F

Double trisomy (48, XXY,+ 21) in a foetus associated with Atrioventricular septal defects (*AVSD*) and agenesis of the corpus callosum. *S. K. Bhattacharya[1], U. Radhakrishna[2], V. Lal[1].* 1) Department of Cytogenetics, Dr. Lal Path Labs. Pvt. Ltd. , New Delhi, India; 2) Department of Obstetrics and Gynaecology, Oakland University William Beaumont School of Medicine, Royal Oak, MI, USA.

The incidence of double aneuploidy in the same individual is relatively uncommon. In addition dual aneuploidies involving both autosomal and sex chromosomes have rarely been reported to be associated with developmental defects such as congenital heart defects (CHD). Most reported cases of double aneuploidy are presented in the form of recurrent miscarriages and abortions. A 20-weeks old male fetus with double aneuploidy associated with atrioventricular septal (*AVSD*)and other developmental defects has been analysed. This is the first child of healthy and non-consanguineous parents, which was aborted after 20 weeks of gestation. The estimated foetal weight was 232 gm (35th percentile). Using 4-D ultrasonography, the cavum septum pellucidum in the brain was not observed. Electroencephalogram showed mildly diffuse cortical dysfunction without abnormal epileptiform discharges. Lateral ventricles were also dilated. Only a thin rim of cerebral cortex was present. Choroid plexus and medial wall of ventricle separation was 7. 2 mm (normal < 3 mm). All the conditions were suggestive of agenesis of corpus callosum. Fetal echocardiographic findings indicated congenital heart disease with atrioventricular septal defects (AVSD). Chromosomal and Fluorescence in situ hybridization (FISH) techniques with LSI21 and DYZ3 alpha satellite and CEP X alpha satellite probes analysis showed a karyotype of 48,XXY,+21: a double aneuploidy of DS and Klinefelter syndrome (KS), very few such cases have been reported in the literature. A detailed genetic and clinical report of the present case and current knowledge on causes and consequences of double aneuploidy in humans will be presented.

2518F

Severity of oro-dental phenotype in Down Syndrome based on chromosome analysis. *E. Severin[1], A. PŭN[1], R. Baltag[1], A. Stan[2], C. Funieru[1].* 1) Carol Davila University of Medicine and Pharmacy, Bucharest, Romania; 2) Genetic Lab - Bucharest, Romania.

The extra chromosome 21 or part of its long arm (including 21q22 band) can come indistinct genetic ways, such as, full trisomy 21, mosaic trisomy 21 or translocation trisomy 21 causing Down's syndrome (DS). The severity of clinical features, including oral-facial abnormalities, varies from an individual with DS to another and it is hard to predict how much a fetus/newborn will be affected as a child or an adult. DS has the highest incidence at birth of any chromosomal abnormality and every paediatric dentist can expect to deal with oro-dental problems relating to DS. The aims of the study were to analyze the oro-dental phenotype of 52 subjects with DS, to determine the pattern of dental anomalies in these cases and to find if the severity of oro-dental features can be predicted based on chromosome analysis. Methods: Children and adults with DS were chromosomally investigated in order to confirm the clinical diagnosis of DS. To characterize the oro-dental phenotypic spectrum of DS persons were performed extra / intra-oral examinations, and radiological evaluations. Results: Based on cytogenetically analysis results, persons with full trisomy 21, mosaicism for trisomy 21 and translocation trisomy 21 were identified. Most oro-dental features including teeth, gingiva, tongue, palate and oclussion anomalies were variable in both frequency and expression. The most common dental anomaly was hypodontia followed by microdontia. All persons with DS presented oclussal problems. Eruption of both dentitions was delayed in mostcases. Conclusion: Variation of oro-dental phenotype is not associated with specifically chromosomal mechanisms of trisomy 21. Clinical relevance: cytogenetic testing is not a good predictor for oro-dental features severity for DS butthe study provides data that could help the dentists to improve their knowledge and clinical practice.

2519F

Trisomy 9p. Rethore syndrome, case report. *R. Silva-Cruz[1], L. Bobadilla-Morales[1,2], HJ. Pimentel-Gutiérrez[2], C. Barba-Barba[2], C. Ortega de la-Torre[2], G. Macías-Salcedo[2], A. Márquez-Mora[2], ME. Velázquez-Rivera[2], JR. Corona-Rivera[1,2], F. Sánchez-Zubieta[2], A. Corona-Rivera[1,2].* 1) Laboratorio de CitoGenética, Instituto de Genética Humana "Dr. Enrique Corona Rivera", CUCS, Universidad de Guadalajara; 2) Unidad de CitoGenética, Servicio Onco-hemato Pediatría, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca.

Introduction. Trisomy 9p, known as Rethore syndrome is a chromosomal abnormality characterized by total or partial duplication of chromosome 9p. It is characterized by delayed closure of sutures, ocular hypertelorism and hypoplastic distal phalanges. The phenotypic variability is large due to the variable size of the trisomic fragment involved. **Objective**: to describe a case of trisomy 9p with conventional and molecular cytogenetic approach, in clinical monitoring. **Clinical report.** Male, 7 months old, mother of 40-year-old with right renal hypoplasia, G1, father of 45-year-old healthy. Nonconsanguineous. He was born at 35 weeks, 1,470 g birth weight, APGAR unknown, spontaneous crying and breathing. Cephalic support at 6 months. FE: head circumference 41 cm, height 61 cm, broad forehead, short neck, trichomegaly of lashes, epicanto, strabismus, bulbous nose, micrognathia, thin lower lip, forked tongue, dysplastic ears, single transverse palmar crease, unique flexion crease in 5th right finger, agenesis of the nipple, umbilical hernia, micropenis, clubfoot. In laboratory tests, T3 3. 4 pq/ml, T4 0. 93 nq/dl, TSH 1. 43 uIU/ml. Hemoglobin 11. 33 g/dl, ALT 28. 20 U/L, AST 37. 20 iU/L. IgG 433 mg/dl, IgA 35 mg/dl, IgM 48 mg/dl, IgE 126. 01 UI/ml. USG abdomen shows left kidney in hemiadomen, normal morphology. Skull tomography without alterations. Karyotype 46,XY,add(9)(p13-p24). nuc ish dup(9)(p13p24)(wcp9+). He goes to early stimulation. Shows improvement in the acquisition of developmental milestones, he attained sitting and emit some sounds and some disyllabic. **Conclusions.** Although it is known that this entity genetic presents variability according to the region of arm p (partial or complete), we can see that our case has phenotypes which are common in Rethore syndrome and some others with low frequencies that often appear in cases considered sporadic.

2520F

Dyskeratosis congenital. A X-linked dominant. Case report. *L. Bobadilla-Morales[1, 2], S. A. Brukman-Jiménez[1, 2], L. Zúñiga-Guzmán[1], S. Borrayo-Dorado[1], J. Franco-Pérez[1], F. Zánchez-Zubieta[2], A. Orozco-Alvarado[1], R. Navarro-Martín del Campo[2], G. Macías-Salcedo[2], A. Corona-Rivera[1, 2].* 1) Instituto de Genética Humana "Dr. Enrique Corona Rivera" CUCS, Universidad de Guadalajara; 2) Unidad de CitoGenética, servicio Onco-hemato Pediatría, Hospital Civil de Guadalajara "Dr. Juan I. Menchaca".

Dyskeratosis congenital is an inherited disorder characterized by muco-cutaneous abnormalities (leukoplakia, reticulate hyperpigmentation and nail dystrophy), bone marrow failure, premature ageing phenotype and an increased risk of malignancy. This phenotype is triggered by known mutations in components of telomerase and telomere shelterin which is clinically and genetically heterogeneous, with X-linked recessive (MIM #305000), autosomal dominant (MIM #127550) and autosomal recessive (AR) (MIM #224230) inheritance patterns. Objective: historically X-linked recessive is the most common inheritance pattern of this disease, our case report suggests a X-linked dominant inheritance pattern. Clinical report: 16 years old female, whose parents do not share clinical phenotype. Hyperpigmentation in the sun exposed areas of the skin appeared when she was 6 years old. At 15 years old, she showed nail dystrophy, reticulate hyperpigmentation and leukoplakia, being candidate to bone marrow transplantation. Weight 39kg (Z= -2. 49), size 1. 50mts (Z= -1. 39), head circumference 53cm (Z= -1. 69), hemoglobin 10. 48g/dl. Bone marrow biopsy: aplastic with fatty infiltration. Karyotype mos46,XX[46]/47,XX,+mar[4]. Chromosomal instability positive. Conclusions: Altogether the clinical manifestations and family tree described in this case are not consistent with the previously reported inheritance patterns of this disease. Being a female patient with severe clinical phenotype, we suggest a X-linked dominant inheritance pattern, which has not been previously reported.

2521F

Our experience at Children's National Health Systems using CytoScan® Dx Assay. *k. Cusmano-Ozog, m. seprish, s. hofherr.* childrens national, washington, DC.

The Molecular Diagnostics Laboratory in the Division of Laboratory Medicine at Children's National Health Systems established chromosomal microarray (CMA) in the summer of 2013 to better support the Division of Genetics and Metabolism with over 7,000 patient encounters a year. In August 2014, we went live with CytoScan® Dx, the first FDA-cleared chromosomal microarray assay. During the first 10 months of offering this assay, 517 samples were received, tested and resulted. Of the 517, 147 (28%) had a positive finding and 370 (72%) were negative. Of the positive results, 111 (21% of all tested) had a copy number variant (CNV) detected and 36 (7% of all tested) had significant area(s) of homozygosity (AOH). For those with a CNV, there were 45 duplications, 51 deletions and 9 complex rearrangements. Additionally, there were six times where aneuploidy was detected: Trisomy 21, Trisomy 18, Trisomy 13, 47,XXY, 47,XYY and 48,XXYY. In addition, there was one case each of mosaic Trisomy 14, mosaic ring X and tetrasomy 18p. The most common deletion syndromes identified were: 22q11. 21 (VCF, n=6), 15q11. 2 (PWAS, n=3), 7q11. 23 (Williams, n=4) and 16p11. 2 (autism, n=3). The remaining deletions and duplications identified were evenly dispersed among the chromosomes. In regards to the inheritance of CNVs detected, thus far 15 are de novo, 29 are maternal and 11 are paternal in origin. Of those with AOH, 18 (3% of all tested) were found to have AOH >2. 5% of the autozygome suggestive of identity by descent and 9 (2% of all tested) were suggestive of uniparental disomy (UPD) or isodisomy. This study demonstrates the high detection rate of CMA, especially in the pediatric population that led initially for it to be recommended as a first-tier test by ACMG and more recently approval by the FDA for the detection of CNVs associated with developmental delay and/or intellectual disability (DD/ID), congenital anomalies, and/or dysmorphic features.

2522F

Discordant Circulating Fetal DNA and Subsequent Cytogenetics Reveal False Negative, Placental Mosaic, and Fetal Mosaic Genotypes. *R. V. Lebo[1], R. W. Novak[1], K. Wolfe[2], M. Michelson[2], H. Robinson[3], M. Mancuso[2].* 1) Dept Pathology, Akron Children's Hosp, Akron, OH; 2) Maternal Fetal Medicine, Akron Children's Hospital, Akron, OH; 3) Genetics, Akron Children's Hospital, Akron, OH.

Noninvasive Prenatal Testing (NIPT) of the circulating fetal DNA in maternal blood (cfDNA, trophoblast DNA, placental DNA) can be rechecked by ultrasound and ordering a fetal sample karyotype by Maternal Fetal Medicine clinics. This abstract reviews two significant additional cases. In the first case, we found twice our previously reported number of abnormal NIPT results now adding a balanced reciprocal translocation that was undetected by NIPT analysis which nonetheless correctly identified trisomy 21. In the second case, further analysis revealed one phenotypically *normal* newborn with a complex mosaic karyotype substantially decreasing the newborn's eventual reproductive fitness. Together these two cases emphasize the prudence of pretest counseling. In addition, the second case establishes the importance of karyotyping the placenta and cord or peripheral blood when inconsistent or mosaic results are identified following a NIPT result with a normal newborn phenotype. Since the first discordant NIPT result, the 112 amniocenteses karyotyped by Akron Children's Hospital Cytogenetics Laboratory include 29 positive testing NIPT cases. These 29 cases (~26%) are attributable to a substantial proportion of positive NIPT screened cases referred to Maternal Fetal Medicine specialists. Cytogenetic analysis confirmed ~60% of these cases tested by NIPT. Our original results plus these new cases confirm (1) a normal NIPT test result followed by a 20 week anatomical ultrasound detected a false negative trisomy 18 NIPT result, (2) a substantial proportion of abnormal NIPT tests identified chromosomal mosaicism that may or may not be confined to the placenta, (3) follow up karyotyping should be ordered on the newborn placenta when the amniocyte karyotype does not confirm the NIPT reported abnormality in order to identify ongoing risk for mosaic symptoms, (4) considering our ~4,000 reported amniocenteses, karyotyping all high risk fetuses will define ~24% of chromosome abnormalities not currently screened by NIPT, and (5) ongoing counseling will facilitate these improved screening results. These Maternal Fetal Medicine referrals demonstrate that positive NIPT results identify abnormal karyotypic frequency as well as a substantial proportion of discordant circulating fetal DNA results.

2523F

Disruption of the MBD5 gene by a small deletion of the 2q23. 1 in a family with mental retardation and language impairment. *Y. G. Oh, J. O. Lee, B. H. Lee, E. J. Seo.* Medical Genetics Center, Asan Medical Center, Univ. of. Ulsan College of Medicine.

The 2q23. 1 deletion syndrome is a recently identified neurodevelopmental disorder characterized by developmental delay, seizures, intellectual disability, and autistic-like behavior. The deletions at the 2q23. 1 range in size from 38 kb to greater than 19 Mb. Methyl-CpG-binding domain 5 (MBD5) is regarded as a causative gene of this syndrome with haploinsufficiency. In this study, we describe a family with MBD5 disruption resulting from a 163 kb deletion. The proband was a 15-year old girl who presented with developmental delay, motor delay, severe mental retardation, language impairment, autistic-like symptom, short stature, normocephaly, and skeletal abnormalities including pes planus and clinodactyly. Conventional chromosomal analysis showed a normal karyotype. Chromosomal microarray using Agilent SurePrint G3 Human 180K revealed a 163 kb deletion at the 2q23. 1, which encompassed exons 1 to 6 of ORC4 and exons 1 to 2 of MBD5. The proband's father and younger sister were found to have the same deletion of the 2q23. 1 using chromosomal microarray. They presented with short stature, mild to moderate mental retardation, and language impairment, but less pronounced features than proband. Interestingly, seizures or craniofacial abnormalities in this family were unremarkable. This result suggested that the haploinsufficiency of MBD5 caused by complete deletion or gene disruption might be closely associated with mental retardation and language impairment among variable clinical manifestations of the 2q23. 1 deletion syndrome.

2524F

Clinical Characteristics of Chromosome Microarray Studies. *E. Purifoy[1], V. Goitia[1], M. Cua[1], C. Wang[2].* 1) Pediatrics, Driscoll Children's Hospital, Corpus Christi, TX; 2) Neurology, Driscoll Children's Hospital, Corpus Christi, TX.

Chromosome microarray study (CMS) is widely utilized in clinical settings. However, interpretation of results can be difficult due to the lack of clear clinical correlates. In order to better understand the clinical utility of CMS we conducted a comprehensive retrospective study and analyzed all CMS performed at Driscoll Children's Hospital (DCH). Patients aged 0-21 years who had a CMS performed between the years 2008-2014 were included. We constructed a searchable electronic database including detailed genotypes and clinical information. Phase I data analysis investigated positive rates in demographic subgroups, different clinical subspecialties, and different clinical diagnostic categories. All subjects were classified into the following diagnostic categories: autism, congenital anomalies, and others (including several other clinical diagnoses). Phase II data analysis focused on the positive results and the clinical characteristics of different genotypes. A total of 807 CMS were identified. The overall positive rate was 29%. There were 529 males (66%) and 278 females (34%). The largest age group was 1-5years (364 subjects). Neurologists ordered the largest number of tests. We found a 15% positive rate for autism, 27% for congenital anomalies and 27% for other diagnoses. Interestingly, there was a significant increase in the rate of positive test results when the subject had two of these diagnoses. The rate of positivity was 56% for subjects with both autism and congenital anomaly, 37% for autism and other, 43% for congenital anomaly and other. In our phase II data analysis we found that chromosomes 15, 16 and X were the most commonly affected chromosomes overall. The chromosomes with the most mutation sites in autism were 10, 15, and 16. For congenital anomalies the most commonly affected chromosomes were 7 and X. Chromosomes 1,2,16, and X had the most mutation sites for subjects with the other diagnoses. A gender comparison of the chromosome mutations sites in autism was performed and revealed that there was a significant difference in the distribution of mutation sites between males and females. Our finding that a patient with two clinical diagnoses significantly increases the likelihood of a positive test has not been reported previously.

2525F

High detection rate of clinically relevant genomic abnormalities in plasma cells enriched using magnetic-activated cell sorting from consecutive patients with multiple myeloma. *S. Chan[1], Z. Singh[2], M. R. Baer[3], A. Z. Badros[3], Y. Huang[4], Y. S. Zou[2].* 1) Department of Human Genetics and Genomic Medicine; 2) Department of Pathology; 3) Department of Medicine and Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, MD; 4) Department of Mathematics and Statistics, University of Maryland, Baltimore County, MD.

Multiple myeloma (MM) is a plasma cell malignancy. Each year approximately 24,050 cases of MM are diagnosed, with a mortality of 11,090 individuals. Chromosome and FISH studies are used to provide prognostic information useful for refining risk stratification and predicting therapeutic response. Chromosome studies are frequently limited by lack of proliferation of plasma cells in vitro, and FISH analysis is hampered by small plasma cell percentages in marrow samples. To alleviate this problem, we implemented a protocol of magnetic-activated cell sorting (MACs) to isolate CD138+ plasma cells followed by FISH study to detect MM-related genomic abnormalities (the MACs FISH method) in July 2013. Before that, FISH was performed on BM samples without plasma cell isolation (the conventional FISH method). 286 consecutive patients diagnosed with MM at the University of Maryland Greenebaum Cancer Center were analyzed by the conventional FISH method (from January 2012 to June 2013) and 180 patients were analyzed by the MACs FISH method (from July 2013 to January 2015). Although demographics and laboratory data of patients from January 2012 to June 2013 and from July 2013 to January 2015 were similar, the MACs FISHmethod identified significantly more abnormal cases (69. 5% vs. 28. 9% cases, P < 0. 01) than did the conventionalFISH method. The MACs FISH method yielded significantly higher detection rates of del(13q14/*RB1*), del(17p/*TP53*), 1q21/*CKS1B* gain, *CCND1*/*IGH*, *FGFR3*/*IGH*, and other *IGH* gene rearrangements (P<0. 01) than the conventional FISH method. The MACs FISH method was able to detect a significantly greater number of abnormal cytogenetic subtypes associated with MM (average 1. 5 vs. 0. 4 per case, P<0. 01) than the conventional FISH method. We conclude that plasma cell enrichment of BM specimens by the MACs FISH method significantly increasesFISHsensitivity for detecting clinically relevant genomic abnormalities in MM. These results, in combination with clinical data, can be of value in improving risk stratification and patient management.

2526F

Molecular cytogenetic analysis of recurrent chromosomal abnormalities in pancreatic cancer. *D. Shabsovich[1,2], L. Fonseka[2], K. Park[2], L. Yang[2], N. Rao[2], C. A. Tirado[2].* 1) MICROBIOLOGY, IMMUNOLOGY, AND MOLECULAR GENETICS, UCLA, SANTA MONICA, CA; 2) Pathology & Laboratory Medicine , UCLA, Los Angeles, CA 90024.

Pancreatic cancer is the fourth leading cause of cancer-related death worldwide, with an average 5-year survival rate of approximately 7 percent. However, due to extensive intratumor and intertumor cytogenetic heterogeneity, there are no cytogenetic assays utilized in the clinical management of the malignancy. Herein, we perform Giemsa banding and fluorescence in situ hybridization (FISH) analysis of 15 pancreatic cancer cell lines using probes specific to the following genomic loci : *ARID1A* (1p36), *TGFBR2* (3p22), *MLH1* (3p22), 7q31, centromere 8, *MYC* (8q24), *CDKN2A* (9p21), *TP53* (17p13), and *SMAD4* (18q21). By conventional cytogenetic analysis, the most common numerical aberrations involve chromosomes 20, 22, 7, 12, 13, 14, 18, X and Y. Additionally, structural aberrations most commonly involved chromsomes 1p, 3p, and 8p. FISH analysis revealed a range of abnormal signal patterns and clonal subpopulations bearing various numerical and structural abnormalities both observed and not observed by conventional cytogenetic analysis. In particular, the average copy numbers of the studied genes were found to be different from that of the ploidy of the overall specimen, suggesting a wide range of locus-specific cytogenetic aberrations present in the malignancy involving genes previously associated with pancreatic and other gastrointestinal cancers. To the best of our knowledge, this is the first study that used FISH to investigate *ARID1A*, *TGFBR2*, and *MLH1* abnormalities in pancreatic cancer. These findings provide insight into the chromosomal and molecular nature of the disease and suggest targets for future investigation. Ultimately, phenotypic correlation of chromosomal abnormalities in pancreatic cancer can provide a framework for the development of diagnostic and prognostic cytogenetic assays that can be utilized in the clinical management of the malignancy.

2527F

Application of fluorescence in situ hybridization (FISH) and immunohistochemistry (IHC) for detecting PTEN loss in diagnostic prostate cancer needle biopsies. *J. A. Squire[1,2,3], C. G. Picanco-Albuquerque[3], C. L. Morais[4], F. L. F. Carvalho[4], S. B. Peskoe[6], J. L. Hicks[4], O. Ludkovski[1], H. Fedor[4], E. Humphreys[7], M. Han[7], E. A. Platz[4,5,6], A. M. De Marzo[4,5,7], D. M. Berman[1,4,7], T. L. Lotan[4,5].* 1) Department of Pathology and Molecular Medicine, Queen's University, Kingston, ON, Canada; 2) Department of Pathology, Ribeirao Preto Medical School, University of Sao Paulo, Ribeirao Preto, Brazil; 3) Department of Genetics, Ribeirão Preto Medical School, University of Sao Paulo, Ribeirão Preto, Brazil; 4) Department of Pathology, Johns Hopkins University School of Medicine; 5) Department of Oncology, Johns Hopkins University School of Medicine; 6) Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health; 7) Department of Urology and the James Buchanan Brady Urological Institute, Johns Hopkins University School of Medicine.

Loss of the phosphatase and tensin homolog (*PTEN*) tumor suppressor gene by genomic deletion occurs in approximately 20-30% of prostate cancers. *PTEN* loss at radical prostatectomy has been shown to correlate with early biochemical recurrence, extracapsular and seminal vesicle invasion, castrate resistant disease, metastasis and prostate cancer-specific death. Differentiation of aggressive from indolent tumors remains a high priority for the appropriate management of prostate cancer and the avoidance of unnecessary treatments and side effects in patients with indolent disease. The prognostic value of PTEN loss has primarily been evaluated by either fluorescence in situ hybridization (FISH) or immunohistochemistry (IHC) using radical prostatectomy specimens. This case control cohort was designed to determine the effectiveness of PTEN loss by IHC in predicting the need to upgrade the Gleason score based on the pre-operative needle core biopsies. In this study a subset of 111 needle core samples were analyzed by *PTEN* FISH to determine the predictive value of FISH to upgrade the Gleason score. *PTEN* deletion were found by FISH in 26. 7% (12/45) of upgraded cases compared with 10. 6% (7/66) of controls (p =0. 04). Tumors with *PTEN* deletions were more likely to be upgraded at radical prostatectomy than those without a deletion, when adjusting for age (odds ratio = 3. 40 (1. 14–10. 11; P=0. 028)). Overall the IHC and FISH assays were highly concordant: of 93 evaluated biopsies scored with protein intact, *PTEN* FISH indicated 89 (95. 7%) had no deletion. Of 18 biopsies with PTEN protein loss 15 showed homozygous or hemizygous *PTEN* gene deletion by FISH. Only 4 biopsies of the 93 (4. 3%) with PTEN protein intact had a deletion by FISH. In these four biopsies regions of uncertainty were systematically studied at the cellular level by FISH to demonstrate that intratumoral variation of *PTEN* deletion could account for variation in immunoreactivity. Collectively, this study demonstrates the complementary advantages of both the IHC and FISH assays for analyzing PTEN loss in histologic sections from newly diagnosed prostate cancer patients.

2528F

Clinical features of *TWIST2* heterozygous carriers in a Mexican patient with a 2q37 deletion syndrome (Albright´s hereditary osteodystrophy-like syndrome). *D. E. Cervantes-Barragan[1], L. J. EDELMANN[4], M. J. GAYTAN-GARCIA[1], L. D. LOPEZ-VELAZQUEZ[2], S. RUIZ-PEREZ[3], I. NAZARENKO[4], B. H. LEE[4,5], J. I. NAVARRETE-MARTINEZ[1], R. J. DESNICK[4].* 1) DEPARTMENT OF GENETICS, HOSPITAL CENTRAL SUR DE ALTA ESPECIALIDAD PEMEX, MEXICO CITY, Mexico City, Mexico; 2) Department of Ophthalmology, Hospital Central Sur de Alta Especialidad PEMEX, Mexico city; 3) Department of Pediatrics, Hospital Regional Ciudad Madero, Tamaulipas, Mexico; 4) Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, New York, USA; 5) Department of Pediatrics, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea.

The 2q37 deletion syndrome or Albright´s hereditary osteodystrophy-like syndrome (OMIM # 600430) is a rare chromosomal disorder reported in more than 115 cases. The cardinal features of the phenotype include developmental delay, obesity, type E brachydactyly, autism and facial dysmorphism. At least 197 genes were mapped in the 2q37 locus, 11 were potentially candidate genes involved in the phenotype, for example *HDAC4, FARP2* and *TWIST2* genes. Homozygous *TWIST2* mutations have been identified in patients with Focal Facial Dermal Dysplasia type 3 or Setleis syndrome (OMIM #227260). The Setleis phenotype includes bitemporal "forceps like" marks, distichiasis, superior triple eye row eyelashes, absence or lower eyelashes and facial dysmorphism. Some carries (*TWIST2* heterozygotes) present with subtle facial dysmporhism and aberrant eyebrows and eyelashes conformation. The exact role of the*TWIST2* phenotype in the 2q37 deletion is poorly understood. We present a 2q37 deletion patient with the clinical features of *TWIST2*heterozygotes. A 7 year 5 month old male patient was referred for moderate developmental delay, scoliosis secondary to a lumbar hemivertebrae, type E brachydactyly, low height and facial dysmorphism (cupped ears and full lips with bow arrow of superior lip). He was the first child of the parents, conceived by *in vitro* fertilization with an unrelated sperm donor. Initial 550-band peripheral blood karyotype revealed a 46,XY,del(2)(q37), confirming an Albright´s osteodystrophy-like syndrome phenotype. High resolution oligonucleotide array-CGH are in progress. Intentional heterozygous onphysical exam, the patient had a right upper eyelashes triple row and distichasis with a V-shaped nasal tip. TWIST2 is a basic-helix-loop-helix transcription factor that is involved in regulation of dermal and bone development during mammalian embryogenesis. *TWIST2* has been proposed as a strong candidate gene for various skeletal malformations involved in 2q37 deletion, including, brachydactyly type E, short limbs, and abnormal vertebrae, but not craniofacial features. Ophthalmologic (eyelashes and eyebrows) and facial (nasal and jaw) features are present in few *TWIST2* heterozygotes. Additional to skeletal involvement, *TWIST2* would explain facial dysmorphysm in patients with 2q37 deletion. Ophthalmologic features of heterozygotesshould be considered in patients with 2q37 deletion.

2529F

Two recurring deletions of 18p with variable expressivity of the 18p minus syndrome. *K. Phillips, R. Burnside, I. Gadi, J. Tepperberg, V. Jaswaney, J. Schleede, A. Penton, H. Risheg, V. Potluri, B. Rush, R. Pasion, H. Taylor, H. Cabral, J. Shafer, L. Kline, M. Johansen, K. Riley, S. Molinari, S. Schwartz, P. Papenhausen.* Cytogenetics, LabCorp, CMBP, RTP, NC.

18p deletion syndrome occurs in 1 in 40,000 births but is common in infants with malformations. Previous studies have mapped clinical features including learning difficulties, ptosis, and short neck to proximal 18p deletions and clinical features such as growth delay, seizures, holoproscencephaly, dystonia, hypotonia, and round face to distal 18p deletions. Holoprosencephaly (HPE4) is associated with loss of function mutations in the *TGIF1* gene. Using SNP microarray, we have analyzed 19 cases of isolated 18p terminal deletions. These cases all reported with terminal deletions and did not include deletions mediated by ring formation, translocations, adjacent inversion duplications, or familial interstitial copy loss variants. Most of the patients were referred only for microarray without accompanying karyotype analysis, and only 2 of 19 patients had parental follow-up analyses done by our lab. Patient results were sorted into 2 groups: those with deletions of the entire p-arm (~15 Mb withnumerous OMIM genes: *USP14->SPIRE1* or *MC2R*) and those with a karyotypically cryptic distal deletion of ~5 Mb (proximal OMIM gene: *DLGAP1*). Patient-age at time of referral ranged from newborn to 62 years, and adult patients were present in both groups. Reasons for referral varied and included cleft lip/palate, holoprosencephaly, encephalopathy, delayed milestones and developmental delay, intellectual disability, FTT, hypotonia, multiple congenital anomalies, heart septal defects, minor anomalies, muscle disorder, and short stature. Break points of distal deletions occurred in the gap between *DLGAP1* telomerically and *EPB41L3* centromerically, and whole p arm deletion breakpoints consistently occurred near the centromere. Although clinical features in patients with 18p deletion syndrome are reported to be dependent upon the size of the deletion and the genes involved, in this study severity of the patient phenotype did not appear to correlate with size of the deletion or breakpoint, as the patients in both groups were phenotypically similar. These data suggest that a limited number of genes predict the phenotype and other genomic and environmental factors likely play an important role in 18p deletion syndrome.

2530F

Cryptic deletion adjacent to the breakpoints of the t(9;22)(q34;q11) translocation in patients with chronic myeloid leukemia. *C. Borjas[1,2], M. Dominguez[1,2], J. González[1,2].* 1) Universidad de Guadalajara, Guadalajara, Guadalajara, Mexico; 2) División de Genética CIBO Instituto Mexicano del Seguro Social.

Introduction Chronic Myeloid Leukemia (CML) is a malignant disorder that results from a reciprocal translocation t(9;22)(q34;q11. 2). Conventional cytogenetic analysis identifies the t(9;22) translocation in 90% of patients. There are factors that determine a poor response to treatment. One such factor could be the sub-microscopic deletion in the derivative 9, which mostly implies the 5'*ABL1*/3'*BCR* sequence. Its incidence is unknown in the Mexican population and its clinical impact. We investigated the frequency of cryptic deletion adjacent to the breakpoints of the t(9:22) translocation by fluorescence in situ hybridization (FISH) in CML patients and evaluate its impact on some clinical and laboratorial parameters. . **Material and methods**We conducted a cross-sectional, descriptive, observational study. We analyzed bone marrow samples as well as the clinical and laboratorial characteristics of the patients with CML at diagnosis. The analysis included GTG-banding and FISH with dual color, dual fusion probes (DC-DF). Statistical analysis was performed using the Student t test for independent samples and Fisher exact test. **Results**Twenty-four CML patients were analyzed, 12 men and 12 women. Mean age was 35. 3 years. The Sokal low risk was present in 38%. The study by GTG banding disclosed the standard t(9;22) in 16 patients (66. 6%), a complex translocation in 6 (25%), a patient with concealed t(9;22) translocation had the karyotype 47,XY,+8 (4. 2%) and a case where metaphases could not been obtained (4. 2%). All patients showed the *BCR/ABL1* fusion by FISH. The *ABL1* cryptic deletion was found in 4 patients (16. 6%). The patient 47,XY,+8 was diagnosed as positive for the BCR/ABL1 fusion and for the deletion of the der(9) sequences by DC-DF FISH. However, a complex arrangement that included a re-translocation *5'ABL1/3'BCR* on the derivative 22 was identified by tricolor-DF FISH. **Discussion** We found that the splenomegaly was significantly higher in men (mean = 13. 8) than females (mean = 7. 0). Also, leukocytosis values showed significant differences when grouping the patients by severity of anemia. No statistical significance between other clinical and laboratorial variables with the cryptic deletion of the *ABL1* sequence was found. **Conclusion** The frequency of cryptic deletion 5'*ABL/3'BCR* in our population is similar to that reported for other populations, and no association was found between the cryptic deletion of *ABL1* with the clinical phenotype of the disease at diagnosis.

2531F

Clinical and molecular characterization of an infant with a 4p16. 3p15. 33 deletion and Xp22. 33p22. 31 duplication. *M. E. S. Colovati[1], S. Bragagnolo[1], R. S. Guilherme[1], A. G. Dantas[1], M. Z. Souza[1], M. F. Soares[2], A. B. A. Perez[1], M. I. Melaragno[1].* 1) Genetics Division, Department of Morphology and Genetics, Universidade Federal de São Paulo, São Paulo, Brazil; 2) Department of Imaging Diagnosis, Universidade Federal de São Paulo, São Paulo, Brazil.

Wolf-Hirschhorn syndrome (WHS, OMIM #194190) is a contiguous gene syndrome with frequency estimated at 1 in 20,000 to 1/50,000 births. Approximately 50-60% individuals have a*de novo*4p16 pure deletion and 40-45% have an unbalanced translocation, *de novo*or inherited from a parent with a balancedrearrangement. We describe a 2 year-old female patient who presented Wolf-Hirschhorn syndrome features with apparent facial assimetry, high forehead, ears tags, left microtia, cleft palate, pectus carinatum, pré-sacral pit, pulmonary stenosis, atrial septal defect, and gastroesophageal reflux. She showed growth retardation and developmental delay, as well as, epibulbar dermoids in the right eye, a typical characteristic of the oculoauriculovertebral spectrum (OAVS). Karyotype was considered normal but high-resolution genomic array analysis showed a ~13 Mb 4p16. 3p15. 33 deletion associated with a ~9 Mb Xp22. 33p22. 31 duplication due to a maternal balanced translocation identified by FISH. The deletion in chromosome 4 includes the WHS critical region, as well as, *BAPX1* and *HMX1* genes that have been associated with OAVS. The patient's and mother's rearrangement breakpoints were mapped by custom genomic array, and sequence analysis of the junction fragments indicated no gene disruption. The mother's der(X) showed 10 additional nucleotides and the patient's and mother's der(4) showed 4 pb deletion. The mother's rearrangement could have been produced by Non Homologous End Joining mechanism (NHEJ), leavinga 'molecular scar' at the DNA junction. In the breakpoints of both chromosomes, repetitive sequences were found (segmental duplications and LTR66 in Xp and Charlie17a in 4p) that could be involved in the rearrangement formation. Thus, the atypical proband clinical phenotype could not be explained by gene disruption but could be due to the presence of sensitive dosage genes in the chromosome regions with 4p monosomy and funcional Xp disomy. This case supports the hypothesis that the 4p region is a candidate for OAVS. Our data revealed the first case of WHS and OAVS features in a patient with functional Xp disomy due to a maternal balanced translocation, a rare cytogenetic finding in females. Financial support: FAPESP.

2532F

A female teenager with sporadic Williams-Beuren syndrome and a familial inv(X)(p22q22). *A. Ramírez[1], H. Rivera[1,2], AI. Vásquez[2], P. Barros[2], L. Figuera[2], MG. Domínguez[2].* 1) Universidad de Guadalajara, Guadalajara, Jalisco, México; 2) División de Genética, CIBO, Instituto Mexicano del Seguro Social, Guadalajara, México.

Introduction: Williams-Beuren syndrome (WBS; OMIM 194050) is typically due to a hemizygous deletion of either 1. 55 or 1. 84 Mb (90% and 8% respectively) on chromosome 7q11. 23with an incidence of 1/7500 [1]. These deletions contain 25 or 28 genes and include the *ELN* gene [2]. The WBS phenotype consists of characteristic facial features, cardiac defects, structural brain abnormalities, and a specific cognitive profile [3]. Pericentric inversions of the X chromosome occur at a frequency of 1/28,000-1/30,000 individuals [4]. We report on a female patient with a 46,X,inv(X)(p22q22) karyotype and a clinical diagnosis of Williams-Beuren syndrome confirmed by FISH. **Clinical report:** This female patient was diagnosed with WBS at 14 years of age. She was born at term to non-consanguineous parents; her birth-weight was 2,700 g. The patient had a typical WBS facies with bitemporal narrowing, epicanthal folds, prominent cheekbones, small and spaced teeth, large mouth, thick lips, prominent eyebrows, bulbous nose, and micrognathia. She also presented short stature, delayed development, supravalvular aortic stenosis (SVAS), visuospatial impairment, and friendly personality. **Methodology:** Chromosomes were analyzed by GTG banding with resolution of 450 bands in 50 metaphases from lymphocyte cultures. FISH assays were carried out with the dual WBS (LSI ELN Spectrum Orange/LSI D7S486, D7S522 Spectrum Green) and Kallmann (LSI KAL Spectrum Orange/CEP X (DXZ1) Spectrum Green), Downers Grove, IL) probes. **Results:** G- banding in the patient and her mother revealed an inv(X)(p22q22), the father karyotype was normal. These breakpoints were confirmed with the Kallmann probe whose signal remained in place; the *ELN* probe revealed a concomitant 7q11. 23 deletion. **Discussion:** About 40 X-chromosome pericentric inversions have been reported including a single case with similar breakpoints ascertained in another Mexican family [5]; since both families came from the same region, they may be related. The co-occurrence of WBS with other alteration is rare; there is a report of a patient with *de novo* typical 7q11. 23 and 22q11. 2 deletions and a combined phenotype of WBS and DiGeorge syndromes [6]. Genetic counseling for carriers of an X-chromosome pericentric inversion should consider the risks of spontaneous abortions, infertility, and recombinant offspring [7]. The concurrence here reported appears to be a fortuitous one.

2533F

Bilateral aniridia due to de novo deletion in 11p. 1-p13 region may have resulted from HPV infction. *S. G. Cevik[1], M. O. Cevik[2].* 1) Ophtalmology, Bursa Sevket Yilmaz Training and Research Hospital, Bursa, Turkey; 2) Adiyaman University School of Medicine Department of Medical Genetics, Adiyaman, Turkey.

WAGR syndrome is an acronym forWilms tumor, aniridia, genitourinary anomalies, and mental retardation that are prevalent in most of the patients of this syndrome. We present a one year old babygirl with bilateralaniridia, an absence of the iris. She had loss ofvisual acuityand increased sensitivity to photophobia. She did not have cataracts, noglaucoma, but hadnystagmus-involuntary eye movements. Current status, she has no Wilms tumor. Her karyotype analysis revealed 46,XX karyotype. However her array CGH analysis revealed a heterozygous 6. 7 Mb deletion in chromosome 11p. 1-p13 region that arr [hg19] 11p14. 1-p13(29,209,743-35,961,997)x1. These region includes genes PAX6, WT1,DCDC1 genes. PAX6 gene is known to be responsible from aniridia. WT1 is helşd responsible from Wilms tumor. On the intact side of the WT1 gene region, we found that there is a homozygous G>C replacement at 267. nucleotide (3'UTR+267G>C). Her familal history included HPV infection in gestational period her mother. .

2534F

Unique Patterns of Allelic Imbalances Found by SNP-based Microarray: A Tale of Two Chimeras. *C. A. Marcou, U. Aypar, E. C. Thorland, N. L. Hoppman, H. M. Kearney.* Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN.

Genome-wide single nucleotide polymorphism (SNP)-based microarrays are powerful tools used to identify copy number imbalances in addition to providing genotype information for multiple polymorphic loci throughout the genome. The inclusion of SNP alleles allows for the recognition of copy neutral absence of heterozygosity (AOH) associated with recessive disease and uniparental disomy (UPD) as well as unique allelic patterns associated with mosaicism and, more rarely, chimerism. We report two cases of constitutional chimerism detected by recognition of unique patterns of SNP allele imbalance. The first case involves 2-year-old dizygotic male/female twin siblings with previously unrecognized twin-twin transfusion. The second case demonstrates diploid/triploid mosaicism (mixoploidy) in a 2-week-old male referred for dwarfism, physical abnormalities, cardiac defects, misshapen ears, and polydactyly. In both cases, characteristic SNP allele patterns allowed for the identification of two genetically unique cell lines, consistent with chimerism, and additional cytogenetic testing characterized the nature of the chimerism. When encountering mixed gender chimerism, it is important to distinguish chimerism limited to the hematopoietic lineage from that with more extensive somatic involvement. The latter may be associated with disorders of sexual development and other abnormal features, where the former may be clinically benign. The clinical consequences, if any, of chimerism/mixoploidy will rely on the percent and tissue distribution of the two independent cell lines. It is likely that chimerism is more common than previously recognized, and will be diagnosed more commonly through routine use of SNP-based microarrays. Therefore, recognition of unique SNP allelic patterns can lead to increased diagnostic utility of SNP-based microarrays in the constitutional setting.

2535F

The importance of intragenic duplications within VPS13B in the causation of Cohen syndrome. *JL. Smith, SW. Cheung.* Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Mutations in the VPS13B gene are associated with Cohen syndrome, an autosomal recessive disorder characterized by psychomotor disability, microcephaly, facial dysmorphism, childhood hypotonia, joint laxity and neutropenia. Types of mutations described in various ethnic backgrounds include nonsense, frameshift, splice site and missense mutations as well as small deletions. MLPA has been used to detect intragenic duplications in a few patients. We report a pair of siblings with homozygous duplications in VPS13B each detected by a different chromosomal microarray platform. A custom 400k Agilent array revealed an 81 kb intragenic gain spanning from intron 3 to 13 in one sibling; the log2 ratios indicated the presence of 4 copies of the region. The sibling's array was performed on the Affymetrix Cytoscan HD platform which showed a similar sized intragenic gain of 89 kb involving the same exons of VPS13B and a copy number state of 4. These intragenic duplications are predicted to disrupt the gene. The Cytoscan array also showed that the gain was in a 6. 7 Mb contiguous region of AOH; review of the Agilent array SNP data indicated the sibling's gain was in a small region of AOH. Agilent arrays were performed on both parents and both were found to carry a heterozygous duplication within VPS13B. Consanguinity was denied; however, both parents are from a small town in Mexico, hence a common ancestor is possible. Clinical assessment and hematological testing of both children are consistent with Cohen syndrome. Review of our data base of over 40,000 Agilent arrays did not detect any other cases of homozygous intragenic gains in VPS13B; however, it did detect four gains judged to be heterozygous. The literature indicates that some patients who fit the clinical description of Cohen syndrome have not been confirmed by laboratory analysis of VPS13B to have a homozygous mutation. These data demonstrate that microarray analysis may be another alternative to consider for full assessment of the mutational status of the VPS13B gene in patients who fit the clinical description of Cohen syndrome.

2536F

Microdeletion 15q13. 2-q13. 3 syndrome: case report of a patient with facial dysmorphisms, developmental delay, autism spectrum disorder and epilepsy. *E. D. F. Carvalho[1, 2,3], K. M. Carvalho[2, 3], M. D. F. Carvalho[1, 2, 3].* 1) Medical Genetics, Unichristus, Fortaleza, Brazil; 2) Medical Genetics, Genpharma, Fortaleza, Brazil; 3) Renorbio, Universidade Estadual do Ceará, Fortaleza, Brazil.

Microdeletions and chromosomal microduplications are increasingly being diagnosed in patients with developmental delay and intellectual disability, and the majority of cases were revealed by Comparative Genomic Hybridization (CGH array). The microdeletion 15q13. 2-q13. 3is not a syndrome with well defined clinical features and with an easy recognition. Their carriers appear to be more prone to a range of clinical manifestations that vary widely and may include learning disabilities, seizures, autism, neuropsychiatric disorders and congenital heart disease. In 75% of cases, the microdeletion is inherited from one parent. The microdeletion can also be present in unaffected individuals, featuring an incomplete penetrance, therefore, the microdeletion should be research in the parents for an appropriate genetic counseling. We report a case of a female child with facial dysmorphisms, developmental delay, autism spectrum disorder and epilepsy who was diagnosed with microdeletion 15q13. 2-q13. 3 syndrome. She is the second child of non consanguineous parents. The prenatal period was marked by mother hypertension that progressed to pre-eclampsia from the fifth month. The antenatal ultrasound performed was normal. She was born at 31 weeks of gestation by Caesarean section and had a birth weight of 1,429 g, a birth length of 39 cm and an occipital frontal circumference of 30cm . Her Apgar score was 7 and 9 at 1 and 5 min, respectively. At 2 years old, and her physical examination disclosed: developmental delay, short stature, facial discrete dysmorphisms (long palpebral fissures, open mouth, anteverted nostrils). The clinical follow-up was marked by epilepsy, autism spectrum disorder and behavioral changes at 3 years age. Echocardiogram, abdominal USG and transfontanellar and G-banded karyotype were normal. Skull MRI revealed discrete ectasia not hypertensive of the lateral ventricles and Esophageal pH monitoring showed pathologic distal gastroesophageal reflux. The CGH array revealed a microdeletion of 1. 5 MB size at 15q13. 2-q13. 3 involving seven genes: TRPM1, KLF13, OTUD7A, CHRNA7, TRPM1. The CHG array of the parents was performed, but the result is not available yet. Our patient shows clinical signs already reported in this microdeletion syndrome, but with additional clinical features not previously reported as facial dysmorphisms and important gastroesophageal reflux which can lead us to think that the phenotype has not yet been fully clarified.

2537F

qPCR confirmation and follow-up testing of copy number variants identified by microarray analysis. *J. Schleede[1], T. Tepperberg[1], S. Schwartz[1], S. Schepis[1], B. Williford[1], B. Burnside[1], V. Jaswaney[1], I. Gadi[1], K. Phillips[1], H. Risheg[2], A. Penton[1], P. Papenhausen[1].* 1) Laboratory Corporation of America® Holdings, Department of Cytogenetics, Research Triangle Park, NC,; 2) Laboratory Corporation of America® Holdings, Houston, TX, 3Dynacare Laboratories, LabCorp, Seattle, WA.

Microarray technology has revolutionized the identification and classification of small copy number variants (CNVs). Fluorescence in situ hybridization (FISH) using BAC probes is widely accepted as the method of choice for confirmation and familial follow-up of CNVs identified by microarray analysis. However, FISH is time consuming, costly, and is not always applicable for the detection of very small CNVs. Real-time quantitative PCR (qPCR) analysis allows for the rapid detection of CNVs at a fraction of the cost of traditional FISH analysis. We have recently added qPCR analysis to our comprehensive microarray follow-up protocol using the Affymetrx® VeriQuest SYBR Green qPCR Master Mix. CNV specific primer pairs are designed using the online resource Primer3Plus while specificity is assayed using the UCSC In-Silico PCR tool. Three normal controls are run in triplicate in conjunction with all samples to ensure both precision and accuracy. qPCR analysis gives reliable and reproducible CNV detection from all sample types including blood, buccal, amnio, CVS, and other tissues. All abnormal prenatal microarrays and CNVs smaller than 50 kilobases can now be quickly and definitively confirmed using this technology. Furthermore, qPCR analysis has become our first round test for all parental follow-up analyses of variants of unknown significance. In the case of *de novo* CNVs we continue to offer parental follow-up FISH analysis to rule out the possibility of a balanced rearrangement with significant recurrence risk. Although one caveat of qPCR testing has been the need to very carefully select primer pairs for complex CNVs that involve small clusters of repeat elements, a clear benefit is the elucidation of chromosome structure associated with genomic gains. We have found that single copy gain by array shows as multi-copy gain by qPCR. This may be the result of various repeat elements in the genome (e. g. small segmental duplications, LINES, SINES, Alu repeats, etc).

2538F

Atypical Angelman Syndrome Caused by a Familial Imprinting Center Deletion. *S. S. Costa, A. C. Krepischi, C. Rosenbeg.* Genetics and Evolutionary Biology Dpt. , University of Sao Paulo, Sao Paulo, SP, Brazil.

Angelman syndrome (AS) is caused by different genetic abnormalities involving the chromosome 15q11-13 region, which is subject to genomic imprinting. Imprinted gene expression in the region is regulated by a bipartite imprinting centre (IC). The AS-IC lies around 35 kb upstream of *SNRPN* exon 1 and has been mapped to a region of just 0. 88 kb. Its prevalence is 1-2 per 10,000 individuals. Most patients are affected by AS due to one of four mechanisms: (1) large interstitial deletion of 15q11-q13; (2) paternal uniparental disomy (UPD) of chromosome 15; (3) imprinting defect (ID); or (4) mutation in the E3 ubiquitin protein ligase gene (*UBE3A*). In the few remaining patients, the mechanism is unidentified. Major characteristics of AS include developmental delay, severe mental retardation with lack of speech, movement ataxia, hyperactivity, seizures, aggressive behavior, and excessive inappropriate laughter. At clinical examination, AS was not considered as a diagnostic hypothesis for two siblings who presented only absent speech and intellectual disability. Ataxia and epilepsy were absent, and EEGs reveal no abnormality and, therefore, the phenotype was not characteristic of Angelman syndrome. Array-CGH (180K - Oxford Gene Technology) disclosed in both patients a deletion of 53 Kb encompassing exons U1 to U6. 5 of the *SNRPN* gene. Further, array-CGH showed that the phenotypically normal mother is a carrier, while the normal daughter did not carry the *SNRPN* deletion. DNA methylation of the *SNURF-SNRPN* exon 1/promoter was tested and the patients showed AS methylation pattern. Our results point to the importance of considering mild and atypical Angelman syndrome phenotypes within the spectrum of intellectual handicap, particularly for clarifying the genetic risk to other family members. Financial support: FAPESP/CAPES.

2539W

Mutations in *KIAA0586* cause lethal ciliopathies ranging from hydrolethalus phenotype to short-rib polydactyly syndrome. *C. Alby[1,2], K. Piquand[1], C. Huber[3], A. Megarbané[4], A. Ichkou[1,2], M. Legendre[5], F. Pelluard[6], F. Encha-Razavi[1,2], G. Abi Tayeh[7], B. Bessieres[2], S. El Chehadeh-Djebbar[8], N. Laurent[8], L. Faivre[8], L. Sztriha[9], M. Zombor[9], H. Szabó[9], M. Failler[10], M. Garfa-Traore[11], C. Bole[12], P. Nitschké[13], M. Nizon[2,3], N. Elkhartoufi[1,2], F. Clerget-Darpoux[1], A. Munnich[1,2], S. Lyonnet[1,2], M. Vekemans[1,2], V. Cormier-Daire[2,3], T. Attié-Bitach[1,2], S. Thomas[1].* 1) INSERM U1163 - Institut Imagine, PARIS, France; 2) Département de Génétique, Hôpital Necker-Enfants Malades, Assistance Publique Hôpitaux de Paris (AP-HP), 75015 Paris, France; 3) INSERM U1163, Laboratory of Molecular and Physiopathological bases of osteochondrodysplasia, 75015 Paris, France; Paris Descartes - Sorbonne Paris Cité University, Imagine Institute, 75015 Paris, France; 4) Medical Genetics Unit, Saint Joseph University, Rue de Damas, B. P. 17-5208 - Mar Mikhaël Beyrouth 1104 2020, Lebanon; 5) Department of Genetics, Poitiers University Hospital, 2 Rue de la Milàtrie, 86021 Poitiers, France; 6) Unité de Pathologie Fœtoplacentaire, Centre Hospitalier Universitaire, Groupe hospitalier Pellegrin, Place Amélie Raba-Léon, 33 076 Bordeaux CEDEX, France; 7) Service de Gynécologie Obstétrique, Hôtel-Dieu de France, Beyrouth BP: 166830, Lebanon; 8) Génétique et Anomalies du Développement, EA4271, Université de Bourgogne, 21000 Dijon, France; 9) Department of Paediatrics, Faculty of Medicine, University of Szeged, Szeged, Dugonics tér 13, 6720, Hungary; 10) INSERM U1163, Laboratory of Inherited Kidney Diseases, Imagine Institute, 75015 Paris, France; Paris Descartes - Sorbonne Paris Cité University, Imagine Institute, 75015 Paris, France; 11) Paris Descartes-Sorbonne Paris Cité University, Imagine Institute, Cell Imaging platform, 75015 Paris, France; 12) Paris Descartes-Sorbonne Paris Cité University, Imagine Institute, Genomic Core Facility, 75015 Paris, France; 13) Paris-Descartes Sorbonne Paris-Cité University, Bioinformatics Core Facility, 75015 Paris, France.

KIAA0586, the human ortholog of chicken *Talpid3*, encodes a centrosomal protein that is essential for primary ciliogenesis and its disruption in animal models causes defects attributed to abnormal hedgehog signaling including polydactyly and abnormal dorsoventral patterning of the neural tube. Here we report homozygous mutations of *KIAA0586* in 4 families with lethal ciliopathies ranging from a hydrolethalus phenotype to short-rib polydactyly (SRP). Consistent with a role of Talpid3 in primary cilia biogenesis, we show defective ciliogenesis in cells derived from affected individuals as well as abnormal response to SHH signaling activation. Whereas centriolar maturation seems unaffected in mutant cells, we observed an abnormal extended pattern of CEP290, a centriolar satellite protein previously identified as a ciliopathy protein. Our data show the crucial role of KIAA0586 in human primary ciliogenesis and subsequent abnormal Hedgehog signaling through abnormal GLI3 processing. Our results thus establish that *KIAA0586* mutations cause lethal ciliopathies.

2540W

Regulation of exocrine and endocrine pancreas by ciliopathy genes. *T. L. Hostelley, S. Lodh, N. A. Zaghloul.* Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD.

Ciliopathies are rare autosomal recessive disorders characterized by dysfunction in genes associated with the primary cilium. Two ciliopathies, Alström Syndrome (Alms) and Bardet-Biedl Syndrome (BBS), are distinct in the highly prevalent rates of obesity. Interestingly, however, they have strikingly different rates of diabetes. The rate is much higher among Alms patients relative to BBS patients. To gain insight into potential genetic mechanisms for this discrepancy we carried out whole transcriptome sequencing on zebrafish depleted of either Alms1 or Bbs1 expression. We identified gene expression changes between these disorders and found 8 genes that were differentially expressed in opposite directions between the two models. Among these, several exocrine pancreas digestive enzymes were found to be downregulated in Alms and/or upregulated in BBS. These changes in gene expression were validated by qRT-PCR. Based on these observations and our previous identification of discrepant β-cell mass between the two models, we hypothesized that the expression of exocrine pancreas enzymes may have an effect on endocrine pancreas function. This might suggest that the development of diabetes in Alms may be due to impaired endocrine function in the pancreas due to the reduced expression of exocrine pancreas enzymes, which is not seen in BBS as a result of increased expression of these enzymes. To test this, we overexpressed each enzyme in transgenic zebrafish embryos, in which β-cell mass can be visualized by fluorescent marker. We also overexpressed these enzymes in Alms1-depleted embryos to determine if overexpression of the exocrine pancreas enzymes could rescue the loss of β-cells observed in those animals. Overexpression of the exocrine pancreas enzymes significantly increased the area of the β-cell mass as well as the number of β-cells, suggesting that these enzymes play a role in regulating β-cell proliferation and endocrine pancreas function. Taken together, these findings suggest that global gene expression analysis can inform mechanisms of disease in rare genetic disorders as well as more common phenotypes associated with them.

2541W

Uncovering the variation in phenotypes arising from mutations causing ciliopathic disorders. *J. P. Whalley, I. Gut, Ciliopathies Exome Sequencing Initiative.* CNAG, Barcelona, Spain.

Ciliopathies encompass rare Mendelian diseases that affect cellular cilia. Interesting, several of the genes found containing causative mutations that result in ciliopathies, have been linked to syndromes affecting different tissues. Notably the gene *CEP290* and; Meckel syndrome type 4 (MKS4 [MIM #611134]), which primarily affects the kidneys, Leber's congenital amaurosis 10 (LCA10 [MIM #611755]), a severe retinal dystrophy and Joubert syndrome 5 (JBTS5 [MIM #610188]), which in the majority of cases, affects the brain. Why a mutation in this gene can lead to differing phenotypes is an interesting area of research. To work on this problem we have been granted access to the 1058 whole exome sequenced samples of the Ciliopathies Exome Sequencing Initiative. We also make full use of databases covering protein interactions and gene expression. We have uncovered two possible explanations on the discrepancy in the phenotypes expressed for mutations in *CEP290*. Firstly this may be linked to the position of the mutation. Joubert syndrome 5 mutations tend to cluster in the second half of the gene, while Meckel syndrome type 4 and Leber's congenital amaurosis 10 mutations are found uniformly across the gene. Enticingly this may be linked to a suspected isoform of the CEP290 protein, which does not include the first 940 amino acids. Another explanation may come from working with genomic sequencing and expression data. *CEP290* interacts with other genes, which may have an effect on the phenotypes observed. A good candidate gene is *PCNT*, which has expression in kidney tissue, but no detectable expression in brain tissue (in contrast to *CEP290* which has high expression in the tissue of the brain and the kidney). Great progress in rare disease research has been made in identifying causative mutations. This research is part of the next step in going further and trying to find out why seemingly very similar causative mutations have such a difference in the type and the intensity of phenotypes of rare diseases observed.

2542W

Functional genome-wide siRNA screen identifies *KIAA0586* as mutated in Joubert syndrome. *S. Roosing[1], M. Hofree[2], S. Kim[1,3], E. Scott[1], B. Copeland[1], M. Romani[4], J. L. Sylhavy[1], R. O. Rosti[1], J. Schroth[1], P. Aza-Blac[5], S. Heynen-Genel[5], T. Ideker[2], B. D. Dynlacht[3], J. E. Lee[6], E. M. Valente[4,7], J. Kim[8], J. G. Gleeson[1], Joubert Syndrome consortium.* 1) Laborator for Pediatric Brain Disease, The Rockefeller University, New York, NY; 2) Department of Computer Science and Engineering and Department of Medicine, University of California, San Diego, USA; 3) Department of Pathology and Cancer Institute, Smilow Research Center, New York University School of Medicine, New York, USA; 4) Mendel Institute, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Italy; 5) High Content Screening Systems, Burnham Institute, La Jolla, USA; 6) Samsung Genome Institute (SGI), Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Seoul, South Korea; 7) Section of Neurosciences, Department of Medicine and Surgery, University of Salerno, Salerno, Italy; 8) Korea Advanced Institute of Science and Technology (KAIST), School of Medical Science and Engineering, Daejeon, South Korea.

Defective primary ciliogenesis or cilium stability forms the basis of human ciliopathies, including Joubert syndrome (JS), with defective cerebellar vermis development. We performed a high-content genome-wide siRNA screen to identify genes regulating ciliogenesis as candidates for JS. We analyzed results with a supervised learning approach, using SYSCILIA gold standard, Cildb3. 0, a centriole siRNA screen and the GTex project, identifying 591 likely candidates. Intersection of this data with whole exome results from 145 individuals with unexplained JS identified six families with predominantly compound heterozygous mutations in *KIAA0586*. A c. 428del base deletion in 0. 1% of the general population was found in *trans* with a second mutation in an additional set of 9 of 163 unexplained JS patients. *KIAA0586* is an orthologue of chick *Talpid3*, required for ciliogenesis and sonic hedgehog signaling. Our results uncover a relatively high frequency cause for JS and contribute a list of candidates for future gene discoveries in ciliopathies.

2543W

A role for BBS4 in regulation of BDNF/TrkB signaling and hypothalamic neuronal specification. *C. Leitch, S. Lodh, N. Zaghloul.* Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland, Baltimore, Baltimore, MD.

Bardet-Biedl Syndrome (BBS) is an autosomal recessive ciliopathy characterized by highly penetrant obesity that is likely a result of defects in centrally mediated regulation of food intake: hypothalamic regulation of satiety. The extent to which regulation of hypothalamic neuronal signaling pathways are dependent on proper BBS protein function or ciliary regulation are unknown. To examine this possibility, we investigated a potential relationship between the BBS protein, BBS4, and a major neuronal signaling pathway involved in regulation of food intake, brain derived neurotrophic factor (BDNF) signaling. We investigated the role of BBS4 in modulation of BDNF signaling through its receptor, TRKB. We found that BBS4 is necessary for proper activation of the TRKB receptor by BDNF and that localization of TRKB to the axoneme of the cilium may play a role. Examination of the relevance of these findings to embryonic production of neuronal populations in zebrafish revealed an important role for BBS4 in modulating the balance between neuronal populations necessary for hunger and satiety, including hunger-inducing NPY/AgRP neurons. These effects could be ameliorated by addition of exogenous BDNF, potentially implicating the pathway in the production of these neurons. Additionally, reduction of BBS4 resulted in increased NPY expression levels, which was also ameliorated by addition of exogenous BDNF. Our findings suggest an important role for BBS4 and ciliary localization in transduction of BDNF signaling through TRKB. These data also suggest a potential role for BBS4 in mediation of the BDNF signaling pathway and in hypothalamic neurogenesis during development.

2544W

The role of the BBSome in visual function during eye development and adulthood. *Y. Hsu, J. E. Garrison, D. Y. Nishimura, K. E. Bugge, C. C. Searby, V. C. Sheffield.* University of Iowa , Iowa city , IA.

The outer segments of photoreceptor cells are specialized sensory **cilia**, and share features with other primary and sensory cilia. The BBSome, a protein complex consisting of BBS1, 2, 4, 5, 7, 8, 9 and 18, is required for the proper localization of proteins important for visual function such as rhodopsin. Human patients with Bardet-Biedl syndrome (BBS) and knockout mouse models of core BBSome components have retinal degeneration. However, it is not clear whether the BBSome plays a developmental role in the formation of photoreceptors in addition to maintaining normal functions of photoreceptors throughout life. Using a mouse model that enables the tamoxifen inducible deletion of BBS8, a core BBSome protein, we test whether the loss of the BBSome in adulthood, after the completion of eye development, disrupts visual function. We found that the inducible deletion of *Bbs8* in adult mice causes significant deterioration of visual functions measured by electroretinogram. Furthermore, we found that the deletion of *Bbs8* in juvenile mice at specific time points of retinal maturation causes rapid vision deterioration comparable to congenital *Bbs8* knockout mice. The data suggest that the BBSome plays a role in photoreceptor development as well as the continual function of photoreceptors after eye development is complete. Finally, using a hypomorphic mouse model where a gene trap is inserted between exon 3 and 4 of *Bbs8*, we investigated the functional relationship between vision and the expression levels of *Bbs8*. We show that visual function is sensitive to the gene expression levels of Bbs8, and that homozygous gene trap mice are able to retain partial visual function due to a low level of *Bbs8* expression. We show that the BBSome is required throughout life for the function and survival of photoreceptors. The deletion of BBSome components during the late stage of eye development causes accelerated vision deterioration compared to the deletion of BBSome components during adulthood, which causes a slower but significant rate of vision deterioration.

2545W

Transcriptional and translational dynamics of cell fate specification in pluripotent stem cells define the unique biology of individual human genomes. *A. Jaishankar, G. Stein-O'Brien, SK. Kim, S. Seo, E. Fertig, J. Shin, J. Chenoweth, D. Hoeppner, C. Colantuoni, R. McKay.* Lieber Institute for brain development, Baltimore, MD.

Human pluripotent stem cells have been proposed as a powerful tool for drug development as well as to define cellular deficits associated with genetic risk for disease. However, these applications require an understanding of factors that control cell fate specificationas well as the ability to consistently distinguish the cellular behavior of individual genomes. Novel bioinformatic analysis ofRNA-seq data obtained from multiple human pluripotent stem cell lines in conditions that support self-renewal or differentiation to mesendodermal and neuroectodermal fates, reveals lineage bias and distinct transcriptional dynamics in individual lines that accurately predicts their differentiation potential. We use high content imaging to map these spatial and temporal changes during lineage specification as human pluripotent stem cells rapidly organize into zones biased to form either the brain or the body. The key signaling pathways regulating stem cells continue to have a major role in brain development and dysregulation of these pathways contribute to risk for neurodevelopmental disorders. The AKT/mTOR signaling pathway regulates pluripotency. We have shown that this occurs in specific domains biased towards different regional states that form the organs and tissues of the new individual. A pharmacological approach will be presented to define how differences in AKT/mTOR signaling regulate the spatial organization of this early epithelium and the subsequent epigenetic trajectories these cells traverse. Applying these concepts, tools and novel data analysis opens the door to interrogate genetic and pharmacology diversity of human brain development from a powerful novel perspective.

2546W

Atypical cadherin FAT1 is essential for optic fissure closure in mouse and zebrafish. *A. George[1], S. BHARTI[1], R. SHARMA[2], F. ONOJAFE[1], S. DUTTA[1], H. McNEILL[3], K. BHARTI[2], B. P. BROOKS[1].* 1) Ophthalmic Genetics and Visual Function Branch, National Eye Institute/NIH, BETHESDA, MD; 2) Section on Epithelial and Retinal Physiology and Disease, National Eye Institute/NIH, Bethesda, MD; 3) Mt. Sinai Hospital, Department of Molecular Genetics, University of Toronto, Canada.

Expression profiling using microarray during mouse optic fissure closure (OFC) revealed that *Fat4* is dynamically expressed at the closing edges of the optic fissure in the developing mouse eye, suggesting that it might have a potential role in the pathogenesis of a potentially-blinding ocular malformation, coloboma. Upon histological analysis of *Fat4*-/- mice, no abnormality was observed in the OFC. On the contrary, a closely related family member *Fat1*-/- consistently exhibited OFC defects. Expression of other members of the FAT family i. e. *FAT2* and *FAT3* was not observed in the developing mouse eye. Expression patterns for *Pax2*, *Pax6*, *Mitf* and *Chx10* were similar between WT and *Fat1*-/- mice, suggesting the lack of early patterning defect. The phospho Histone-3 positive cells were significantly increased in the *Fat1*-/- mice compared to WT but the total number of cells in the optic cup remained equal. Activated Caspase-3 levels were relatively unchanged. As compared to WT, increase in cytoplasmic β-catenin staining was observed in the *Fat1*-/- mice retinal pigmented epithelium flat-mounts. To further delineate the possible mechanism we shifted our focus to zebrafish. Consistent with the findings in mice, morpholino mediated knock-down of *fat1* resulted in coloboma whereas it was not observed with *fat4* knock-down zebrafish embryos. The abnormal phenotypes, including cystic pronephros and brain defects were consistent with previous reports and suggests a possible role for Hippo signaling. Defects in Hippo signaling results in nuclear translocation of *yap1* resulting in growth defects. Concomitant knock-down of *yap1* together with *fat1* resulted in partial rescue of the OFC defects. Interestingly, over-expression by mRNA injection and morpholino mediated knock-down of *yap1* resulted on OFC defects. Mutations in *YAP1* and deletion of *SCRIBBL* gene, both of which play an important role in Hippo signaling, have already been reported to cause coloboma in humans, though the underlying mechanism remains obscure. .

2547W

The *Gfi-Cre* knock-in mouse as a tool for inner ear gene expression analysis. *M. Matern[1], B. Milon[1], R. Elkon[2], Y. Song[1], R. Hertzano[1].* 1) University of Maryland, Baltimore, Baltimore, MD; 2) Tel Aviv University, Tel Aviv, Israel.

Hearing depends on the proper functioning of hair cells, which are specialized neuroepithelial cells present in the inner ear. In adult mammals, cochlear hair cells are post-mitotic, meaning that once they are damaged they cannot be regenerated. Therefore, most forms of traumatic hearing loss from noise or ototoxic drug exposure result in a permanent loss of hair cells. Additionally, congenital non-syndromic sensorineural hearing loss can result from mutations in over 150 genes, many of which are expressed in hair cells. For years, the mouse model organism has proved to be an invaluable tool for understanding the genetics of hearing loss. However, molecular analyses of hair cell specific changes that occur in response to noise, drugs or genetic manipulations have been limited by the heterogeneous nature of the auditory sensory epithelium in which the hair cells comprise less than 2% of cells. More recently, RiboTag mice and transgenic mice expressing fluorescent markers have made tissue and cell type-specific analyses of gene expression possible. Specifically, the *Gfi1-Cre* knock-in mouse is commonly used for hair cell-specific recombination to analyze conditional knockout mice in the auditory field. Here we performed a comprehensive analysis of these mice to determine the specificity and utility of the *Gfi1-Cre* driver in analyzing hair cell-specific gene expression in the newborn inner ear. Specifically, we crossed *Gfi1-Cre* mice with both Ai14 and RiboTag mice to perform hair cell transcriptome and translatome analyses. Our results show that *Gfi1-Cre* mice produce a broader pattern of gene recombination than previously described. This is likely due to a more extensive expression pattern of the *Gfi1* transcription factor in the developing mouse otocyst. We further compare the inner ear transcriptomes of both wild type and heterozygous *Gfi1-cre* mice to determine whether the omission of one copy of *Gfi1* affects overall gene expression. Finally, we compare the hair cell translatome as obtained using the *Gfi1-Cre*/RiboTag mice with transcriptomes of sorted hair cells and critically discuss our results. This study is the first to evaluate the utility of the *Gfi1-Cre* mice as a tool for hair cell-specific genomic analysis in the inner ear. While our results show that the expression of *Gfi1-Cre* is not specific to hair cells as previously described, these mice are still a valuable tool for inner ear research.

2548W

Molecular genetic characterization of TGF-β mechanism in embryonic rescue for the secondary palate with overexpression of Smad2 using genetic models. *H. A. Al Omer[1, 3], G. E-H. Gawish[2, 3].* 1) Prince Abdulrahman Advanced Dental Institute (PAADI), Riyadh, Saudi Arabia; 2) Medical Genetics, College of Medicine, Al Imama Muhammad Ibn Saud Unversity; 3) Oral Biological Medical Science, University of British Columbia, Canada.

Cleft palate is a common birth defect in the human population. In analyses of non-syndromic cleft palate, a linkage to TGF-β3 has been shown. Signalling of TGF-β is mediated in the cell through Smad2. The aim of this study was to understand the mechanism of palatal fusion in the rescue mice compared to wild-type. The heads of embryos at age (E14. 5) of rescued, Smad2 overexpression and wild-type models were embedded in paraffin after genotyping and fixation. Serial 7 micron sections were studied for detection of apoptosis and epithelial mesenchymal transition using immunohistochemistry. Images were captured with confocal laser microscope. Activation of Smad2 was studied with phospho-Smad2 antibody, and the level of Smad2 in each embryo normalized with immunoblotting. We observed that TGF-β3 null mice developed a secondary palatal cleft while the TBF-β3 null mice that had also inherited the K14-Smad2 gene had fusion of the secondary palate. The effect of the K14-Smad2 expression was analyzed in the medial edge epithelium of the rescue mice; the MEE had a much higher ratio of cells with cleaved caspase, a marker of apoptosis, than in the control fused palates. The increase in apoptosis was correlated with increased p-Smad2 in the same cells while p-Smad2 in the control mice with normal palatal fusion was not associated with high levels of apoptotic MEE. We concluded that Smad2 overexpression might rescue the cleft in the secondary palate of mice by increasing apoptosis of epithelial cells in the middle seam. Thus the mechanism of rescue is not identical to the events that normally occur during palatal fusion.

2549W

Modeling distal hereditary motor neuropathies in zebrafish. *A. Kondo, B. Mitchell, L. L. Smith, A. H. Beggs, V. A. Gupta.* Genetics/Genomics, Boston Children's Hospital, Boston, MA.

The distal hereditary motor neuropathies (dHMN) are a heterogeneous group of diseases that are characterized by degeneration, loss of lower motor neurons, and skeletal muscle atrophy. Some forms of dHMN may have sensory abnormalities and overlap with axonal forms of Charcot-Marie-Tooth disease (CMT2). dHMNs are classified in to several forms based on mode of inheritance and phenotype (dHMN I-dHMNVII and X-linked dHMN). DHMN-VI or spinal muscular atrophy with respiratory distress type 1 (SMARD1) is a severe, autosomal-recessive infantile neuropathy. The most prominent symptoms are diaphragmatic paralysis and predominant involvement of the upper limbs and distal muscles. SMARD1 is caused by mutations in the gene immunoglobulin μ binding protein 2 (*IGHMBP2*). IGHMBP2 is an ATP dependent helicase that unwinds RNA and DNA duplexes and finally regulates the protein translation. Recent studies have identified *IGHMBP2* mutations underlying CMT2 and axonal sensorimotor neuropathies, expanding the clinical spectra of this gene mutations. To advance the understanding of SMARD1 and associated disorders, we have created a vertebrate zebrafish animal model of IGHMBP2 deficiency. Transient knock-down of zebrafish *ighmbp2* resulted in impaired motor function and muscle defects as observed in human patients. To study the long-term effects of IGHMBP2 deficiency and develop therapies, we have utilized a CRISPR/Cas9 system to create a knockout zebrafish model. Targeting three different exons yielded high rates of mutagenesis, resulting in insertions or deletions in *ighmbp2*. Mutant embryos exhibit motor neuron abnormalities and weak skeletal muscles in IGHMBP2 deficiency. These abnormalities are observed during early development of zebrafish embryos (2 dpf) suggesting that IGHMBP2 is required for development and function of primary motor neurons and skeletal muscles as mutations in human patients are also associated with early congenital abnormalities. The study of IGHMBP2 in zebrafish offers a promising model for the study of dHMN and will lead to a clearer understanding of the cellular and molecular abnormalities underlying this disease.

2550W

A novel mouse model of Uveal Coloboma due to genetic rearrangement on Chr13. *R. P. Alur[1], F. I. Onojafe[1], A. Dutra[2], E. Pak[2], J. Thomas[3], M. Fruttiger[4], W. D. Richardson[5], S. Pieke-Dahl[6], P. F. Hitchcock[6], B. P. Brooks[1].* 1) National Eye Institute, National Institutes of Health, Bethesda, MD, 20892; 2) NHGRI, NIH, Bethesda, MD, 20892; 3) NISC, NHGRI, NIH, Bethesda, MD, 20892; 4) UCL Institute of Ophthalmology, London, U. K; 5) Wolfson Institute for Biomedical Research, University College London, London, U. K; 6) University of Michigan, Ann Arbor, MI.

Statement of purpose: Uveal coloboma is a potentially blinding congenital ocular malformation caused by failure of the optic fissure to close during the 5th week of human gestation and at around 11. 5 days in the mouse. Many genetic studies in human and mouse have suggested genetic heterogeneity and incomplete penetrance. Known genes explain a minority of cases, implying the role of other developmentally regulated genes. In this study, we characterize the RICO (**R**etinal & **I**ris **CO**loboma) mouse and propose a model for genetic rearrangement. **Methods used: The** RICO mouse model was created by the transgenic insertion of a human VEGF (hVEGF) expression vector. The insertion site was localized with FISH using commercially-available BAC clones for mouse chromosome 13 and the original transgenic plasmid. Embryos and mice were genotyped using a combination of FISH and PCR/TaqMan assay and were phenotyped grossly and histopathologically. The insertion site was evaluated using a combination of Comparative Genomic Hybridization (CGH), BAC library construction, shot gun sequencing of two BAC clones and whole genome sequencing of RICO genomic DNA. Further, Array Star mouse lncRNA (long non-coding RNA) array was performed to study the regulatory elements in this genomic region. **Summary of results:** Uveal coloboma in C57BL/6J-RICO mice was generally bilateral (≥60%; N=310), affecting the iris, retina/choroid, and optic nerve. Transmission was semi-dominant with nearly 100% penetrance and homozygous embryonic lethality. The *NSE-hVEGF* transgene inserted on mouse chromosome 13C was transcriptionally silent. CGH array of heterozygote RICO DNA showed no large duplications or deletions on Chr 13. From the BAC library, shotgun sequencing and whole genome sequencing, we propose a model for the duplication, genetic insertion/re-arrangement responsible for the phenotype. Further, analysis of the lncRNA array results with focus on re-arranged genomic region showed down regulation of a lncRNA element in the close proximity of Nr2f1 (3. 64 fold change and intergenic in nature) in the homozygous RICO mouse. Further confirmation of the lncRNA elements by qRT-PCR and investigations into the functions of these regulatory elements are underway.

2551W

Normal palate development requires *Prdm16* expression in the mandible, not the palate. *B. C. Bjork, I. B. Emerson, K. Lundry, L. Pitstick.* Department of Biochemistry, Midwestern University, Downers Grove, IL.

The *Prdm16* transcription factor gene functions in many different tissues and biological processes, including hematopoietic and neuronal stem cell maintenance and brown adipose tissue cell fate determination. *Prdm16*is expressed in both the palate, mandible and tongue,and *Prdm16* loss-of-function mutations cause cleft secondary palate (CP). CP in these mutants occurs secondary to micrognathia due to failed palate shelf elevation similar to human Pierre Robin Sequence-type CP. To determine the role of *Prdm16* during palate formation, we carried out tissue-specific ablation experiments using a novel conditional gene trap allele of *Prdm16, Prdm16cGT*, and palate mesenchyme-specific (*Osr2::cre*) and mandible-specific (*Hand2::cre*) *cre*-expressing strains. Palate-specific ablation of *Prdm16* did not adversely affect palate shelf elevation or fusion; however, we observed micrognathia and CP inmandible-specific knockout embryos. We also observed that *Prdm16* null mutant heads with mandibles and tongues removed undergo normal palate shelf elevation and fusion in an *in vitro* suspension culture system. These results confirm that *Prdm16* expression in the mandible and not the palate, is necessary for normal palate formation. Furthermore, we tested if *Prdm16* expression in the mandible is sufficient for normal palatogenesis via mandible-specific rescue of *Prdm16* expression to the mandible only in an otherwise null *Prdm16cGT* embryo. These crosses yielded both *Prdm16* null mutant embryos with cleft palate and micrognathia and *Prdm16* null mutant embryos with *Prdm16* expression returned to the mandible only that exhibit rescue of the CP phenotype. This result further demonstrates a palate-extrinsic role for *Prdm16* during palate formation. Beyond understanding the developmental mechanism of clefting in *Prdm16* mutants, we are interested in identifying genes and pathways within which *Prdm16* acts during normal mandible and palate formation. Screening of mandible markers via whole-mount *in situ* hybridization identified *Goosecoid* (*Gsc*) as being down-regulated in the mandibular arch in *Prdm16* null mutant embryos. We are examining the possibility that *Gsc* is a direct transcriptional target of PRDM16 using *in vitro* luciferase reporter assays and Chromatin Immunoprecipitation (ChIP). In addition, we are using a Next-Generation sequencing approach to determine global expression differences that occur as a consequence of *Prdm16* loss in the developing mandible.

2552W

Mitochondrial deficiency by Gtpbp3 depletion affects further development in zebrafish. *D. Chen, F. Li, Q. Yang, M. Tian, Y. Chen, M. Guan.* Institute of genetics, Zhejiang University, Hangzhou, Zhejiang, China.

Mitochondria are eukaryotic organelles that generate most of the energy in the cell by oxidative phosphorylation. Mitochondrial respiratory chain deficiencies exhibit a wide variety of clinical phenotypes, which could be caused by either mutations in the mtDNA or mutations in nuclear genes coding for mitochondrial proteins. We have previously reported the identification and characterization of human *MSS1* homolog, *GTPBP3*, the first identified vertebrate gene related to mitochondrial tRNA modification, and found the mutations in human *GTPBP3* manipulate the penetrance of deafness associated with mitochondrial 12S rRNA mutation. The Gtpbp3 is the mitochondrial GTPase evolutionarily conserved from bacteria to mammals. Here, we reported a novel Gtpbp3 depletion animal model using antisense morpholinos in zebrafish. A marked decrease of mitochondrial ATP generation accompanied by increased levels of apoptosis and reactive oxygen species were observed in *gtpbp3* knockdown zebrafish embryos. The Gtpbp3 morphants exhibited melanin, oedema, curved tails, bleeding and most of them died within 5 days post fertilization. Taken together, our results suggest that zebrafish Gtpbp3 is a structural and functional homolog of human and yeast Gtpbp3, and the impairment of mitochondrial ATP generation caused by Gtpbp3-deficiency may lead to zebrafish embryonic development arrest in an apoptosis dependent way.

2553W

A Forward Genetics Approach to Discover Modifiers of Holoprosencephaly and Cleft Lip/Palate. *K. A. Geister[1], A. E. Timms[1], S. Ha[1], R. J. Lipinski[2], D. R. Beier[1].* 1) Center for Developmental Biology and Regenerative Medicine, Seattle Children's Research Institute, Seattle, WA; 2) Department of Comparative Biosciences, University of Wisconsin-Madison School of Veterinary Medicine, Madison, WI.

The discovery of modifier genes in human and mouse has seen slow progress. In mice, this is largely due to the reliance on inherent variability among inbred strains, which makes identifying the causal variant, even in highly resolved mapped loci, quite challenging. We have developed a fast and cost-effective method for genetic mapping of causal ENU-induced mutations using Next Generation Sequencing that combines single nucleotide polymorphism (SNP) discovery, mutation localization, and identification of causal sequence variants. This approach precludes the need for an outcross to facilitate mapping, as we use the ENU-induced variants as SNP markers. This strategy allows us to perform both primary screens and modifier screens on fully inbred lines. We have successfully mapped causal mutations using this approach, and we are currently testing the efficacy of this strategy with regard to modifier discovery using two C57BL/6 congenic strains as our sensitized lines. One is a loss-of-function allele of the SHH transcriptional effector *Gli2*, which can be modified by prenatal ethanol exposure to generate holoprosencephaly-like phenotypes. The other is a spontaneous loss-of-function allele of *Pibf1* that causes cleft lip/palate and ciliopathy-like phenotypes. Both alleles could reveal modifiers of genes involved in craniofacial, brain, and skeletal development as well as those involved in the formation and function of the primary cilium. In the simplest cases, whole-genome sequencing will reveal areas of homozygosity shared between mutants with a consistent "modified" phenotype, and allow us to narrow our focus to the induced variants included in these homozygous regions. We anticipate that we may ascertain more complex genetic interactions as well as new ENU-induced mutant phenotypes that model human birth defects.

2554W

Distinct requirement for the basal plate in the regionalization of the posterior diencephalon. *B. Lee, Y. Jeong.* Genetic engineering, Kyunghee University, Yongin-si, Kyunggi-Do, South Korea.

The developing diencephalon is the posterior region of the embryonic forebrain, and undergoes dynamic regionalization including its subdivision along the dorsoventral and anteroposterior axes. The alar plate of the posterior diencephalon is partitioned into three different prosomeres (designated p1-p3), which develop into the pretectum, thalamus, and prethalamus, respectively. In the present study, we show the developmental outcomes of genetic ablation of cell populations from the diencephalic basal plate. The strategy for conditionally regulated cell ablation is based on the targeted expression of the diphtheria toxin gene (DTA) to the diencephalic basal plate via tamoxifen- induced, Cre-mediated recombination of the ROSADTAallele. Our results show that activation of DTA leads to specific cell loss in the basal plate of the posterior diencephalon, and disrupted early regionalization of distinct alar territories. In the basal plate-deficient embryos, the p1 alar plate displayed reduced expression of subtype-specific markers in the pretectum, whereas p2 alar plate failed to further subdivide into two discrete thalamic subpopulations. We also report that these defects lead to abnormal nuclear organization at later developmental stages. Our data have strong implications for increased understanding of the interactive roles between discrete diencephalic compartments.

2555W

Reeling in the Effect of Decreased Sulfate on Embryonic Development and Risk of Autism Spectrum Disorder through Disruption of *slc13a1* in Zebrafish. *C. G. Perry, L. M. Yerges-Armstrong, A. R. Shuldiner, N. A. Zaghloul.* Division of Endocrinology, Diabetes, and Nutrition, and Program for Personalized and Genomic Medicine, University of Maryland School of Medicine, Baltimore, MD.

Autism spectrum disorder (ASD) affects approximately 1 in 68 children in the United States. The heritability of ASD is estimated to be 50%, suggesting that both genetic and environmental risk factors contribute significantly to the development of ASD. Many of these risk factors have yet to be identified; however, previous studies suggest that sulfate deficiency may be an important risk factor. In fact, such investigations have shown that blood sulfate concentration is reduced in 92% of children with ASD compared to age-matched controls (0. 55 vs. 4. 90 nmol/mg protein, respectively). Though not routinely measured, sulfate is an essential micronutrient integral in the biotransformation of multiple compounds, such as hormones, neurotransmitters, proteoglycans, and xenobiotics, via sulfation. Interestingly, the solute carrier family 13, member 1 *(SLC13A1)* gene, which encodes an apical membrane, sodium-sulfate cotransporter (NaS1) that mediates sulfate (re)absorption across renal and intestinal epithelia, resides in the Autism Susceptibility Locus 1 (*AUTS1*) on chromosome 7q. We previously reported on two rare, non-linked, nonsense variants in *SLC13A1* (c. 34C>T, p. R12X and c. 144G>A, p. W48X) that are enriched in the Amish and associated with hyposulfatemia (P=9x10-20). Mammals with homozygous loss-of-function variants in *Slc13a1* display developmental and growth abnormalities, consistent with the critical role of sulfation in the regulation of growth and development. With this in mind, we injected zebrafish embryos with morpholino targeting *slc13a1*, which resulted in a significant decrease in *slc13a1* mRNA expression, and assessed embryonic growth and neurological development at 48 hours post fertilization. Strikingly, *slc13a1* morphants display 3- to 7-fold higher frequencies of gross, tail, and gastrointestinal abnormalities, along with a 14% reduction in body length, compared to injected controls (P<0. 0001). These results are consistent with abnormal development in tissues known to require sulfate for proper development. Furthermore, *slc13a1* morphants exhibit a 22% decrease in eye area, a 7% decrease in midbrain width, and a 5% increase in forebrain:midbrain width, compared to injected controls (P<0. 0005). Together, these results suggest a role for *slc13a1* and/or sulfate in embryonic growth and neurological development, offer insight into sulfate's role in human physiology and disease, and provide support for additional phenotyping of human R12X and W48X carriers.

2556W

Brain region specific roles for CHD7, the gene mutated in human CHARGE syndrome. *J. M. Skidmore[1], M. A. Durham[2], H. Yao[1], D. M. Martin[1,3].* 1) Dept Pediatrics, Univ Michigan, Ann Arbor, MI; 2) Baylor College of Medicine, Houston, TX; 3) Dept Human Genetics, Univ Michigan, Ann Arbor, MI.

Impairments in central nervous system (CNS) processing of peripheral sensory functions (vision, hearing, balance and olfaction) are common in autism spectrum disorders (ASDs). CHARGE syndrome, a congenital anomaly disorder and leading cause of multisensory impairment and deafblindness, is one of many genetic forms of ASD. The underlying mechanisms leading to ASD behaviors in CHARGE are not known; however, CNS developmental abnormalities including arrhinencephaly, holoprosencephaly, corpus callosum agenesis/dysgenesis and cerebellar defects are common. CHARGE is caused by heterozygous mutations in CHD7, an ATP dependent chromodomain remodeling protein that regulates gene expression. *Chd7* heterozygous mutant mice exhibit defects in development of inner ear and craniofacial structures similar to those in humans with CHARGE. In mice, *Chd7* is expressed in the developing and mature brain regions (olfactory bulb, hypothalamus, midbrain, hindbrain, and cerebellum), and in areas of adult neurogenesis (dentate gyrus of the hippocampus and subventricular zone). Prior studies identified roles for *Chd7* in neurogenesis in the developing mouse inner ear and olfactory placode, where CHD7 is expressed in neural stem cells and regulates proneural gene expression, and in the CNS forebrain subventricular zone, where CHD7 regulates neural stem cell proliferation and fate. Here we present data from conditional mice with loss of one or both copies of *Chd7* in the brain. We also report on the generation and characterization of a novel Chd7-iCre transgenic line for lineage tracing and CHD7-specific deletions. *Nestin-Cre*-mediated *Chd7* deletion in neural progenitors leads to mild anxiety phenotypes, whereas *Wnt-1Cre*-mediated conditional deletion of *Chd7* in the midbrain/hindbrain is postnatal-lethal and results in cerebellar hypoplasia and reduced foliation. Loss of *Chd7* by *Wnt1-Cre* also leads to attenuation of *Wnt1* and *Fgf8* expression domains at the embryonic day 11. 5 midbrain/hindbrain boundary, consistent with roles for *Chd7* in early midbrain-hindbrain patterning. Expression analysis suggests dual origins of CHD7 cells in the cerebellar ventricular zone and rhombic lip. Temporal deletion of *Chd7* in developing granule cellswith Atoh1-CreERT2 mice is underway. Our results provide novel insights into roles for *Chd7* in brain and cerebellar development, and may shed light on mechanisms by which CHD7 haploinsufficiency leads to ASD and other cerebellar mediated behaviors.

2557W

B-cell CLL/Lymphoma 9-like (Bcl9l) and eye development. *D. Wu, A. Anderson, A. Slavotinek.* University of California San Francisco, San Francisco, CA.

B-cell CLL/Lymphoma 9 (*Bcl9*) and B-cell CLL/Lymphoma 9-like (*Bcl9l*) mediate the interaction between β-catenin and two Pygopus proteins, Pygo1 and Pygo2, in Wnt signaling in vertebrates. The interaction between Bcl9 and Bcl9l and Pygo occurs through two highly conserved protein domains called HD1 and HD2. Studies in mice using two 'knock in' models to remove the function of either HD1 or HD2 showed that deletion of HD1 blocked the recruitment of Pygo proteins to the β-catenin transcriptional complex, but enabled β-catenin binding and thus signaling through β-catenin (Cantù et al. , 2014). Deletion of HD2 blocked β-catenin binding, but was predicted to enable Pygo2 binding. Transgenic mutant mice without capacity for HD1 signaling for both Bcl9 and Bcl9l had absence of the lens placode and lens (Cantù et al. , 2014), implying that the Pygo proteins were relevant to lens development. We are interested in the role of Wnt signaling in eye development and chose to further examine the effects of loss of function for Bcl9l to determine if reduced function for this gene alone is relevant to lens formation. We first used a splice site antisense morpholino targeting *bcl9l*, the *Danio rerio* orthologous gene, to screen for eye defects associated with loss of function. Morphant larvae showed a consistent phenotype with a smaller eye, with eye size being dose dependent and not rescued by co-injection of p53 antisense morpholino. Sectioning of morphant larvae showed a smaller lens, although the remainder of the eye appeared normally formed. We then used Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) to make fish with stable loss of bcl9l function, and have generated one mutant that is a homozygous for a frameshift mutation prior to the HD1 domain. This fish appears normal externally, but the second *Danio rerio* orthologue of Bcl9l has not been targeted and reduced function for this gene or *bcl9* may be required for a phenotype. We also investigated the effects of *BCL9L* on Wnt signaling *in vitro*, and found that targeting *BCL9L* using siRNA decreased β-catenin signaling at the RNA level without effects on *PYGO2* or *BCL9* expression in ARPE19 cells from human retinal pigment epithelium. These preliminary results support a role for *Bcl9l* on lens development acting through β-catenin, and our studies are ongoing. Cantù et al. Genes Dev. 2014;28:1879.

2558W

KLF14 involves in controlling inflammation in the white adipose tissue. *C. Tayama[1], E. Takanashi[2], J. Tomikawa[1], H. Okita[3], K. Hata[1], T. Okamura[2], K. Nakabayashi[1].* 1) Dpt. of Maternal-Fetal Biol. , Nat. Res. Inst. for Child Health and Development, Tokyo, Japan; 2) Dept. of Infectious Dis. , Res. Inst. , Natl. Ctr. for Global Health and Med. , Tokyo, Japan; 3) Dept. of Pathology, Keio University School of Medicine, Tokyo, Japan.

KLF14 is imprinted in both mice and humans. eQTL and GWA studies have suggested that KLF14 acts as a master regulator of key metabolic genes in adipose tissue, and demonstrated that cis-regulatory SNPs in the *KLF14* locus are associated with susceptibility to multiple metabolic diseases including T2D. We revealed by qRT-PCR analyses that *Klf14* was highly expressed in the embryos at 9. 5pdc or later and only in the white adipose tissue (WAT) among 15 adult tissues examined. Maternal allele-specific expression of *Klf14* in WAT was consistently maintained in all adult stages examined. We obtained a *Klf14* knock-out mouse line from KOMP. *Klf14*-/- mice showed no morphological and growth abnormalities. We fed *Klf14*-/- mice (n=11) and wild-type littermates (n=7) with a high-fat diet for 8 weeks (from 9 to 16 weeks of ages) and examined their metabolic phenotypes. No significant difference was observed in the items inspected such as body weight, GTT, and ITT. We subsequently subjected the WAT samples from *Klf14*-/- and WT mice with high-fat diet treatment to a gene expression array analysis (n=5 each). Both of GO and GSEA analyses for differentially expressed genes, 1033 up- and 1441 down-regulated genes with t-test *P*-value < 0. 05, revealed that only the latter was enriched with genes related to inflammation and immune cell activation, indicating that obesity-induced inflammation characterized by macrophage infiltration was alleviated in *Klf14*-/- WAT. The expression levels of key pro- and anti-inflammatory adipokines were also altered in *Klf14*-/- WAT. Our results suggest that KLF14 involves in controlling inflammation in WAT possibly through transcriptional regulation of adipokine genes.

2559W

Genetic dissection of globozoospermia reveals a role for centrosomal proteins and regulation of the ubiquitin-proteasome system. *E. C. Oh, S. C. Brodar, N. Katsanis.* Center for Human Disease Modeling, Duke University, Durham, NC.

The ciliopathies are clinically complex multisystemic disorders that have informed our understanding of the cilium and the centrosome. Given that ciliopathy males are infertile, not always because of a structural failure of the sperm flagellum, and that loss of intraflagellar transport proteins in mice results in flagella and spermatid head defects, we asked whether other sperm maturation syndromes such as globozoospermia might be associated with centrosomal dysfunction. Computational analysis of the ten known globozoospermia loci showed a direct link of 7/10 with the centrosome, while cellular localization of five disease proteins were visualized at this organelle in spermatogonial cells. Expression studies during spermatogenesis identified a common profile between the known globozoospermia genes and five ciliary/centrosomal proteins, including PCM1, an interactor of ciliopathy proteins. Ablation of *Pcm1* in the mouse caused male infertility characterized by acrosomeless spermatazoa, defects in nuclear shaping, and impaired maintenance of the mitochondrial sheath. The overlap in mouse and human phenotypes prompted us to sequence *PCM1* in 36 infertile male patients and we discovered mutations that segregated with the phenotype and performed as pathogenic alleles in a zebrafish model. Since protein and organellar degradation is necessary for the maturation of germ cells and that the centrosome is a known hub for proteasomal degradation, we show that globozoospermia proteins are ubiquitinated, they interact with components of the 26S proteasome, and are required for protein clearance. Finally, we show that developmental phenotypes in zebrafish mutants for globozoospermia proteins and sperm defects in *Pcm1*-/- mice can be ameliorated by treatment with a chemical proteasomal agonist. Taken together, our data support a centrosomal hypothesis for globozoospermia and suggest that targeting proteasomal function might be of clinical utility in this disorder, and possibly other ciliopathies.

2560W

Corpus callosum anomalies in fetuses: from fetal pathology to NGS and reverse phenotyping. *T. Attie-Bitach[1,2], C. Alby[1], L. Boutaud[1,2], V. Malan[1,2], L. Mouthon[1], A. Achaiaa[2], C. Gordon[1], N. Bahi-Buisson[1,3], J. Oliviera[1], K. Piquand[1], P. Nitschke[4], C. Bole[5], P. Sonigo[6], A. Millischer-Bellaiche[6], M. Moutard[7], C. Depienne[8,9], D. Heron[8,9], S. Lyonnet[1,10], Y. Ville[11], S. Thomas[1], F. Encha-Razavi[1,2], M. Vekemans[1,2].* 1) INSERM U1163, Institut Imagine, Paris Descartes, Paris, FranceFrance; 2) Service d'Histologie-Embryologie-CytoGénétique, Hôpital Necker-Enfants Malades, APHP, Paris, France,; 3) Service de Neuropédiatrie, Hôpital Necker-Enfants Malades, APHP, Paris, France; 4) Plateforme de Bioinformatique, Institut Imagine, Paris Descartes, Paris, France; 5) Plateforme de Génomique, Institut Imagine, Paris Descartes, Paris, France; 6) Service de Radiologie Pédiatrique, Hôpital Necker-Enfants Malades, Paris, France; 7) Service de NeuroPédiatrie, Hôpital Trousseau, APHP, Paris, France; 8) Service de Génétique, Hôpital Pitié-Salpêtrière, APHP, Paris, France,; 9) INSERM U975, UPMC, Hôpital Pitié-Salpêtrière, Paris, France; 10) Service de Génétique, Hôpital Necker-Enfants Malades, APHP, Paris, France; 11) Service de Gynécologie-Obstétrique, Hôpital Necker-Enfants Malades, APHP, Paris, France.

Corpus callosum (CC) is the major brain commissure connecting the homologous areas of both hemispheres at the midline. CC malformations (CCM) are the most frequent brain malformations with an incidence of 1/4000 newborn, often associated with chromosomal anomalies or mendelian syndromes. Recurrence is observed in 5 % of cases. Children with CCM have an uncertain neuro-developmental outcome. Therefore, counseling remains challenging, especially prenatally. We systematically reviewed the data of 138 fetuses with CCM as isolated or associated autopsy findings in our center. We first completed the cytogenetic analysis by a CGH array then applied whole exome or targeted exome sequencing of a panel of genes involved in CCM in human or animal models. Combined fetal imaging, fetal autopsy, cytogenetic and molecular analysis allowed the identification of the cause in 44 fetuses, 30 % of fetuses with 15% of chromosomal anomaly and 15 % of mendelian disorder. In undiagnosed cases, we used exome sequencing in 10 trios and a targeted high throughput sequencing strategy including 423 genes in 64 fetuses. NGS allowed a diagnosis in 12% of cases, such as PDH deficiency (*PDHA1*), PCH with ACC (*AMPD2*), genitopatelar (*KAT6B*), Primrose (*ZBTB20*), Coffin-Siris (*ARID1A* and *ARID1B*), Chudley Mac Cullough syndrome (*GPSM2*). Interestingly some diagnosis were not possible antenatally due to the absence of specific signs, but reevaluation of fetalpathological data (reverse phenotyping) allowed to support NGS findings. Several novel candidate genes are currently being studied. Since the completion of the Human Genome Project in 2003, the rapid development of CGH and NGS sequencing have allowed the identification of many new genes involved in human disorders involving the CC development. The application of NGS in clinical diagnosis holds much promise and will hopefully enable the identification of pathogenic variants and hence influence diagnosis, genetic counseling and the identification of new genes involved in CCM. All together, combined fetal imaging, fetal necropsy, cytogenetic and molecular analysis allowed the identification of the cause in 40 % of our fetal cohort. The additional anomalies after autopsy found in 17% of fetuses prenatally described as having an isolated CCM and the necessary reverse phenotyping following NGS sequencing underlines the importance of fetal autopsy following pregnancy terminations.

2561W

Network analysis of genome-wide selective constraint reveals a gene network active in early fetal brain intolerant of mutation. *J. Choi[1], P. Shooshtari[1], K. E. Samocha[2,3,4,5], M. J. Daly[2,3,4], C. Cotsapas[1,2,3,4,6].* 1) Department of Neurology, Yale School of Medicine, New Haven CT, USA; 2) Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; 3) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 4) Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA; 5) Program in Genetics and Genomics, Biological and Biomedical Sciences, Harvard Medical School, Boston, Massachusetts, USA; 6) Department of Genetics, Yale School of Medicine, New Haven CT, USA.

Using robust, integrated analysis of multiple genomic datasets, we show that genes depleted for non-synonymous *de novo* mutations form a subnetwork of 72 members under strong selective constraint. We further show this subnetwork is preferentially expressed in the early development of the human hippocampus and is enriched for genes mutated in neurological, but not other, Mendelian disorders. We thus conclude that carefully orchestrated developmental processes are under strong constraint in early brain development, and perturbations caused by mutation have adverse outcomes subject to strong purifying selection. Our findings demonstrate that selective forces can act on groups of genes involved in the same process, supporting the notion that adaptation can act coordinately on multiple genes. Our approach provides a statistically robust, interpretable way to identify the tissues and developmental times where groups of disease genes are active. Our findings highlight the importance of considering the interactions between genes when analyzing genome-wide sequence data. .

2562W

BREATH: A web-accessible database of normal human and mouse lung development developed by the Molecular Atlas of Lung Development Program (LungMAP). *N. C. Gaddis[1], J. L. Levy[1], Y. Xu[2], M. Duparc[1], G. P. Page[3], J. A. Whitsett[2], S. M. Palmer[4], R. F. Clark[1], LungMAP Data Coordinating Center (DCC).* 1) RTI International, Research Triangle Park, NC; 2) Cincinnati Children's Hospital Medical Center, Cincinnati, OH; 3) RTI International, Atlanta, GA; 4) Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC.

The lung is a complex organ with high cellular heterogeneity, and research is needed to define the interactive gene networks and dynamic crosstalk among multiple cell types that coordinate normal lung development. Significant knowledge gaps exist in the understanding of lung development from late fetal to perinatal stages and through early childhood, a critical period when the diverse lung cells go through terminal differentiation and maturation and when alveoli form. Coordination and integration of information from disparate assays and experimental approaches is critically needed to help the pulmonary research community uncover the molecular determinants of lung development. The goal of LungMAP is to build an open-access resource providing a comprehensive molecular atlas of late-stage lung development in humans and mice, making otherwise dispersed data, reagents, and protocols freely available to the research community. We have created the Bioinformatics REsource ATlas for the Healthy lung (BREATH) database, applying novel data management and bioinformatics approaches to manage high-throughput multidimensional experimental data. BREATH is built on a triple store database backbone, integrated with anatomical ontologies for lung development, and provides access to novel web-based tools for the analysis and visualization of data generated by the four Research Centers and Human Tissue Core of the LungMAP Consortium. The current version of BREATH contains confocal immunofluorescence images, *in situ* hybridization images, and 3D confocal and uCT images of developing mouse lung at several time points, as well as single-cell RNAseq data from mouse lung cells. The LungMAP website (www. lungmap. net) provides an entry portal to the BREATH database and tools for exploring and interacting with the lung images and RNAseq data. Future versions of BREATH will incorporate additional data types, including proteomic, lipidomic, metabolomic, and epigenetic data, and novel tools for cross-datatype analysis. A better understanding of the basic molecular pathways that regulate normal lung development will enable development of innovative treatments that successfully manipulate, enhance and improve lung injury repair and regeneration.

2563W

Functional validation in a zebrafish model of *ZFHX4* mutations that cause a novel recessive disorder with progressive ophthalmologic, neurodevelopmental, and endocrine dysfunction. *Y. Jing[1], E. McCormick[2], E. Place[1,2], M. Peng[2], M. Consugar[1], C. Seiler[3], Q. Zhang[1], E. Pierce[1], M. Falk[2], X. Gai[1].* 1) Ocular Genomic Institute, Massachusetts Eye and Ear Infirmary, Harvard Medical School, Boston, MA, USA; 2) Division of Human Genetics, Department of Pediatrics, The Children's Hospital of Philadelphia and University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA; 3) Zebrafish Core Facility, The Children's Hospital of Philadelphia Research Institute, Philadelphia, Pennsylvania, USA.

Background. *ZFHX4* (zinc finger homeobox4), a recently identified 397 kDa transcription factor, that plays an important role in human neural development. A prior report identified CNVs involving this gene in patients with intellectual disability and congenital bilateral ptosis. Here, we report compound heterozygous *ZFHX4* mutations in a patient with multi-system disease identified by whole-exome sequencing (WES). Phenotypes from *ZHFX4* loss of function was studied by morpholino (MO) knockdown in zebrafish. **Case Description**. The proband was a 28-year-old Caucasian man who presented global developmental delay, mild intellectual disability, lifelong ataxia, spasticity, hyperreflexia, progressive resting tremor, nystagmus since 6 months old, complete blindness since age 18 attributable to both optic atrophy and progressive pigmentary retinopathy, inferior cerebellar vermis mild hypoplasia, type I diabetes mellitus , hypothyroidism, severe liver dysfunction in infancy and early childhood, hypocholesterolemia, and short stature. Following enrollment in CHOP IRB study #08-6177, WES was performed on the proband and his unaffected patents. Bioinformatics analysis identified two novel, biparentally-inherited, compound heterozygous mutations in highly conserved regions of *ZFHX4* (c. 7025C>T;p. T2342I) and c. 9233A>G;p. Q3078R). **Results of functional analyses.** The large gene size prevented cDNA rescue. Zebrafish *ZFHX4* morphants exhibited body axis curvature, edema and smaller eye size on 3 day post fertilization (dpf). Histologic examination showed normal eye lamination with well-defined inner and outer segments in control larvae eyes at dpf 6 but. morphants revealed defective photoreceptor morphogenesis, with shortened, and disorganized outer segments, which was further confirmed by transmission electron microscopy Morphants also exhibited defect in sarcomere assembly orientation, showing smaller somites in muscle development. Furthermore, optokinetic response analysis demonstrated an impaired visual behavior in morphant larvae. **Conclusions.** *ZFHX4* gene mutations could cause an autosomal recessive multi-system disorder with global developmental delay, intellectual disability, cerebellar ataxia, spasticity and tremor, blindness due to progressive retinitis pigmentosa and optic atrophy, hearing loss, and multiple endocrine problems (diabetes mellitus, hypothyroid, hypocholesterolemia, and short stature).

2564W

MicroRNA in Cell Fate Determination. *L. Li[1], Y. K. Suen[1], S. Gu[2], H. H. Cheung[1], W. N. Law[1], D. D. Cao[1], W. Y. Chan[1].* 1) The Chinese University of Hong Kong - Chinese Academy of Sciences Guangzhou Institute of Biomedicine and Health Joint Laboratory on Stem Cell and Regenerative Medicine, School of Biomedical Sciences, Faculty of Medicine, The Chinese University of Hong Kong; 2) Department of Human and Molecular Genetics, Baylor College of Medicine, Houston, Texas, USA.

Human induced pluripotent stem cells (hiPSC) have the potential to differentiate into all somatic cell types. Because the introduction of these cells does not evoke immune rejection and the use of these cells does not encounter any ethical issues, iPSCs have become the most promising source of cells for therapy as well as for developmental studies. microRNAs (miRNA) are known to regulate gene expression post-transcriptionally. Previous studies have shown that miRNA can induce de-differentiation and trans-differentiation of cells. These observations suggest that the possibility of manipulating hiPSC differentiation with miRNAs is high. In order to understand the role of miRNAs in lineage-specific differentiation, we have established a platform to differentiate the same iPSCs to three germ layers, namely, hepatocyte for endoderm, nephron progenitors for mesoderm, and neural progenitors for ectoderm. Profiling of miRNA expression in samples collected from different time points (day0, 3, 6 and 10) during differentiation with microarrays allowed us to perform vertical and horizontal comparisons among the three representative lineages of the three germ layers. A combination of target gene prediction (TargetScan and miRDB) and functional annotation (DAVID Functional Annotation Tool) related to development were used to select candidate miRNAs, the expression of which was confirmed by qPCR. By following this strategy, we identify 6 candidate miRNAs potentially regulating neural progenitor differentiation, which are miR-125-5p, -23c, -362-5p, 3714, -4465, -532-5p. In further analysis, PUFY3, ANPEP and PAFAH1B1 were predicted to be regulated by miR-125-5p in neural development, which need further confirmation by luciferase reporter assay. As miR-125-5p increased continuously from day0 to day 10, both indicated by array and qPCR, it is expected to induce neural progenitors formation. The roles of candidate miRNAs in lineage specification including three lineages will be further studied via functional analysis including gain- and loss-of-function assays, and generation of miRNA-specific expressing transgenic mice. We hope to reveal the miRNA(s) that is located in the center of the network of cell fate decision and lineage specification.

2565W

Characterizing neuronal differentiation of human iPSCs by single cell RNA-seq reveals distinct developmental trajectories. *J. D. Robertson[1], E. Oni[2], A. Halikere[4], A. Tiethof[2], M. Swerdel[2], A. Toro-Ramos[2], J. C. Moore[3], Z. Pang[4], R. Hart[2], L. A. Goff[1,5].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD; 2) Department of Cell Biology and Neuroscience, Rutgers University, Piscataway, NJ; 3) Department of Genetics, Rutgers University, Piscataway, NJ; 4) Child Health Institute of New Jersey, Robert Wood Johnson Medical School, New Brunswick, NJ; 5) Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD.

Precise characterization of the transcriptional kinetics of gene expression in a heterogeneous mixture of cells is crucial for identifying key pathways affected in neurodevelopmental diseases. However, in vitro time course experiments on differentiating cells designed to reveal relevant genes, pathways, and expression dynamics are confounded by variation in cell identity, and progress through differentiation. Average gene expression values do not reflect the diversity of intermediate cell types, or alternative cell fates that may be present in culture. We used single cell RNA-seq on mNgn2-induced neurons (iN) from human iPSCs in order to detect differentially regulated genes at 0, 2, 5, and 14 days of differentiation. Next, to order individual cells according to their differentiation state, we hand selected a set of 625 curated genes known to be differentially expressed between neural progenitors and mature neural cell types. We performed independent component analysis on the expression profiles of these genes to establish a 'pseudotemporal' ordering of these cells reflecting their individual progress through differentiation. We identified two discrete developmental paths or trajectories that occur during iN differentiation. One trajectory was significantly enriched for classical markers of sensory neurons such as ILS1, NEUROD1, NTRK1, and NTRK3 while the other was enriched for motor neuron development markers such as BCL6, CDH13, CD44, CRIM1, DIAPH3, and PDLIM1. This suggests that the single factor mNgn2 is capable of inducing multiple neuronal cellular identities in vitro. In addition to this cell fate heterogeneity, we also detected significant differences in the developmental trajectories between iN neurons and neurons differentiated by culture in neural induction media (PSC) alone. Interestingly, on day 0, ~1% had already spontaneously begun differentiation and by day 5 post-mitotic, intermediate, and undifferentiated cell types were detected. This approach can be used to understand the precise effects of disease-associated mutations that lead to defects in neuronal differentiation, and identify the affected phase of differentiation. We are able to detect transient intermediate cell types, key genes and pathways that are differentially affecting during neuronal differentiation, and key transition points that affect cell fate and variability in phenotypic outcomes which would not be revealed by sequencing of bulk RNA.

2566W

Incorporated Cesium-137 in Pregnant Women and Congenital Anomaly Rates in a Chornobyl Impacted Polissia Region of Ukraine. *W. Wertelecki[1], L. Yevtushok[2,3], A. Koerblein[2], N. Zymak-Zakutnia[2,4], O. Komov[5], I. Kuznietsov[6], S. Lapchenko[2], Z. Sosyniuk[2,3].* 1) OMNI-Net for Children, Mobile, AL, USA; 2) OMNI-Net for Children, Rivne, Ukraine; 3) Rivne Province Regional Medical Diagnostic Center, Rivne, Ukraine; 4) Khmelnytsky City Perinatal Center, Khmelnytsky, Ukraine; 5) Rivne Province State Sanitary-and-Epidemiology Service, Rivne, Ukraine; 6) Human and Animal Physiology Department, Eastern-European University, Lutsk, Volyn, Ukraine.

Neural tube defects (NTD), microcephaly and microphthalmia (M/M) population-based rates are persistently higher in the Chornobyl radiation impacted Polissia region of the Rivne Province of Ukraine and are among the highest in Europe. The aim of our investigation was to compare regional differences in Rivne Province of specific congenital anomaly (CA) rates and patterns of incorporated cesium-137 by ambulatory patients and pregnant women. **Methods:** Analysis of yearly population rates (2000-2012) of specific CA and whole body counts (WBCs) of incorporated cesium-137 above 100 Bq of 34786 adult ambulatory patients (2001-2010), including repeated WBCs, and 3492 pregnant women (2011-2013), restricted to first WBC determinations. **Results:** NTD and M/M population rates were persistently and statistically significantly higher in Polissia than in non-Polissia in contrast to cleft lip with or without cleft palate and Down syndrome which were similar across regions. The average WBCs of male and female ambulatory patients in Polissia were 2640 and 2223 becquerel (Bq) respectively, and in non-Polissia ambulatory patients were 507 and 435 Bq respectively. Among pregnant women from Polissia and non-Polissia, the mean WBCs were 2767 and 738 Bq, respectively, and the specific activity was 43. 5 and 11. 7 Bq per kilogram body weight (Bq/kg), respectively. **Conclusions:** The results confirm the co-occurrence of persistent significantly elevated NTD and M/M CA rates and higher incorporation of cesium-137 by pregnant women in Polissia compared to non-Polissia. A direct association of cesium burden (Bq/kg) and of other nuclides with NTD and M/M risk still needs to be determined.

2567W

Full-length cDNA Sequencing of Alternatively Spliced Isoforms Provides Insight into Human Cancer. *T. Clark, T. Hon, E. Tseng.* Pacific Biosciences, Menlo Park, CA.

The majority of human genes are alternatively spliced, making it possible for most genes to generate multiple proteins. The process of alternative splicing is highly regulated in a developmental-stage and tissue-specific manner. Perturbations in the regulation of these events can lead to disease in humans (1). Alternative splicing has been shown to play a role in human cancer, muscular dystrophy, Alzheimer's, and many other diseases. Understanding these diseases requires knowing the full complement of mRNA isoforms. Microarrays and high-throughput cDNA sequencing have become highly successful tools for studying transcriptomes, however these technologies only provide small fragments of transcripts and building complete transcript isoforms has been very challenging (2). We have developed a technique, called Iso-Seq™ sequencing, that is capable of sequencing full-length, single-molecule cDNA sequences. The method employs SMRT® Sequencing from PacBio, which can sequence individual molecules with read lengths that average more than 10 kb and can reach as long as 40 kb. As most transcripts are from 1 – 10 kb, we can sequence through entire RNA molecules, requiring no fragmentation or post-sequencing assembly. Jointly with the sequencing method, we developed a computational pipeline that polishes these full-length transcript sequences into high-quality, non-redundant transcript consensus sequences. Iso-Seq sequencing enables unambiguous identification of alternative splicing events, alternative transcriptional start and polyA sites, and transcripts from gene fusion events. Knowledge of the complete set of isoforms from a sample of interest is key for accurate quantification of isoform abundance when using any technology for transcriptome studies (3). Here we characterize the full-length transcriptome of paired tumor/normal samples from breast cancer using deep Iso-Seq sequencing. We highlight numerous discoveries of novel alternatively spliced isoforms, gene-fusion events, and previously unannotated genes that will improve our understanding of human cancer. (1) Faustino NA and Cooper TA. *Genes and Development*. 2003. 17: 419-437(2) Steijger T, et al. *Nat Methods*. 2013 Dec;10(12):1177-84. (3) Au KF, et al. *Proc Natl Acad Sci U S A*. 2013 Dec 10;110(50):E4821-30.

2568T

Conventional and Pagetic Giant Cell Tumor of Bone: are they distinct entities? *G. Divisato[1], T. Esposito[1], L. Pazzaglia[2], D. Merlotti[3], M. S. Benassi[2], D. Rendina[4], L. Gennari[3], F. Gianfrancesco[1].* 1) Institute of Genetics and Biophysics, National Research Council of Italy, Naples, Italy; 2) Laboratory of Experimental Oncology, Rizzoli Orthopaedic Institute, Bologna, Italy; 3) Department of Clinical, Surgical and Neurological Sciences, University of Siena, Siena, Italy; 4) Department of Clinical and Experimental Medicine, Federico II University, Naples, Italy.

The alteration of bone turnover could result in the uncontrolled proliferation of mesenchymal spindle-like stromal cells, as in the case of Giant cell tumor of bone (GCT), an aggressive bone tumor also characterized by pathological multinucleated osteoclast-like giant cells that rarely could arise on tissues affected by Paget's Disease of Bone (GCT-PDB). A recent study showed that conventional GCT is due to recurrent somatic mutations in *H3F3A* gene, that encodes a histone variant H3. 3, in the mesenchymal stromal cells. Otherwise, the genetic and molecular basis of GCT-PDB remain to be defined. We analysed the genetic profile of mesenchymal stromal cells selected "ad hoc" by Laser Capture MIcrodissection (LCM), isolated from heterogenous tissue of conventional GCT to reduce the contamination of the surrounding cells identifying somatic mutations in *H3F3A* gene in 38 out of 44 cases (86%). In contrast, the analysis of patients with Paget's disease of bone (PDB) associated with giant cell tumor did not show any mutation in *H3F3A* gene, at both somatic and germline level, suggesting a different genetic background. We recently reported an extended Italian family in which 4 out of 14 PDB affected members developed multiple GCTs at pagetic skeletal sites. Whole exome sequencing, in this family identifies a missense mutation in a novel uncharacterized zinc finger protein, that interacts with an histone reader to form a transcription coregulator complex. These evidences suggested that GCT, associated or not with PDB, is due to mutations in different genes, both acting in nuclear context and also showed a different histological background, suggesting that common or different molecular signatures are responsible for these two phenotypes.

2569W

A method for characterization of breast cancer samples from archived formalin-fixed, paraffin-embedded (FFPE) tissues. *P. Apopa[1], M. Orloff[1,2].* 1) Dept. of Epidemiology, University of Arkansas for Medical Sciences, Little Rock, AR; 2) Winthrop P. Rockefeller Cancer Institute, University of Arkansas for Medical Sciences, Little Rock, AR.

Development of molecular tools in diagnosis and prognosis of cancer is critical to the understanding and management of the disease. The gold standard for genetic comparative studies and genotyping is DNA sequencing. Using archival collection of formalin-fixed, paraffin-embedded (FFPE) breast cancer samples with known clinical outcomes is a good resource for retrospective studies. Unfortunately the preparation, preservation and age of FFPE samples have led to DNA fragmentation making it very challenging to obtain a good quality and decent sized DNA fragments for downstream analyses. This study sought to investigate the suitability of using archival tissues to assess previously identified and novel mutations through direct DNA sequencing. Further, we sought to validate the FFPE findings in matched frozen fresh tissues from the same individuals. Using FFPE samples, we isolated genomic DNA and miRNA. DNA purity and quality was determined by spectroscopy. Similarly extraction of DNA and miRNA was done in matched fresh frozen tissues. DNA was amplified using specific primers targeting exonic regions, exon/intron junctions and 3'UTR of the *OLFML3* and *EGFR.* The amplicons were gel purified and sequenced by Sanger method. 20 (FFPE and fresh frozen tissue) samples were sequenced. We determined that by using amplicons not exceeding 300bps, we can get clear sequences from FFPE samples. We identified non-synonymous coding mutations in both *EGFR* and *OLFML3*. A high concordance was also observed in FFPE and matched frozen samples. Like previous studies based on the small fragment size of miRNA, it was not challenging to replicate miRNA expression profiles in FFPE breast cancer tissues compared to previously reported expression profiles. Specifically, miR135A, miR144 and miR451 were relatively overexpressed in FFPE breast cancer tissues validating previously reported findings in fresh frozen tissues. FFPE samples can therefore be effectively used to identify mutations and regulatory elements on breast cancer patient samples. Further, this can also aid in identifying suitable biomarkers for late stage breast cancer. The age of the FFPE sample will influence the quality of nucleic acid product.

2570T

Systematic identification of significantly mutated regions reveals a rich landscape of functional molecular alterations across cancer genomes. *C. L. Araya[1], C. Cenik[1], J. A. Reuter[1], G. Kiss[2], V. S. Pande[2], M. P. Snyder[1], W. J. Greenleaf[1].* 1) Genetics, Stanford University, Stanford, CA; 2) Chemistry, Stanford University, Stanford, CA.

Cancer genome sequencing studies have primarily focused on identifying cancer-driver genes from the increased accumulation of protein-altering mutations. However, the positional distributions of coding mutations, and the 79% of somatic variants in exome data that do not alter protein sequence or RNA splicing, remain largely unstudied. We employed density-based clustering methods on ~4,700 exomes from 21 tumor types to detect variably-sized significantly mutated regions (SMRs). SMRs reveal recurrent alterations across a diverse spectrum of coding and non-coding elements, including microRNAs, transcription factor binding sites, and untranslated regions that are individually mutated in up to ~15% of samples in specific cancer types. The discovery of non-coding SMRs constituting the second, third, and fourth most frequent non-coding alterations described in any cancer-type to date, and validated in independent cohorts and reporter assays, firmly establish the utility of our approach for evaluating pathogenic, exon-proximal non-coding variation. Mapping SMRs to protein structures revealed spatial clustering of somatic mutations at known and novel cancer-driver domains and molecular interfaces. These include recurrent alterations at a histone H3. 1 interface, with an associated potential impact on patient survival. We demonstrate cancer-specific SMR mutation frequencies within PIK3CA domains, as exemplified by a previously uncharacterized α-helix that is recurrently and directionally mutated in endometrial cancers. Molecular dynamics indicate mutations in this region alter regulatory subunit (PIK3R1) binding through modulation of a salt-bridge network thereby describing a novel mechanism of pathogenic alteration in a key oncogene. SMRs often associated with changes in gene expression and signalling. For example, NFE2L2 alterations affecting KEAP1-binding interfaces are associated with consistent transcriptional signatures across bladder, endometrial, lung squamous cell carcinomas, and head and neck cancers. Yet, mutation frequencies in SMRs demonstrate that distinct protein regions can be differentially mutated among tumor types, and display differential signaling as evidenced in TP53 and PIK3CA. The functional diversity of SMRs underscores both the varied mechanisms of oncogenic misregulation and valuable opportunities in complementing extant, gene- and pathway-level with unbiased, multi-scale analysis of genomic variation to identify disease drivers. .

2571W

Evaluation of candidate risk loci in familial lung cancer families linked to 6q. *J. E. Bailey-Wilson[1], C. L. Simpson[1], A. Musolf[1], S. M. Pinney[2], M. de Andrade[3], C. Gaba[4], P. Yang[3], M. You[5], A. G. Schwartz[6], D. Mandal[7], Y. Liu[8], M. R. Spitz[8], E. Y. Kupert[5], C. I. Amos[9], M. W. Anderson[5].* 1) Computational and Statistical Genomics Branch, NIH/NHGRI, Baltimore, MD; 2) University of Cincinnati College of Medicine, Cincinnati, OH; 3) Mayo Clinic, Rochester, MN; 4) University of Toledo Dana Cancer Center, Toledo, OH; 5) Medical College of Wisconsin, Milwaukee, WI; 6) Karmanos Cancer Institute, Wayne State University, Detroit, MI; 7) Louisiana State University Health Sciences Center, New Orleans, LA; 8) Baylor College of Medicine, Houston, TX; 9) Geisel School of Medicine, Dartmouth College, Lebanon, NH.

We have published evidence of linkage of familial lung cancer (FLC) risk to a region on 6q (PMID: 15272417, 20215501). Wilson et al. (doi:10. 1038/onc. 2013. 396) found that *EYA4* (in our linkage region) was frequently inactivated biallelically in sporadic lung cancer (LC), displays tumor suppressor gene-like properties and affects DNA repair, and that 5 of 17 common SNPs in this 0. 3Mb region showed nominal association to FLC ($p<0.05$) although these associations are not significant after correction for multiple testing. Here, we sequenced 75 individuals from our 9 most strongly 6q-linked families, using Illumina technology and a custom Agilent kit to capture 37Mb of chromosome 6 from 130Mb to 167Mb. To detect sharing of variants among affecteds in each family while allowing for phenocopies, singlepoint affected-only linkage analysis was performed on lung cancer affection status for all sequence variants using the Elston-Stewart algorithm implemented in the R package paramlink assuming penetrance of 0. 01, 0. 1 and 0. 1 for dd/Dd/DD and a disease allele (D) frequency of 0. 01, at theta=0. The highest total LOD score (summed across families) in the 37Mb targeted region was not near *EYA4*. No family had a variant in or near EYA4 that was on the linked haplotype or had a LOD score close to the maximum family-specific LOD score in the region. The highest total LOD score in the region was found at a missense coding variant, rs41267809 at 6:160953642-SNV (LOD=1. 099), in the lipoprotein-A, *LPA*, gene. This signal was driven almost entirely by a single family. This variant has a frequency ranging from 1-3% in European-derived populations in 1000 Genomes and 3 heterozygotes for this variant were observed in a replication sample of 55 independent FLC cases (5. 4%). We have previously published evidence that a rare variant in PARK2 may be a risk allele in one of these families although there are also other candidate regions in this family. Other families have their strongest linkage signals in other non-coding regions including intronic variants in ARID1B. Additional annotation of non-coding variants is ongoing in these families. Thus, EYA4 is unlikely to be harboring functional variants that explain the linkage to 6q in these families. LPA is a candidate in one family. Non-coding elements or coding variants in different genes across families may explain the multipoint linkage signal in this cancer candidate gene-rich region.

2572T

DKK1 is upregulated in mesenchymal stem cells of patients with Multiple Myeloma. *O. Bashti[1, 2], M. Mashadikhan[2], A. Saremi[2], M. Ahmadvand[2], M. Noruzinia[1, 2].* 1) Dept. Medical Genetics, Tarbia Modares University, Tehran, tehran, Iran; 2) Sarem Cell Research Center, Tehran, Iran.

Multiple Myeloma Mesenchymal stem cells contribute primarily in osteolytic lesion due to their impaired osteogenic potential. Bone lesions in patients with multiple myeloma are in addition a sign of enhanced osteoclast formation of MM MSCs. Recent studies have reported a significant role of wnt signaling in aberrant osteogenic activity of MM MSCs. In this research the role of DKK1 as one of the wnt signaling pathway was investigated in MM MSCs compared to normal MSCs. Material and methods:BM nucleated cells (BMNC) from 6 normal donors and 6 MM patients were plated, cultured ina-MEM medium aftercentrifugation. MSCs of passage 3 were put through flowcytometry RT PCR and in vitro differentiation to confirm presenceof MSCs. RT PCR for transcripts of genes involved in wnt signaling RUN2 and DKK1, and VEGF were performed to compare the two patients and normal groups. Results:Immunophenotyping confirmed absence of myeloma cells. RT PCR, invitro differentiation and flowcytometry confirmed the presence of MSCs. RUNX2 was expressed in mesenchymal stem cells of MM patients. However the expression was significantly lower in patient than normal group ($p<0.05$). On the other hand DKK1 showed an increase in expression MM MSCs compared to normal MSCs($P<0.05$). VEGF shows an increase in expression in MM MSCs. Discussion:Bone marrow Mesenchymal stem cells are among factors in MM microenvironment which play crucial roles directly and indirectly. We show in this research that MM MSCs are different from normal MSCs at least in wnt signaling pathway. Genome wide analysis has shown that MM MSCs are different in expression of many genes involved in osteoblastogenesis. It has been shown by our team and others that these two cells are epigenetically different. In total, our results confirm the role of MM MSCs and wnt signaling in the pathogenesis of the bone lesions in MM patients. Epidrug therapy in mesenchymal stem cells targeting wnt to normalize their expression could enhance the efficacy of the current treatments of osteolytic lesions in these patients.

2573W

Advances in Breast Cancer Biomarker Discovery Methods. *K. Bramlett[1], J. Schageman[1], V. Bagai[1], K. Lea[1], J. Gu[1], J. Wittliff[2].* 1) Life Sciences Solutions Group, Thermo Fisher Scientific, Austin, TX 78744; 2) Dept of Biochemistry and Molecular Biology, Institute for Molecular Diversity and Drug Design, University of Louisville, Louisville, Kentucky 40292.

Forecasting clinical behavior and therapeutic response of human cancer currently utilizes a limited number of tumor markers in combination with characteristics of the patient and their disease. Although few tumor markers and molecular targets exist for evaluation, the wealth of information derived from recent sequencing advancements provides greater opportunities to develop more precise tests for diagnostics, prognostics, therapy selection and monitoring in the future. The objectives of this study are to decipher miRNA and mRNA expression profiles of laser capture microdissection (LCM)-procured carcinoma cells and intact serial sections of breast cancer tissue samples using next generation sequencing (NGS) methods. Our hypothesis is that miRNA signatures discerned from specific carcinoma cell populations more precisely correlate with behavior than that provided by conventional biomarkers from intact tissue samples. Additionally, we hypothesize the data generated in this study will present mRNA signatures informative for breast cancer research and support our miRNA findings through suggesting relevant miRNA:mRNA target associations. De-identified frozen breast cancer research samples of primary invasive ductal tumors of known grade and biomarker status containing 35-70% tumor were selected from an IRB-approved Biorepository. Comparison of expressed miRNAs from intact tissue sections with those of cognate tumor cells procured by LCM revealed, in general, that smaller defined miRNA gene sets were expressed in LCM isolated populations of carcinoma cells. In addition to miRNA sequencing, targeted RNA sequencing with the Ion AmpliSeq™ Transcriptome Human Gene Expression Kit was used to capture mRNA expression information. Data presented here demonstrates high mapping rates for targeted mRNA (>91% of reads) and miRNA (> 88% of reads) libraries. We also demonstrate high technical reproducibility between multiple libraries from the same tumor sample for both mRNA ($R>0.99$) and miRNA ($R>0.97$) libraries. Results suggest potential miRNA:mRNA target associations identified in our set of breast cancer research samples. These data provide insights into breast cancer biology that may lead to new molecular diagnostics and targets for drug design in the future as well as an improved understanding of the molecular basis of clinical behavior and potential therapeutic response.

2574T

Recrutiment mechanisms for the establishment of an African American breast cancer cohort for genetic studies in Alabama. *M. R. Chandler, N. D. Merner.* Department of Drug Discovery and Development, Harrison School of Pharmacy, Auburn University, 2316 Walker Bldg. , Auburn, AL.

African American (AA) women have a higher incidence rate of breast cancer (BC) under the age of 40 when compared to Caucasians. Considering that an early age of onset is a hallmark of hereditary BC, our laboratory has developed protocols to recruit AA BC individuals/families in order to identify the genetic risk factors that are contributing to this BC disparity. The Auburn University Institutional Review Board (IRB) approved two recruitment protocols (#14-232 and #15-111); #14-232 involves hospital recruitment through collaboration with East Alabama Medical Center (EAMC), and #15-111 is a community-based recruitment (CBR) mechanism. The enrollment criteria include individuals (1) diagnosed with BC (at any age) who have a family history or (2) diagnosed with BC under the age of 45 without a family history. Participation requires a blood draw. At the EAMC Cancer Center, a research nurse consents patients once identified by EAMC oncologists (recruitment began on February 11, 2015). CBR targets medically under-served individuals who are not usually offered the chance to participate in research projects; it is a strategic mechanism that aims to reach out to the community in order to educate and build trust (recruitment began on March 23, 2015). DNA is extracted from each blood sample and stored in a DNA bank (IRB #14-335). To date, 69 DNA samples are stored in the DNA Bank from 36 BC families; 18 probands (average age of onset: 38; age range: 23-52) have been recruited through EAMC, and 18 probands (average age of onset: 48; age range: 34-60) have been recruited through CBR. CBR has been the most effective mechanism of AA recruitment (11/18) compared to EAMC (4/18). Three key AA families have been recruited through CBR with a total of 8 (3 affected), 6 (5 affected) and 4 (3 affected) individuals recruited per family. All probands are being screened for 87 known or candidate BC susceptibility genes using an Agilent Technologies custom-designed HaloPlex probe capture kit followed by massively parallel sequencing on an Illumina platform. Mutation-negative families will be exome sequenced in order to identify novel susceptibility genes; no AA exome sequencing studies have been published to date. This effort marks the start of a unique AA cohort from the under-studied Alabama population, and demonstrates how we plan to contribute to the understanding of AA BC genetics and disparities.

2575W

Replication of cervical cancer GWAS-identified variants in Taiwanese population. *T. Chang[1], Y. Yang[2,3], Y. Lee[1,4,5], T. Chen[2], W. Lin[1], S. Chang[1].* 1) Med Res Dept, Mackay Memorial Hosp, New Taipei, Taiwan; 2) Dept of Gynecology and Obstetrics, Mackay Memorial Hosp, Taipei, Taiwan; 3) Dept of Gynecology and Obstetrics, Taipei Medical University, Taipei, Taiwan; 4) Pediatrics, Mackay Memorial Hosp, Taipei, Taiwan; 5) Pediatrics, Taipei Medical University, Taipei, Taiwan.

Genetic epidemiological studies show that genetic factors contribute significantly to cervical cancer carcinogenesis. Recently, a few genome-wide association studies (GWAS) have revealed new genetic polymorphisms associated with cervical cancer susceptibility. The aim of this study is to replicate four GWAS-identified single nucleotide polymorphisms (SNPs) which were associated with cervical cancer in Chinese women. The rs13117307 *C/T*, rs8067378 *A/G*, rs4282438 *T/G*, and rs9277952 *G/A* SNPs were genotyped in 507 cervical squamous cell carcinoma (CSCC) patients and 507 age/sex matched healthy controls by using the Pre-Developed TaqMan Allelic Discrimination Assay. The presence and genotypes of HPV in CSCC patients were determined by PCR. In our case-control study, only rs4282438 *G/G* genotype and *G* allele were found to significantlyassociate with CSCC risk (OR = 1. 96, 95% CI = 1. 42-2. 72, *P* = 4. 18 x 10-5; OR = 1. 41, 95% CI 1. 14-1. 75, *P* = 1. 32 x 10-3, respectively). In addition, significant increases of *G/G* genotype and *G* allele frequencies were observed between HPV-16 positive CSCC patients and controls (OR = 2. 25, 95% CI 1. 55-3. 28, *P* = 1. 75 x 10-5; OR = 1. 61, 95% CI 1. 24-2. 06, *P* = 2. 73 x 10-4, respectively). Although we validated the GWAS-associated rs4282438 SNP in our Taiwanese population, the result was inconsistent with the Chinese one which showed rs4282438 *G* allele is protective against cervical cancer (OR = 0. 75, 95% CI 0. 71-0. 79, *P* = 4. 52 x 10-27). In conclusion, our findings provided convincing evidence that rs4282438 *T/G* significantly associated with CSCC susceptibility.

2576T

Germline cytotoxic lymphocytes defective mutations in Chinese lymphoma patients. *X. Chen, Y. Zhang, F. Wang, W. Teng, Y. Lin, X. Han, F. Jin, Q. Yin, C. Tong, H. Liu.* Hebei Yanda Ludaopei Hospital, Langfang, Hebei, China.

PurposeIt has been reported that some lymphoma patients may harbor mutations in PRF1, UNC13D, STX11, STXBP2 or SH2D1A, which caused function defects of cytotoxic lymphocytes. Data of the association between genetic defects and the development of lymphoma in Chinese patients is limited to date. The aim of the present study is to detect germline cytotoxic lymphocytes defective mutations in Chinese lymphoma patients. **Methods** Patients with either Hodgkin or non-Hodgkin lymphomas were collected. All coding exons and flanking sequences of UNC13D, PRF1, STXBP2, STX11, SH2D1A, and XIAP were amplified and sequenced. **Results**A total of 90 patients were collected (age, 3-60 years), including 48 males and 42 females. Mutations were observed in 24 of them (26. 67%). Sixteen patients carried mutations in UNC13D, including twelve with monoallelic mutations, one with homozygous mutation and three with compound heterozygous mutations. Seven patients had PRF1 mutations, including four with monoallelic mutations, one with homozygous mutation and two with compound heterozygous mutations. One patient carried STX11 monoallelic mutation. All mutations were confirmed to be germline derived. UNC13D c. 2588G>A/p. G863D mutation was detected in nine patients (10. 00%) and in only five of 210 controls (2. 38%). It was predicted to be pathogenic and the 1000 genomes database showed it reserved only in Chinese and Japanese populations. **Conclusion**The present study provides evidence that impaired cytotoxic machinery may represent a predisposing factor for the development of lymphoma. Moreover, it describes a distinct mutation spectrum in Chinese lymphoma patients, with mutations in UNC13D the most frequent. In addition, we find UNC13D c. 2588G>A mutation to be a founder mutation in Chinese patients.

2577W

Germline and somatic HIF2A mutations associated with Phaeochromocytoma and Paragangliomas. *G. R. Clark[1], D. M. Walsh[2], G. Kirby[2,3], M. A. Simpson[4], R. C. Trembath[4], E. R. Woodward[2,3], E. R. Maher[1].* 1) Medical Genetics, University of Cambridge, Cambridge, Cambridgeshire, United Kingdom; 2) Centre for Rare Diseases & Personalised Medicine, University of Birmingham, Birmingham, United Kingdom; 3) West Midlands Regional Genetics Service, Birmingham Women's Hospital, Birmingham, United Kingdom; 4) Division of Genetics and Molecular Medicine, King's College London, London, United Kingdom.

Phaeochromocytomas (PCCs) and paragangliomas (PGLs) are rare autonomic nervous system tumours, which typically hypersecrete catecholamines. A notable feature of PCCs/PGLs is the high incidence of inherited cases, such that germline mutations in at least 12 susceptibility genes can be detected in 30-40% of all cases. Recent gene expression profiling studies have shown that a large proportion of inherited cases show activation of the hypoxic gene response pathways. It has also been shown that loss of function mutations in *VHL* and *SDH* subunit genes lead to the stabilisation of the hypoxia-inducible factor (HIF) proteins. These HIF proteins are a family of transcription factors (HIF-1, HIF-2 & HIF-3) that bind to and activate multiple genes associated with angiogenesis, glycolysis and cell growth. The stability of the HIF proteins is dependent on two specific proline residues located in the oxygen dependent degradation domain, that are hydroxylated by the oxygen dependent prolyl hydroxylases (PHDs). Once hydroxylated, the HIF proteins can be targeted for proteasomal degradation by the VHL complex. Recently, several groups have reported gain of function mutations in *HIF2A/EPAS1* in PCC/PGL cases, which result in the stabilisation of the HIF-2 protein. To date, the majority of the *HIF2A* mutations have been detected as somatic mutations, with little evidence of germline mutations. Here we report the finding of six novel variants, four of which were found present in the germline, and cluster around the secondary hydroxylation site. We also present functional data, which provides further evidence that these variants are pathogenic.

2578T

Identification of candidate target regions physically interact with prostate cancer risk loci. *MJ. Du[1], L. Tillmans[3], JZ. Gao[4], N. Sahr[2], TZ. Yuan[1], RL. Dittmar[1], T. Wang[2], SN. Thibodeau[3], L. Wang[1].* 1) Department of Pathology and Cancer Center, Medical College of Wisconsin, Milwaukee, WI 53226; 2) Division of Biostatistics, Institute for Health & Society, Medical College of Wisconsin, Milwaukee, WI 53226; 3) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55905; 4) Beijing 3H Medical Technology Co. Ltd. , Beijing, China, 100176.

Genome wide association studies have identified more than 80 common variants that are associated with prostate cancer risk. However, the vast majority of these SNPs lie in noncoding regions of the genome. To testwhether these risk SNPs regulate their target genes through long-range chromatin interactions, we applied a capture-based 3C-seq to investigate possible *cis*-interactions at ten prostate risk loci in six cell lines. We identified statistically significant interactions between risk region-contained regulatory elements and their putative target genes, such as: CAPG at 2p11. 2, RFX at 6q22. 1, NFASC at 1q32. 1, PCAT1, MYC and PVT1 at 8q24. 1. The most significant interaction was between capture fragment (chr2:85769616-85778503-85781283) at 2p11. 2 and the promoter region of the gene CAPG. This interaction was not tissue-specific because we observed it in all cell lines tested. However, the interactions at 6q22. 1 loci with RFX and interactions at 1q32. 1 with NFASC were prostate tissue specific because they were only found in prostate cell lines. To further determine whether the tag SNPs rs2028898 (2p11. 2), rs10089608 (8q24. 21), rs339331 (6q22. 1), rs4245739 (1q32. 1) were associated with expression levels of target genes *CAPG, RFX, NFASC, PCAT1* and *MYC*, we carried out expression quantitative trait locus (eQTL) analysis using RNA-seq data on 476 normal prostate tissues. We observed suggestive eQTL signal at rs699664 (a proxy for rs2028898, r2 = 0. 99) for *CAPG*, rs339331 for *RFX*, rs4245739 for *NFASC*, and rs10086908 for PCAT1 (p value<0. 05). To validate these interactions, we applied 3C-qPCR and 4C-seq and confirmed their interactions at risk regions of 2p11. 2 and 8q24. We found that these significant interactions were characterized by DNase I hypersensitive and active histone modifications that correlate with active enhancers (H3K27ac, H3K4me1), active promoters (H3K4me3) and/or binding sites for specific transcription factors. Our results demonstrate that the capture-based 3C-seq along with eQTL analysis may facilitate identification of functional SNPs and their candidate genes at prostate cancer risk regions. Further understanding genetic effect and biological mechanism of these chromatin interactions will shed light on the newly discovered regulatory role of the risk locus in PC etiology and progression.

2579W

Discovery of Rare Variants in Prostate Cancer Susceptibility via Custom Microarray Genotyping. *N. Emami[1], T. Hoffmann[1], J. Mefford[1], K. Lindquist[1], C. Cario[1], C. G. Tai[1], E. Wan[1], S. Wong[1], J. Gollub[2], A. Finn[2], D. Aaronson[4], J. Presti[4], L. Habel[3], E. Jorgenson[3], L. Sakoda[3], M. Kvale[1], P. Kwok[1], C. Schaefer[3], N. Risch[1,3], S. Van Den Eeden[3], J. Witte[1].* 1) University of California, San Francisco, San Francisco, CA; 2) Affymetrix, Inc. , Santa Clara, CA; 3) Division of Research, Kaiser Permanente, Northern California, Oakland, CA; 4) Department of Urology, Kaiser Oakland Medical Center, Northern California, Oakland, CA.

The influence of rare genetic variants on common disease susceptibility is currently an area of wide interest. In light of the missing heritability of complex human traits and the ongoing search for causal germline biomarkers of disease, it has been hypothesized that rare variants with high effect sizes may contribute considerably to complex trait heritability and underlie the signals of common single nucleotide polymorphisms (SNPs) discovered in genome wide association studies (GWAS) via linkage disequilibrium (LD). Accordingly, it has been shown that moderate penetrance rare variants can drive synthetic associations of tagging common polymorphisms and a number of rare variants have been implicated in common disease susceptibility. However, the minor allele frequencies of putatively causal rare variants presents certain challenges to their discovery. While fine mapping all significant, common GWAS lead SNPs through next generation sequencing may detect all variants at a given locus, the current cost of sequencing precludes the usage of cohort sizes large enough to assess the significance of rare variants. Moreover, although certain efforts have demonstrated the feasibility of imputing rare variants using GWAS data, incomplete LD presents uncertainty and obstacles to their imputation. Previously we described the design of a custom Affymetrix Axiom microarray for directly assaying putative rare variants and fine mapping loci previously associated with prostate cancer (PrCa), a highly heritable, common disease in men and ideal complex trait for studying rare variant association due to appreciable missing heritability. Using our custom array, 416,047 markers were genotyped in a population of 8,480 prostate cancer cases and 8,128 controls nested within the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH), ProHealth and California Men's Health Study. Preliminary results replicate previously discovered PrCa susceptibility loci at 7p15. 2, 8q24. 21, 10q11. 23, 11q13. 3, 17q12, and 19q13. 33, including significant and suggestive p-values for SNPs in the intermediate (1% < MAF < 5%) and rare (MAF < 1%) allele frequency ranges in single marker association tests. Future conditional analyses may distinguish whether these associations are independent signals, and to what extent rare variants drive heritability and synthetic associations at PrCa susceptibility loci.

2580T

Hormone-related pathway genes and risk of breast cancer subtypes in African American women. *S. A. Haddad[1], K. L. Lunetta[2], E. A. Ruiz-Narváez[1], J. T. Bensen[3], C. C. Hong[4], L. E. Sucheston-Campbell[4], S. Yao[4], E. V. Bandera[5], L. Rosenberg[1], C. A. Haiman[6], M. A. Troester[3], C. B. Ambrosone[4], J. R. Palmer[1].* 1) Slone Epidemiology Center at Boston University, Boston, MA; 2) Department of Biostatistics, Boston University School of Public Health, Boston, MA; 3) Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY; 5) Cancer Prevention and Control, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ; 6) Department of Preventive Medicine, Keck School of Medicine, University of Southern California/Norris Comprehensive Cancer Center, Los Angeles, CA.

Background: Breast cancer (MIM 114480) in women of African ancestry (AA) has been relatively understudied. Given extensive evidence that steroid hormones affect breast cancer risk, it is possible that variants in steroid hormone pathway genes may contribute to susceptibility in this and other populations. *Methods:* We investigated variation in steroid hormone genes in relation to risk of breast cancer, overall and by subtype, in the African American Breast Cancer Epidemiology and Risk (AMBER) Consortium, which includes data from four large studies of AA women – the Carolina Breast Cancer Study, the Women's Circle of Health Study, the Black Women's Health Study, and the Multiethnic Cohort Study. Genotyping and imputation yielded data on 143,934 SNPs in 308 candidate genes for 3663 breast cancer cases (1098 estrogen-receptor negative (ER-), 1983 ER+, 582 ER unknown) and 4687 controls. Gene-level associations were evaluated using the adaptive rank truncated product (ARTP) statistic. *Results:* The most significantly associated genes were *GHRH* (MIM 139190), *CALM2* (MIM 114182), *CETP* (MIM 118470), and *AKR1C1* (MIM 600449) for overall breast cancer (nominal p ≤0. 01); *NR0B1* (MIM 300473), *IGF2R* (MIM 147280), *CALM2*, *CYP1B1* (MIM 601771), and *GRB2* (MIM 108355) for ER+ breast cancer (p ≤0. 02); and *PGR* (MIM 607311), *MAPK3* (MIM 601795), *MAP3K1* (MIM 600982), and *LHCGR* (MIM 152790) for ER- disease (p ≤0. 02). Evaluation of all SNPs with pairwise r2 <0. 8 in the top genes identified 10 common SNPs (in *CALM2*, *CETP*, *NR0B1, IGF2R, CYP1B1*, and *PGR)* that were significantly associated with overall, ER+, or ER- breast cancer after gene-level correction for multiple testing. Rs11571215 in *PGR* (progesterone receptor) was the SNP most strongly associated with ER- breast cancer; this SNP is monomorphic in 1000 Genomes European ancestry samples. *Conclusion:* We identified six genes in steroid hormone pathways that contain common variants associated with breast cancer in AA women, including the PGR gene for ER- breast cancer. Although there were several associated SNPs, these may not be the causal variants. Future work should consider the entire gene.

2581W

A comparative analysis of network mutationburdens across 21 tumor types predicts newcandidate cancer genes. *H. Horn[1,2], M. S. Lawrence[2], J. Xin Hu[1,2,3,4], N. Ilic[2,5], Y. Shresta[2], E. Kim[2,5], A. Kamburov[2], A. Kashani[1,2], W. C. Hahn[2,5], J. S. Boehm[2], G. Getz[2], K. Lage[1,2].* 1) Department of Surgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA; 2) Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA; 3) Center for Biological Sequence Analysis, Technical University of Denmark, DK2800 Copenhagen, Denmark; 4) NNF Center for Protein Research, University of Copenhagen, DK2200 Copenhagen, Denmark; 5) Department of Medical Oncology, DanaFarber Cancer Institute, Boston, MA 02215, USA.

Heterogeneity across cancer makes it difficult to find driver genes with intermediate (2-20%) and low frequency (<2%) mutations, and we are likely missing entire classes of networks (or pathways) of biological and therapeutic value. Here, we quantify the extent to which cancer genes across 21 tumor types have an increased burden of mutations in their immediate gene network derived from functional genomics data. We formalize a classifier that accurately calculates the significance level of a gene's network mutation burden (NMB) and show it distinguishes both classic cancer genes and new driver genes in the majority of tested tumors. Our approach predicts 62 genome-wide significant candidate cancer genes - 44% of which point to new cancer biology. NMB identifies proportionally more (4x) low-frequency cancer genes than gene-based tests, and provides molecular clues in patients without established driver mutations. Our quantitative and comparative analysis of pan-cancer networks across 21 tumor types gives new insights into the biological and genetic architecture of cancers and illustrate an approach to augment discovery from existing cancer genomes that will become more powerful as datasets improve in the future.

2582T

Identifying pancreatic cancer susceptibility genes in the Czech Republic. *M. Janatova[1], M. Borecka[1], J. Soukupova[1], M. Vocka[2], P. Soucek[3], Z. Kleibl[1].* 1) Institute of Biochemistry and Experimetal Oncology, First Faculty of Medicine, Charles University in Prague, Prague 2, Czech Republic; 2) Department of Oncology, First Faculty of Medicine, Charles University in Prague and General University Hospital in Prague, Prague 2, Czech Republic; 3) The National Institute of Public Health, Prague 10, Czech Republic.

Pancreatic cancer (PC) is the sixth most frequent cancer and the fifth most frequent cause of death in the Czech Republic. It has one of the worst prognoses of any type of cancer. Early detection and subsequent surgical treatment combined with chemotherapy can improve 5-year survival up to 20%. Inherited genetic factors play an important role in PC risk but the genetic background of pancreatic cancer is highly heterogeneous. Recently, *PALB2* was identified as a breast and pancreatic susceptibility gene. It plays a role in DNA repair. The MRN (*MRE11*, *RAD50*, *NBN*) complex is also important for DNA damage response pathway. Heterozygous mutations in *NBN* predispose to various cancers, especially breast cancer. The Slavic common founder mutation c. 657del5 represents the majority of *NBN* mutations in the Central or Eastern Europe. We examined peripheral blood DNA samples of 196 unselected PC patients. All *PALB2* exons and flanking intron-exon boundaries were analyzed by direct sequencing. Exon 6 of the *NBN* gene was studied by HRM analysis. Identified mutations were analyzed in a group of 915 non-cancer control samples by HRM analysis. We identified three truncating mutations in the *PALB2* gene (3/196; 1. 5%). Only one mutation was found in control samples (1/915; 0. 1%; P=0. 02). We revealed six carriers of the *NBN* c. 657del5 mutation (6/196; 3. 1%), whereas two mutations were found in controls (2/915; 0. 2%; P=0. 0005225). Our results provide new evidence that the *PALB2* and *NBN* genes contribute to PC susceptibility in a significant proportion of unselected PC patients in the Czech Republic. Thus, both genes should be implemented in multi-gene panel testing. Genetic testing of PC patients is one of the approaches for identification of high-risk individuals. The mutation carriers with PC can undergo a specific therapy and their relatives can be offered appropriate screening methods for associated tumors. Nevertheless, further studies are needed to precisely estimate the risk of PC for the *PALB2* and *NBN* mutation carriers. Subsequent segregation analysis and penetrance determination will help to improve a clinical management for the unaffected carriers. Supported by grants: IGA MZCR NT 14006-3/2013, and NT13343-4/2012, PRVOUK-P27/LF1/1, and SVV-UK 260148/2015.

2583W

Germline genetic variants in men with prostate and other primary cancers. *A. Johnson[1], P. Pilie[1], E. Koeppe[2], M. Dayno[3], K. Zuhlke[1], L. Okoth[1], E. Stoffel[2], K. Cooney[1].* 1) Internal Med Hem/Onc, Univ Michigan, Ann Arbor, MI; 2) Internal Med GI, Univ Michigan, Ann Arbor, MI; 3) Medical School, Univ Michigan, Ann Arbor, MI.

Prostate cancer (PCa) has a significant heritable component and men with early age of diagnosis and/or positive family history of the disease have been studied to identify PCa susceptibility genes. Men with PCa and one or more additional primary cancers may also be more likely to harbor a germline susceptibility allele but these individuals have not been routinely studied in this context. Here we screen germline DNA from men with PCa as well as >1 additional primary cancer(s) for germline variants in a set of cancer-related genes. One hundred three men with PCa and >1 additional primary cancer were selected from the University of Michigan Prostate Cancer Genetics Project or Cancer Genetics Clinic. Each individual met at least one additional inclusion criteria: 1) early age of onset of first malignancy (age<55), 2) rare tumor type or atypical presentation of a common tumor including male breast cancer, sarcoma, pancreatic cancer, and/or 3) three or more primary malignancies. Average age of first cancer diagnosis was 50. 2 yrs and the average age of PCa diagnosis was 55. 2 yrs. The most common non-PCas were melanoma (34 cases), bladder (17 cases), kidney (13 cases), head and neck (10 cases), lung (10 cases), and colorectal (9 cases). Germline DNA from the 103 multiple primary subjects was sequenced using the GeneRead DNASeq Human Comprehensive Cancer Panel (Qiagen, Valencia, CA) and Illumina HiSeq 2000 (Illumina, San Diego, CA) platform. Within the coding regions of 160 cancer genes ~3500 variants were detected, including 5 inframe coding indels, and 524 missense SNPs. Protein truncating variants (2 nonsense SNPs, 7 frameshift indels) were identified in 9 of the 103 (8. 7%) subjects, including 1 individual with variants in both *BRCA2* and *MLH1*, 3 with *BRCA2*, 2 with *ATM,* and 1 each with *BRIP1*, *PALB2*, and *FGFR3*. All truncating variants were verified through Sanger sequencing and genotyped in family members with DNA available. Amongst truncating variant carriers, the most common non-PCas were thyroid (3 cases), bladder (2 cases) and kidney (2 cases). Taken together, the prevalence of protein truncating germline mutations in this cohort of men with multiple-primary cancers including PCa is 9%. Given the fact that most of these men do not fulfill clinical criteria for genetic testing based on their personal or family cancer history, larger studies should be performed to fully characterize the spectrum of germline mutations in these high risk men.

2584T

Variation in genes used for multigene expression profiling is associated with pretreatment cognitive function in postmenopausal women with breast cancer. *T. A. Koleck[1], C. M. Bender[1], S. M. Sereika[1,2], Y. P. Conley[1,3].* 1) School of Nursing, University of Pittsburgh, Pittsburgh, PA; 2) Graduate School of Public Health, Department of Biostatistics, Epidemiology, and Clinical Translational Science, University of Pittsburgh, Pittsburgh, PA; 3) Graduate School of Public Health, Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA.

Research suggests that variability in changes in cognitive function (CF) experienced by breast cancer (BC) survivors before they begin systemic adjuvant therapy may be related to cellular level BC characteristics. In order to test this hypothesis using a genetic association approach, we first identified biologically-plausible candidate genes from across five commercially available multigene expression profiles for BC, which are used clinically to characterize the biology of BC cells; genes utilized in three or more profiles were prioritized for investigation. Thus, the purpose of this study was to examine relationships between variation in genes used for multigene expression profiling and pretreatment CF in women with BC. Eight CF factors were assessed at pretreatment (i. e. , after primary surgery but prior to initiation of adjuvant therapy) in two cohorts of postmenopausal women diagnosed with early-stage BC prescribed chemotherapy plus anastrozole (n=56) or anastrozole alone (n=85) and a cohort of healthy, age- and education-matched controls (n=84) using a comprehensive neuropsychological test battery. Subjects were genotyped for 14 functional/tagging single nucleotide polymorphisms (SNPs) of the *CCNB1*, *CENPA*, *MELK*, *MYBL2*, and *ORC6* genes using DNA extracted from blood or saliva. Multiple linear regression was used to determine if the presence or absence of one or more variant alleles accounted for variability in CF factors. Analyses controlled for age, estimated intelligence and levels of depressive symptoms, anxiety, fatigue and pain. The strongest findings from these analyses relate to three functional SNPs in the 5' regulatory region of the *CCNB1* gene. The presence or absence of variant alleles (rs164390: TT+GT vs. GG; rs350099: CC+CT vs. TT; rs350104: TT vs. CC+CT) was found to be significantly (p<. 001 to p=. 088) associated with poorer pretreatment CF in BC survivors prescribed chemotherapy plus anastrozole or anastrozole alone compared to healthy controls. Furthermore, all *CCNB1* SNP by prescribed treatment interactions were associated with the same four CF factors - visual memory, visual working memory, verbal memory and executive function. Our results suggest that host variation in genes known to affect BC tumor aggressiveness may influence CF in women diagnosed with BC. In particular, differences in *CCNB1* expression may impact pretreatment cognitive performance in women diagnosed with BC, especially as it relates to memory and executive function.

2585W

Identification of rare genetic variation predisposing to lobular breast cancer. *C. Petridis[1,2], R. Roylance[3], E. J. Sawyer[1], M. A. Simpson[2].* 1) Research Oncology, Guy's Hospital, King's College London, London, United Kingdom; 2) Medical and Molecular Genetics, Guy's Hospital, King's College London, London, United Kingdom; 3) Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, London, United Kingdom.

Lobular breast cancer is the second most common type of breast cancer and accounts for about 15% of all cases. We have recently shown evidence of a distinct genetic aetiology compared to the common ductal breast cancer with both the identification of a lobular specific common variant susceptibility locus, and the identification of germline *CDH1* truncating mutations in ~8% of females with bilateral lobular disease. Here we aim to further evaluate the role of rare and low frequency variation in the genetic susceptibility to lobular breast cancer. To identify potential predisposition genes we have undertaken exome sequencing using an extreme phenotype study design. Affected females with early onset (<45), bilateral disease or family history of breast cancer have been selected from the GLACIER (Genetics of LobulAr Carcinoma In situ in EuRope) cohort (N=2535). The GLACIER study (MREC 06/Q1702/64) has ascertained patients from throughout the UK with the aim of understanding genetic predisposition to lobular breast cancer. 47 females met the inclusion criteria. Germline DNA was extracted from peripheral blood and exome sequencing performed using Agilent SureSelect exome capture and 2x100bp paired end reads on Illumina HiSeq2000. Two individuals had rare truncating alleles in *BRCA2* that have previously been reported as pathogenic (c. C5645A:p. S1882X, and c. C5655A:p. C1885X) and three harbour rare truncating alleles in *CDH1* (c. 2398delC:p. R800fs, c. G1942T:p. E648X, and c. 48+1G>A). We have undertaken a gene based case-control analysis using the data from the remaining 42 cases and 536 European female controls. Variants included in these analyses were selected on the basis of population frequency (MAF<1%), and predicted functionality (truncating and CADD>30). The analysis revealed 10 genes that are the site of rare putative functional variation in 2 or more affected individuals, and in which no equivalent variation is observed in the control group. To evaluate the relevance of these genes in lobular breast cancer disease population we are resequencing the coding regions (and intronic portions flanking splicing sites) in the entire GLACIER cohort (2535 cases and 2121 controls).

2586T

Association Between DNA Repair Gene Variants and Prostate Cancer Susceptibility using the iCOGS genotyping array. *E. Saunders[1], T. Dadaev[1], D. Leongamornlert[1], A. Amin Al Olama[2], S. Benlloch[2], K. Govindasami[1], K. Muir[3], D. Easton[2], R. Eeles[1], Z. Kote-Jarai[1], The PRACTICAL Consortium.* 1) Oncogenetics, Institute of Cancer Research, Sutton, Surrey, United Kingdom; 2) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Strangeways Research Laboratory, Cambridge, United Kingdom; 3) Institute of Population Health, University of Manchester, Manchester, United Kingdom.

Prostate cancer (PrCa) is the most frequently diagnosed male cancer in developed countries and amongst the leading causes of cancer related death. More than 100 common, low penetrance susceptibility variants have been identified by genome-wide association studies (GWAS), in addition to a small number of rare variants giving rise to greater risk; the majority of which are in DNA repair genes. We have previously genotyped 211,155 SNPs on a custom chip (iCOGS) in blood DNA from 21,780 PrCa cases and 21,727 controls of European ancestry from the PRACTICAL Consortium. This dataset has been recently imputed to the 1000 genomes phase 3 reference panel using SHAPEIT and IMPUTE2. In a literature review, we identified 179 genes with a primary function in DNA repair. SNPs within a 20kb flanking region of our gene set according to GENCODE release 19 were selected from the imputed iCOGS dataset. After removal of imputed variants with an IMPUTE2 INFO<0. 3, 98,255 DNA repair gene SNPs were available for single variant and gene-level association analyses. Analyses were adjusted for the first eight principal components and study groups, significance levels were set using the Bonferroni correction (single = P<5×10-7, gene = P<2. 7×10-4). Single variant analysis for association of DNA repair gene variants with PrCa identified only the previously reported association with *RAD51B* (Chr14q24; P = 1. 29x10-10). Several other genes showed non-significant association peaks, however no other variants were within one order of magnitude of the Bonferroni correction threshold. Gene-level analysis identified a novel significant association with the *MSH5* gene (Chr6p21; P = 1. 68x10-4) using the SKAT-C test. *MSH5* is involved in DNA mismatch repair and is a reported lung cancer susceptibility locus. This potential PrCa susceptibility locus warrants additional follow-up in independent datasets. We also investigated association between DNA repair gene variants and phenotypic characteristics of PrCa; to examine whether any of these variants could confer risk towards developing aggressive disease rather than increasing the risk of developing PrCa *per se*. No significant associations were observed at either the variant or gene levels; however these analyses were constrained by small numbers between the comparison groups.

2587W

Biochemical Pathways Involved in Betel Nuts Induced Oral Squamous Cell Carcinoma. *Y. S Shih[1], A. V Vu[3], T. L Liang[3], A. Z Zhang[4], C. C Chen[2], S. S Saharti[2], Y. E Elshimali[3].* 1) La Sierra University, Riverside, CA; 2) UCI Irvine, CA 92697; 3) UCLA Los Angeles, CA 90095; 4) Diamond Bar High School 21400 Pathfinder Rd, Diamond Bar, CA 91765.

Areca nut and betel quid chewing are traditional habits practiced throughout East and Southeast Asia with serious health consequences. Long term use has been associated with oral submucosal fibrosis, pre-cancerous oral lesion, and squamous cell carcinoma of the oral cavity and esophagus. Betel nuts contain several known carcinogenic compounds in high concentration,including hydroxychavicol, safrole and arecoline. Arecoline, a nicotinic acid-based alkaloid is the primary active ingredient in betel nuts responsible for CNS effectsand is a powerful carcinogen. Areca nut extract (ANE), when added to cultured human oral mucosal epithelium, can result in micronuclei formation, DNA damage, and apoptosis of the cells. One study concluded that over 17% of leukoplakia eventually developed into oral squamous cell carcinoma,suggesting the premalignant component of leukoplakia. Different studies have suggested that genetic components of oral squamous cell carcinoma with certain SNP genotypes, predispose one to increased-susceptibility and/or resistance towards progression to malignancy. Due to the complex interactions between environmental and genetic factors which result in the pathogenesis of betel nut induced oral squamous cell carcinoma, we use multiple approaches to analyze potentially relevant biochemical pathways in the pathogenesis. First, we conducted a comprehensive literature review of relevant genes implicated in pathogenesis of betel nut induced squamous cell carcinoma, regardless of the types of studies involved. Using the KEGG Pathway Database, we then identified relevant biochemical pathways where individual genes are involved, especially pathways that appeared more than 3 times in at least 2 independent studies. Overall, 5 biochemical pathways were identified including the MAPK signaling pathway, ErB signaling pathway, PI3K-Alk signaling pathway, AMPK signaling pathway, and VEGF signaling pathways. Currently, we are designing a targeted capture system for genes on these 5 pathways. Using 5 oral squamous cell carcinoma tissues and their associated normal controls from paraffin embedded tissue blocks, we will use the custom capture system to perform RNA sequencing on the Illumina MiSeq sequencer to identify potential mutations and fusion transcripts. Identifying these mutations may provide insight for identification of possible driver mutations.

2588T

How big of a role does *TP53* play in *BRCA1/2* negative familial breast cancer? Data from the SIMPLEXO consortium. *T. P. Slavin[1], K. N. Maxwell[2], S. Hart[3], V. Joseph[4], K. Schrader[5], L. Guidugli[3], T. Thomas[4], R. Moore[3], C. Hu[3], B. Wubbenhorst[6], S. M. Domchek[2], M. E. Robson[7], P. Radice[8], P. Peterlongo[8], J. Ford[9], J. Garber[10], C. Szabo[11], S. Neuhausen[12], K. Offit[7], K. L. Nathanson[6], F. J. Couch[3], J. N. Weitzel[1,12], SIMPLEXO Consortium.* 1) Department of Medical Oncology, Div of Clinical Cancer Genetics, City of Hope, Duarte, CA; 2) Department of Medicine, Division of Hematology-Oncology, University of Pennsylvania, Philadelphia, PA; 3) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 4) Clinical Genetics Research Lab, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 5) Hereditary Cancer Program, British Columbia Cancer Agency, Vancouver, BC; 6) Department of Medicine Division of Translational Medicine and Genetics, University of Pennsylvania, PA; 7) Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 8) IFOM, the FIRC Institute of Molecular Oncology, and Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; 9) Division of Oncology, Stanford University School of Medicine, Stanford, CA; 10) Center for Cancer Genetics and Prevention, Dana Farber Cancer Institute, Boston, MA; 11) National Institutes of Health, Bethesda, MD; 12) Beckman Research Institute of City of Hope, Duarte, CA.

Current known common, low risk polymorphisms and rare, moderate, and high risk alleles in cancer predisposing genes account for approximately half of the hereditary relative risk of breast cancer. Identification of individuals at high hereditary risk of breast cancer may lead to precision screening, prevention, and therapeutic strategies for patients and their families. To identify novel genes that confer an inherited risk of breast cancer, whole exome sequencing was undertaken in 333 *BRCA1/2* negative independent families with at least one primary breast and ovarian cancer or with a proband and at least two first to third degree relatives with breast cancer under age 70. This approach led to a validation set of 564 candidate genes that were evaluated using next-generation sequencing in 2331 *BRCA1/2* negative individuals with familial breast cancer. Analysis of the validation data revealed multiple individuals with rare variants in the Li-Fraumeni (LFS) associated gene *TP53*. A custom bioinformatics pipeline and the ACMG variant classification guidelines were used to identify and classify the variants. Twenty-nine variants found at less than 0. 1% minor allele frequency in the ExAC and EVS databases were identified in 33 individuals (1. 4%). Of these, 18 individuals were found to have 18 likely pathogenic/pathogenic mutations (0. 77% of the overall population), of which four were truncating and 14 were missense mutations in the DNA binding domain or tetramerization domain of p53. Ten of the 14 missense mutations were identified as pathogenic or likely pathogenic in Clinvar. Three missense mutations were classified as likely pathogenic due to *in silico* models, presence in the DNA binding domain and transactivation functional data from the TP53 IARC database. Finally one missense mutation was classified as likely pathogenic based on literature review. Of the 11 remaining rare variants, nine were classified as variants of uncertain significance (VUS) and two were classified as likely benign polymorphisms. Overall, these results show that *TP53* is a rare (0. 77% herein) but important contributor to familial breast cancer in *BRCA1/2* negative individuals.

2589W

Fine-mapping of the differentiated thyroid cancer risk susceptibility loci at 9q22 and 14q13 in populations from Metropolitan France and from New Caledonia. *c. Tcheandjieu[1,2], F. Lesueur[2,3], M. Sanchez[1], T. Truong[1,2], P. Guenel[1,2].* 1) INSERM, Villejuif, France; 2) University of Paris SUD, Le Kremlin Bicetre, France; 3) Institut Curie, Paris, France.

Context: Recently, GWAS have identified several susceptibility loci in differentiated thyroid carcinoma (DTC); the most robust associations were reported at loci 9q22 (rs965513 and rs1867277) and 14q13 (rs942289 and rs116909734). It remains however unclear whether these SNPs represent independent signals of association or if there are other independent signals within each locus. We aimed to replicate the 4 previously identified SNPs at 9q22 and 14q13 in a population of European descent in Metropolitan France and in a very high risk population of Melanesians in New Caledonia (NC). We also sought to identify novel DTC susceptibility SNPs and to find out whether SNPs at 9q22 and 14q13 can explain the differences in incidence rate reported between the 2 populations **Methods**: We included subjects from 2 case-control studies, conducted in Metropolitan France and in NC. We genotyped and imputed 81 and 561 SNPs respectively at loci 9q22 and 14q13 in 625 cases and 776 controls of European descent and 244 cases and 189 controls of Melanesian ethnicity. Odds ratio (OR) were performed using logistic regression assuming a log-additive model. We conducted conditional and haplotype analysis to identify independent signals and used bioinformatics tools to predict function of SNPs. **Results**: In Europeans, we replicated the association between previously reported GWAS SNPs and DTC at 9q22 (rs965513 [OR=1. 52, p=6. 10-6] and rs1867277 [OR=1. 38, p=3. 10-4] and at 14q13 (rs944289 [OR=0. 41, p=5. 10-4] and rs116909374 [OR=1. 97, p=3. 10-3]). In Melanesian, only rs944289 (OR=0. 74, p=1. 10-3) was replicated but magnitude of ORs for rs965513 and rs1867277 were similar to those observed in Europeans. At locus 9q22, no other SNP was significantly associated with DTC after conditioning the analysis on rs965513 and functional analysis suggestedrs10759944 in strong LD with rs965513 as a potential functional SNP. At locus 14q13 we reported a significantly high risk haplotype rs944289-rs116909374-rs999460 (OR=0. 51, p=3. 10-6) in Europeans. In Melanesians, the most significant signal was observed for an independent SNP rs1755774 (OR=0. 57, p=2. 10-5) which is in LD with rs2787423, a potential functional SNP. **Conclusion**: We confirmed the previously reported associations of DTC with SNPs at loci 9q22 and 14q13 and identified new interesting susceptibility SNPs in both ethnic groups. SNPs at both loci cannot account for the difference in incidence rates observed between Europeans and Melanesians.

2590T

Whole-exome sequencing in Taiwanese nasopharyngeal carcinoma (NPC) families to detect NPC susceptibility genes and pathways. *G. Yu[1], W. Hsu[2], M. Yeager[1], CGR. Cancer Genomics Research Lab[1], C. Wang[2], P. Lou[2], S. Diehl[3], C. Chen[4], A. Hildesheim[1], A. Goldstein[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) National Taiwan University, Taiwan; 3) Rutgers School of Dental Medicine, Newark, NJ; 4) Academia Sinica, Taiwan.

NPC is a cancer that is known to be caused by infection with Epstein-Barr virus (EBV), a ubiquitous DNA virus that establishes lifelong latency in infected individuals. While >90% of the world population is infected with EBV during childhood or early adult life, only a very small fraction of infected individuals develop NPC. It is believed that, in addition to EBV infection and other environmental factors, genetic predisposition plays an important role in the pathogenesis of NPC. However, little is known about the genetic basisi of NPC. In this study, we sequenced 259 exomes from 99 NPC families with 2-4 cases by Nimblegen v2. 0 and v3. 0 exome capture array and the Illumina HiSeq2000 sequencer. Reads were aligned using Novoalign and variants were called using GATK software. We prioritized 9 genes with EBV/NPC related function (literature-based) and rare deleterious variants co-segregating in >=2 families (frequency <0. 5% in dbSNP, 1000 genomes or NHLBI Exome Sequencing Project exome variant server; no segmentally duplicated genes/variants; not in a repeat region; predicted as deleterious by Combined Annotation Dependent Depletion (CADD) score greater than 15; at least 10 reads and fitting a dominant segregation model in families with multiple NPC patients sequenced). In addition, we performed pathway enrichment analyses for 611 prioritized rare deleterious variants by Ingenuity Pathway Analysis and identified significant enrichment of Notch signaling pathway genes in NPC cases (P=0. 001). Notch signaling pathway is critical for cell survival, growth and metastasis. In total, 13 genes (including NOTCH1 et. al) were identified as candidate NPC susceptibility genes and will be evaluated further in a large NPC case-control study (~3000 samples) from the same Taiwanese population that generated the families.

2591W

Germline and Somatic *SDHx* Alterations in Apparently Sporadic Differentiated Thyroid Cancer. *Y. Ni[1,9,10], S. Seballos[1], S. Ganapathi[1], D. Gurin[1], B. Fletcher[1], J. Ngeow[1,4], R. Nagy[5,7], R. Kloos[6,7], M. Ringel[6,7], T. LaFramboise[1,8,10], C. Eng[1,2,3,8,10].* 1) Genomic Medicine Institute, Lerner Research Institute, Cleveland Clinic, Clevland, Ohio; 2) Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio; 3) Stanley Shalom Zielony Nursing Institute, Cleveland Clinic, Cleveland, Ohio; 4) Division of Medical Oncology, National Cancer Center, Singapore; 5) Divisions of Human Genetics, Department of Medicine, The Ohio State University, Columbus, Ohio; 6) Divisions of Endocrinology and Metabolism, Department of Medicine, The Ohio State University, Columbus, Ohio; 7) Comprehensive Cancer Center, Arthur G. James Cancer Hospital and Richard G. Solove Research Institute, The Ohio State University, Columbus, Ohio; 8) Department of Genetics and Genome Sciences, Case Western Reserve University School of Medicine, Cleveland, Ohio; 9) Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, Ohio; 10) CASE Comprehensive Cancer Center, Case Western Reserve University, Cleveland, Ohio.

Along with breast and endometrial cancers, thyroid cancer is a major component cancer in Cowden syndrome (CS). Germline variants in *SDHB/C/D* (*SDHx*) genes account for subsets of CS/CS-like cases, conferring a higher risk of breast and thyroid cancers over those with only germline *PTEN* mutations. To investigate whether *SDHx* alterations at both germline and somatic levels occur in apparently sporadic breast cancer and differentiated thyroid cancer (DTC), we analyzed *SDHx* genes in the following four groups: i) 48 individuals with sporadic invasive breast adenocarcinoma for germline mutation; ii) 48 (expanded to 241) DTC for germline mutation; iii) 37 pairs DTC tumor-normal tissues for germline and somatic mutation and mRNA expression levels; and iv) 476 the Cancer Genome Atlas thyroid carcinoma dataset for validation. No germline *SDHx* variant was found in a pilot series of 48 breast cancer cases. As germline *SDHx* variants were found in our pilot of 48 thyroid cancer cases, we expanded to three series of DTC comprising a total 754 cases, and found 48 (6%) with germline *SDHx* variants (P<0. 001 compared with 0/350 controls). In 513 tumors, we found 27 (5%) with large somatic duplications within chromosome 1 encompassing *SDHC*. Both papillary and follicular thyroid tumors showed consistent loss of *SDHC/D* gene expression (P<0. 001), which is associated with earlier disease-onset and higher pathological-TNM stage. Therefore, we conclude that both germline and somatic *SDHx* mutations/variants occur in sporadic DTC but are very rare in sporadic breast cancer, and overall loss of *SDHx* gene expression is a signature of DTC.

2592T

Identification of Retrotransposon Insertion Mutations in Hereditary Cancer. *Y. Qian, D. Mancini-DiNardo, T. Judkins, H. C. Cox, C. Daniels, J. Holladay, M. Ryder, B. Coffee, K. R. Bowles, B. B. Roa.* Myriad Genetics, Inc. , Salt Lake City, UT.

Background: Traditional assays or techniques may underestimate the presence of certain types of large rearrangement mutations. Retroelements (RE) such as Alu and L1 elements are located ubiquitously throughout the human genome, and may cause disease if they result in genetic instability or direct disruption of gene function by inserting into a critical domain. **Methods:** We designed a targeted testing strategy that allows us to clearly characterize these large insertions in patients for whom hereditary cancer genetic testing was performed. **Results:** In this study, we identified 17 retroelement insertions in the exonic regions of 7 genes, which are *APC, ATM, BRCA1, BRCA2, MLH1, MSH2* and *PALB2*. As expected, the personal and family history of cancer was similar to those patients carrying other pathogenic truncating mutations. Strikingly, 9 of the 15 Alu insertions were identified in *BRCA2* exonic regions, four of which are located in exon 11, the largest exon in *BRCA2*, suggesting that this locus may be a hotspot of retroposon insertions of the genome. Interestingly, a few Alu insertions in *BRCA2* are enriched in certain ethnic populations. The individuals with c. 5007_5008insAlu (5235insAlu) share the same haplotype and were only identified in patients with recorded Latino/Caribbean ancestry. Similarly, 19 out of 29 patients carrying the c. 3407_3408insAlu (3635insAlu) mutation were recorded as having African ancestry. For the remaining 10 patients, 7 were of unknown ethnicity and three patients reported Latino or Native American ancestry; however, all the patients shared a common haplotype for *BRCA2*. This suggests that the two Alu insertions might be founder mutations in Latino/Caribbean and African populations, respectively. **Conclusions**: Current technologies are limited in their ability to identify RE insertions. However, when the presence of a large insertion is suspected, targeted follow-up studies are imperative to precisely characterize and classify such large rearrangements.

2593W

Integrated analysis of case-control and tumor-normal pair whole-exome data in breast cancer. *F. Hu, Y. Yu, S. Sivakumar, H. Hu, P. Scheet, X. Wu, C. Huff.* Epidemiology, MD Anderson Cancer Center, Houston, TX.

Genetic susceptibility has a well-established role in the etiology of breast cancer, but the contribution of rare genetic variants to breast cancer risk is underexplored. To address this gap, we conducted an analysis of whole-exome sequencing data to identify and characterize rare coding variants contributing to breast cancer susceptibility. We used VAAST 2. 1 to perform gene-based association tests from whole-exome sequencing data in 783 European ancestry breast cancer cases from TCGA and 1,451 European ancestry population controls from various sources. We constrained the analysis to variants with population minor allele frequencies of less than 1%. Our results replicated four well-established breast cancer-gene associations at p < 0. 05, *BRCA1* (p = 0. 003), *BRCA2* (p = 0. 01), *PALB2* (p = 0. 04), and *CHEK2* (p = 0. 04). In addition, we observed nominally significant associations for two genes previously identified in breast cancer GWAS, *RECQL5* (p = 0. 04), and *LGR6* (p = 0. 03). We also observed evidence of an association with *FANCC* (p = 0. 02), which has been previously implicated in breast cancer susceptibility. Our single marker analyses also identified two nominally significant variants in the above genes, G998E in *PALB2* (OR = 1. 86, p < 0. 005) and S26F in *FANCC* (OR = 3. 47, p = 0. 001). To further characterize putative germline susceptibility variants, we analyzed tumor-normal pair data from the 783 cases to test for evidence of somatic-germline interaction. We detected loss of heterozygosity events using hapLOH and tested for evidence of one-hit and two-hit models of carcinogenesis using SGI, reporting the results for each candidate gene as well as the top signals in a genome-wide analysis. Our results provide further support for the role of rare exonic variation in breast cancer susceptibility.

2594T

Identification and Functional Characterization of ESR1 Mutations in Breast Cancer. *A. Bahreini[1,2], P. Wang[2], S. Puhalla[4], R. Gyanchandani[3], T. Ambros[4], R. Hartmaier[3], B. Kurland[5], P. Lucas[6], H. Bittar[6], R. Hamilton[7], A. Matthews[6], J. Leone[8], N. Davidson[6], K. Weiss[6], R. Watters[3], M. Nikiforova[6], A. Stern[9], A. Brufsky[6], A. Lee[1,2,4], S. Oesterreich[2,3].* 1) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Magee-Women Research Institute, University of Pittsburgh Cancer Institute, Pittsburgh, PA; 3) Department of Pharmacology & Chemical Biology, University of Pittsburgh School of Medicine, Pittsburgh, PA; 4) University of Pittsburgh Medical Center, Women's Cancer Program at Magee-Womens Hospital of UMPC, Pittsburgh, PA; 5) Biostatistics, University of Pittsburgh Cancer Institute, Pittsburgh, PA; 6) University of Pittsburgh School of Medicine, Pittsburgh, PA; 7) Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA; 8) University of Iowa, Iowa city, IA; 9) Department of Computational & Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA.

Mutations in the estrogen receptor (*ESR1*) occur at very low rates in primary breast tumors, but are frequently observed in endocrine resistant metastatic breast cancer. Preliminary studies indicate that mutant *ESR1* is partially resistant to current endocrine therapies. To understand the role of *ESR1* mutations in tumor progression, we employed digital droplet PCR to search for six previously reported *ESR1* mutations in primary tumors (n=43), bone metastases (n=12), brain metastases (n=38), and circulating free (cf) DNA (n=29) from 121 patients. *ESR1* D538G, Y537S, and Y537C were detected in 3/43 primary specimens, 3/38 brain specimens, 1/12 bone specimens, and 7/29 cfDNA, however, *ESR1* S463P, K303R, or Y537N were not observed. Multiple mutations were found in two patients. *ESR1* mutations identified in multiple blood draws were also detected in matched metastatic specimens and the frequencies correlated with disease progression, suggesting the potential value of *ESR1* mutations as a biomarker to monitor response to systemic endocrine therapy. The most frequently detected mutation (*ESR1* D538G) was also introduced into cell line genomic DNA using CRISPR/CAS technology to characterize phenotypic features of the ER mutant. Co-culturing D538G mutant and parental wild-type cells confirms the evolution of ESR1 mutants under anti-ER treatments. Our study shows that *ESR1* mutations are a measurable biomarker and a potential mechanism for endocrine resistance in breast cancer.

2595W

Targeted germline DNA sequence in men with family history of Prostate Cancer (PROFILE STUDY). *C. Cieza-Borrella[1], C. Mikropoulos[1], D. A. Leongamornlert[1], S. Saya[1], E. Bancroft[1], E. J. Saunders[1], T. Dadaev[1], E. Castro[1,2], Z. Kote-Jarai[1], R. A. Eeles[1,2], The PROFILE Study Steering Committee.* 1) The Institute of Cancer Research, 123 Old Brompton Road, London, UK; 2) The Royal Marsden NHS Foundation Trust, Fulham Road, London, UK.

Prostate Cancer (PrCa) is one of the most heritable solid tumours and mutations in DNA repair genes have been shown to predispose to its development. Genetic testing for *BRCA1* and *BRCA2* mutations is widely available for breast cancer, but with the introduction of next generation sequencing (NGS) it is feasible to screen for mutations in a panel of candidate genes and correlate these with PrCa risk. In this study, we included 86 men enrolled on the PROFILE PILOT study, a prospective PrCa screening study using prostate biopsy in asymptomatic men aged 40-69 with significant PrCa family history (FH). Blood samples were taken for PSA measurement and DNA extraction. gDNA samples were enriched for 22 tumour suppressor genes (the BROCA panel) using a custom Agilent SureSelect bait library. Sequencing was performed on the Illumina Miseq platform and reads aligned to the '1000 Genomes project' Phase 1. Sanger sequencing was used to confirm identified sequence variants in patients and their relatives where DNA was available. In total we identified 105 missense mutations, 54 synonymous variants, two nonsense mutations, one splice site mutation and one in-frame deletion with a population MAF of <5%. Within these, two likely pathogenic (Class 4) mutations were found: a splice-site variant in *BRCA1* (rs80358033) and a non-synonymous missense *BRCA2* mutation (rs80359104). The *BRCA1* mutation carrier refused prostate biopsy however the *BRCA2* mutation carrier was diagnosed with PrCa (GS 3+4). A novel *BARD1* stop-codon mutation was identified in a man with a benign biopsy outcome. This cohort will undergo 10-years follow-up to determine future PrCa incidence and in total will recruit 350 men with PrCa FH. This study shows the potential clinical utility of NGS in men with PrCa FH for: (i) assessing risk in patients and their family members and, (ii) potential future PrCa treatment options.

2596T

LncRNA expression influence DNA damage response in acute lymphoblastic leukemia. *R. Gioia[1, 3], M. Ouimet[1, 3], M. Lajoie[1], S. Drouin[1], C. Richer[1], R. Vidal[1], D. Sinnett[1, 2].* 1) Division of Hematology-Oncology, Research Center, Sainte-Justine Hospital, 3175 Chemin de la Cote-Sainte-Catherine, Montreal, QC, H3T 1C5, Canada; 2) Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, QC, H3T 1C5, Canada; 3) These author collaborated equally to this work.

Acute lymphoblastic leukemia (ALL) represents 75% of cases of childhood leukemia, with approximately 3,000 children in the US and 5,000 in Europe diagnosed annually. Most cases involve chromosomal and genetic abnormalities, occurring spontaneously in genes that are important regulator of proliferation and cell death. According to current estimates, long non-coding RNAs (lncRNAs) account for up to 75% of the transcribed genome. However, our knowledge about their functions is scarce. We have initiated a study to identify and characterize lncRNAs involved in childhood ALL. We identified more than 800 lncRNAs whose expression is significantly up- or downregulated in patient samples as compared to CD19+ primary B-cells. We have characterized the role of three downregulated lncRNAs in the development of the disease by overexpressing them in pre-B leukemic cell lines. Two of these, located very close to each other in the genome, increased apoptosis levels after treatment with camptothecin, a DNA-damaging agent, without a concomitant increase in cell death, suggesting a role in DNA damage response. Supporting this, phospho-H2A. X levels in these cell lines are significantly increased compared to controls when exposed to camptothecin. Finally, cells lines in which the third lncRNA was overexpressed show significantly lower proliferation rate. We are currently further investigating the mechanistic underpinning of these effects to identify pathways in which these lncRNAs are implicated. .

2597W

Functional characterization of melanoma risk-associated locus on chromosome band 1q21. 3. *MA. Kovacs[1], J. Choi[1], T. Zhang[1], M. Xu[1], L. Aoude[2], M. Gartside[2], D. Gorkin[3], S. Loftus[4], W. Pavin[4], NK. Hayward[2], KM. Brown[1].* 1) Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA; 2) QIMR Berghofer Medical Research Institute, Herston, Brisbane, QLD, Australia; 3) Ludwig Institute for Cancer Research, 9500 Gilman Drive, La Jolla, CA 92093, USA; 4) Genetic Disease Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

Cutaneous melanoma (CM) is the most aggressive form of skin cancer, accounting for nearly 50,000 deaths worldwide each year. Understanding CM genetic predisposition may shed light on pathways involved in CM pathophysiology and could reveal novel drug targets or biomarkers that may facilitate early detection. A recent genome-wide association study by MacGregor et al. identified a locus on chromosome band 1q21. 3 that significantly correlates with melanoma risk (OR = 0. 89). Fine mapping and conditional analyses have revealed that the association signal is most likely explained by a single genetic variant. The purpose of our study has been to pinpoint the functional variants in the 1q21. 3 locus, which are expected to be in tight LD with the region's lead SNP rs3768013, and to determine the CM-predisposing gene(s) regulated by these variants. To identify functional variants, we developed a bioinformatics tool using R that could annotate variants according to their potential to function as *cis*-regulatory variants. We took into account signature enhancer histone marks and open chromatin regions using histone ChIP-seq and DNaseI hypersensitivity sequencing data available from ENCODE and the Roadmap Epigenomics Project. Our annotation revealed 24 variants tightly linked to rs3768013 that have a high probability of functioning as *cis*-regulatory variants. We are using allele-specific protein binding assays to determine which of these variants could play a functional role in CM susceptibility. To elucidate the biological mechanism by which these variants mediate disease risk, we performed eQTL analysis in melanoma cell lines, as well as in tumor samples from The Cancer Genome Atlas (TCGA) skin melanoma project. Our analysis revealed that increased expression of *GOLPH3L* significantly correlates with the CM risk allele. Because *GOLPH3L* is a paralog of *GOLPH3*, a potent oncogene targeted for amplification in a number of cancers (Scott et al.), we evaluated whether ectopic *GOLPH3L* overexpression could transform human primary melanocytes in conjunction with oncogenic *BRAFV600E*. In sum, we have employed a generalizable system of variant ranking to parse through the vast set of risk-associated variants in the 1q21. 3 GWAS locus to identify a more limited set of variants that may be functional. We have demonstrated the possibility that these variants might regulate *GOLPH3L* expression in an allele-specific manner, indicating a potential novel oncogene involved in CM etiology.

2598T

Exome sequencing for detection of pathogenic rare variants in familial papillary thyroid cancer. *A. L. Camilleri[1], S. G Wilson[1,2,3], J. P. Walsh[1,3], L. C. Ward[3], J. Goldblatt[4].* 1) School of Medicine and Pharmacology, University of Western Australia, Crawley, Australia; 2) Department of Twin Research and Genetic Epidemiology, King's College London, London, UK; 3) Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands, Australia; 4) School of Paediatrics and Child Health, University of Western Australia, Crawley, Australia.

Thyroid cancer is the most common endocrine malignancy, and the reported incidence is increasing. Papillary thyroid cancer (PTC), accounts for 80% of thyroid malignancies. Most PTCs occur in younger adults and therefore occur during peak employment and reproductive years. Despite low mortality rates, local recurrence occurs in up to 20% of patients. For this group of patients, re-operative surgery can cause significant morbidity and apart from radioactive iodine ablation, there is no adjuvant therapy available. Familial PTC (FPTC) occurs in several multi-organ cancer syndromes, but also occurs in isolation with features of a Mendelian autosomal dominant disorder with incomplete penetrance; the gene or genes involved are largely unknown. Approximately 3-7% of all PTC cases are thought to be familial. We performed an exome sequencing study on individuals from 13 families with FPTC. Individuals were recruited from an endocrinology practice and study subjects provided a family cancer history which was confirmed directly by pathology or otherwise from patient records. Whole exome sequencing (WES) was performed on DNA extracted from blood collected for 27 individuals using the Illumina TruSeq Exome Enrichment Kit for exome capture and the Illumina HiSeq2000 platform for sequencing. Results were annotated using ANNOVAR and were analysed using VarSifter which allows easy visualisation and manipulation of the data, and VAAST which ranks the genes from most damaging to least damaging based on the composite likelihood ratio test (CLRT). Exome sequencing yielded an average 76. 6 Mb of DNA target sequence per sample with mean coverage achieved across all targets (20,794 genes across the genome) of 76. 8x. Analysis of the data from one family revealed mutations in the following genes: *ACIN1, LRRC49, CSPG4* and *CHEK2*. One gene variant was identified that was of particular interest in *CHEK2*, located on chromosome 22. The *CHEK2* gene in one FPTC family contained a 1 bp deletion of a G nucleotide at position 29091857 on chromosome 22, causing a frameshift mutation. In summary, our strategy for rare variant detection led to the identification of a potentially damaging mutation which may contribute to FPTC risk in this family. Whilst it has been suggested that this mutation contributes to a risk of PTC, this is the first report of a *CHEK2* truncating mutation in an FPTC family and studies are ongoing to understand the biologic implications.

2599W

Interleukin 7 gene region is associated with mesothelioma survival in Australian males. *G. Cadby[1], F. Brims[2], A. Reid[3], I. Dick[4], C. Robinson[4], P. Melton[1], E. Moses[1], B. Robinson[2,4,5], B. Musk[2,4], J. Creaney[4,5]*. 1) Centre for Genetic Origins of Health and Disease, University of Western Australia, Crawley, WA; 2) Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Nedlands, WA; 3) Epidemiology and Biostatistics, School of Public Health, Curtin University, Bentley, WA; 4) National Centre for Asbestos Related Diseases, School of Medicine and Pharmacology, University of Western Australia, Crawley, WA; 5) The Australian Mesothelioma Tissue Bank, Sir Charles Gairdner Hospital, Nedlands, WA.

Malignant mesothelioma (MM) is an incurable malignant tumour of the pleura or peritoneum with increasing incidence worldwide. There is currently no known effective treatment for MM which is uniformly fatal with a median survival time of ~12 months. A role for genetic variation in MM survival has been observed in various murine genetic backgrounds; however no published studies have examined the germline genetic contribution to human MM survival. The aim of this study was to perform a genome-wide association study of MM survival. Clinical data and DNA were obtained from 371 men with a confirmed diagnosis of MM who attended Sir Charles Gairdner Hospital in Perth, Western Australia. Genome-wide genotyping was performed on the Illumina 660W Quad Array and approximately 500,000 single nucleotide polymorphisms (SNPs) were available for testing after quality control checks. Survival analysis was performed using Cox regression assuming proportional hazards. Age at diagnosis and the first three principal components were included as covariates in the analysis. Median survival was 13 months for those who had died (n=361) and 52 months for those with censored observations (n=10). Average age at diagnosis was 67. 3 years. No single SNP achieved formal genome-wide significance using a Bonferroni correction for multiple testing ($p < 9. 6x10-8$). A region including the Interleukin 7 (*IL7*) gene showed some evidence for association with MM survival (*IL7* rs6993386: hazard ratio=1. 56; p=1. 46x10-7). The IL7 protein (encoded by the *IL7* gene) functions primarily as a growth and anti-apoptotic factor for B and T cells, and has previously been implicated in the progression of haematopoietic and solid malignancies. Further analyses are planned, including replication analysis in independent cohorts of Italian and American MM cases.

2600T

Identifying inherited genetic variation associated with molecular phenotypes in colon cancer. *D. Chang[1], T. Bhangale[1], R. Graham[1], S. Bezieau[2], H. Brenner[3], A. T. Chan[4], J. Chang-Claude[5], S. J. Chanock[6], T. A. Harrison[7], M. Hoffmeister[3], L. Hsu[7], T. J. Hudson[8], P. A. Newcomb[7], J. D. Potter[7], M. L. Slattery[9], E. White[7], U. Peters[7], R. Gentleman[10].* 1) Genentech, South San Francisco, CA USA; 2) Service de Génétique Médicale, CHU Nantes, Nantes, France; 3) Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany; 4) Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 5) Division of Cancer Epidemiology, German Cancer Research Center, Heidelberg, Germany; 6) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA; 7) Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 8) Ontario Institute for Cancer Research, Toronto, Ontario, Canada; 9) Department of Internal Medicine, The University of Utah School of Medicine, Salt Lake City, UT, USA; 10) 23andMe, Mountain View, CA, USA.

Molecular characterization of colon cancer tumors includes assaying microsatellite instability (MSI), methylation (e. g. CpG island methylator phenotype, CIMP), as well as somatic mutations in the tumor. In a parallel effort, GWAS have identified inherited genetic variants involved in colorectal cancer (CRC) risk. The goal of this study is to assay whether variants associated with CRC risk are also associated with tumor-related phenotypes. To carry out a GWAS on tumor phenotypes, we used an array of tumor and adjacent non-tumor colon data available from The Cancer Genome Atlas (TCGA) project. We carried out a GWAS for each of the four binary tumor phenotypes relevant to CRC: CIMP status, MSI status, and presence of somatic mutations in *APC*, and *TP53*. After removing samples of non-European ancestry, 375 individuals remained. Each of the tumor phenotypes was regressed on genetic variants to test for association between inherited genetic variation and tumor-related phenotypes. As phenotype data across the four phenotypes were not available for all individuals, there were 301-375 samples in each GWAS depending on the phenotype. None of the 66 SNPs associated with CRC in the NHGRI GWAS catalog and for which we had data on, were significantly associated with a tumor phenotype. We next applied two complementary gene-association methods, MixMAP and VEGAS that can increase power by reducing the number of tests and leveraging information across multiple SNPs within a region to detect gene-level association. We tested for association across 20,873 genes with tumor phenotypes in the TCGA dataset and for association with CRC risk in the GECCO dataset (a GWAS of over 20,000 samples). Seventeen genes were associated with CRC risk in the GECCO dataset at a significance threshold adjusted for 20,873 tests ($P<2. 4x10-6$ across the 17 genes). Testing these 17 genes for association in the TCGA data, we found that *CLDN4* and *WBSCR27* are associated with *TP53* mutation status (PMixMAP $< 2. 9x10-5$ for both genes); and *FOLH1B* is associated with MSI status (PMixMAP=3. 07x10-4). All are significant at a threshold adjusted for 17 tests. In conclusion, incorporation of tumor molecular data with inherited genetic variation can provide relevant CRC subtypes in which variants associated with CRC risk may play a role. Larger sample sizes are needed to confirm these findings.

2601W

Using genetic variants and classical epidemiology risk factors in a risk prediction model for lung cancer among female non-smokers. *L. Chien[1], I. Chang[1,2], P. Yang[3], C. Chen[4], C. Hsiung[1].* 1) Division of Biostatistics and Bioinformatics,Institute of Population Health Sciences, National Health Research Institutes, Zhunan, Taiwan; 2) National Institute of Cancer Research, National Health Research Institutes, Zhunan, Taiwan; 3) Division of Pulmonary Medicine, Department of Internal Medicine, National Taiwan University Hospital and College of Medicine, Taipei, Taiwan; 4) Genomics Research Center, Academia Sinica; and Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan.

It is estimated that 25% of lung cancer cases arise in never-smoking individuals. Death caused by lung cancer in never-smokers ranks as the seventh most common cause of cancer death worldwide. Epidemiology studies have shown that the incidence of lung cancer in never-smoking women is particularly high in Asia and lung adenocarcinoma is the most common histology subtype among never-smoking lung cancer cases. Several SNPs (single nucleotide polymorphisms) have been identified to be associated with lung cancer susceptibility among never-smokers by genome-wide association studies (GWAS). It is of interest to assess the value of these loci in the context of genetic counselling and population screening for lung cancer. We constructed a multivariate logistic regression model for the computation of absolute risk of a female developing lung cancer within a 5-year period, based on age, classical epidemiological risk factors and these known susceptibility SNPs and illustrate some of its possible uses. Data used in this model development included those from cohort GELAC (genetic epidemiology study of lung adenocarcinoma in Taiwan), on which GWAS were based, and female lung ADC incidence rate based on Taiwan Cancer Registry and census data. GELAC is a case-control study; 1996 lung cancer cases and 1405 healthy controls were used in the model development. Since smoking rate in women is low, between 3-5% for the past decades, we assume females are never-smokers in building the model. This model is potentially useful in designing lung cancer screening programs. For example, applying this model to females with age between 55 and 74 and considering an absolute risk of 0. 83 (1. 28) % as cutoff, 10 (3) % of the population could be identified in which 32 (15) % of the cases would arise. If confirmed in validation studies, it could provide an estimate of lung cancer risk among never-smoking female in Taiwan to guide discussions and decisions regarding prevention and surveillance. Using risk distributions, we also compared our model with the model built on covariates without SNPs, providing an assessment of the value of these susceptibility genetic variants. For example, only 11. 4% of the lung cancer cases would arise from the most risky 3% population if they are identified by the model without any SNP. This study provides a framework to evaluate from the risk assessment viewpoint the value of conducting more GWAS that look for new lung cancer susceptibility loci.

2602T

The genetic contribution to risk of Acute Lymphoblastic Leukemia (ALL) differs across the lifespan in patients treated with unrelated donor allogeneic hematopoietic cell transplant (URD-HCT). *A. Clay[1], T. Hahn[2], L. Preus[1], Q. Zhu[3], L. Yan[3], Q. Liu[4], Q. Hu[3], S. Liu[3], CB. Ambrosone[1], X. Zhu[5], CA. Haiman[6], DO. Stram[6], LC. Pooler[6], X. Sheng[6], D. Tritchler[4], S. Battaglia[7], D. Weisdorf[8], S. Spellman[9], M. Pasquini[10], PL. McCarthy[2], K. Onel[11], LE. Sucheston-Campbell[1].* 1) Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, NY; 2) Department of Medicine, Roswell Park Cancer Institute, Buffalo, NY; 3) Department of Biostatistics and Bioinformatics, Roswell Park Cancer Institute, Buffalo, NY; 4) Department of Biostatistics, SUNY-Buffalo, Buffalo, NY; 5) Center for International Blood and Marrow Transplant Research, Medical College of Wisconsin, Milwaukee, WI; 6) Department of Preventive Medicine, University of Southern California, Los Angeles, CA; 7) Department of Pharmacology and Therapeutics, Roswell Park Cancer Institute, Buffalo, NY; 8) Department of Medicine, University of Minnesota, Minneapolis, MN; 9) Center for International Blood and Marrow Transplant Research, Minneapolis Campus, Minneapolis, MN; 10) Department of Medicine, Medical College of Wisconsin, Milwaukee, WI; 11) Department of Pediatrics, The University of Chicago, Chicago, IL.

Introduction: Genome-wide (GW) significant SNPs associated with pediatric ALL are in *GATA3* (rs3824662), *IKZF1* (rs6964969), *CDKN2A/B* (rs3731217/rs3218005), *ARID5B* (rs10821936), *CEBPE* (rs7157021) and *BMI1-PIP4K2A* (rs7088318), while adolescent young adult (AYA) ALL associates with only the *GATA3* SNP. We tested these SNPs for association with ALL in pediatric (<20 yrs), AYA (20-40 yrs) and adults (>40 yrs) treated with URD-HCT, and with survival. **Methods:** Cohort 1 (C1) and Cohort 2 (C2) controls included 2,229 and 889, respectively, unrelated healthy donors genotyped on Illumina OmniExpress chip along with 504 (C1) and 101 (C2) ALL cases of European Ancestry. C1 and C2 are from the Center for International Blood and Marrow Transplant Research registry. SNPs were tested for association with ALL overall and by age. SNP association with death due to ALL or TRM 1 year post URD-HCT was measured with competing risk models. **Results:** Overall (C1/C2) and age-stratified models (C1 only) are presented (Table) due to the significance of age in all SNP models (p<10e-38). The A allele in rs3824662 was the only SNP associated with hazard of death (Transplant Related Mortality HR=. 66 ±. 16, p=. 008).

SNP	Gene	Overall (C1/C2)		Pediatric		AYA		Adult	
		OR (95%CI)	P	OR (95%CI)	P	OR (95%CI)	P	OR (95%CI)	P
rs3824662	GATA3	1. 6 (1. 4,1. 9)/ 1. 7 (1. 2,2. 4)	6. 4-09/ 0. 002	1. 6 (1. 2,2. 0)	0. 001	1. 4 (1. 1,1. 8)	0. 004	1. 9 (1. 5,2. 5)	2. 6e-07
rs11980379 (LD=1. 0 rs6964969)	IKZF1	1. 5 (1. 3,1. 7)/ 1. 4 (1. 0,1. 9)	4e-07/ 0. 05	1. 8 (1. 4,2. 3)	6. 5e-07	1. 4 (1. 1,1. 7)	0. 006	1. 2 (1. 0,1. 6)	0. 1
rs3218005	CDKN2B	1. 5 (1. 2,1. 9)/ 1. 7 (1. 1,1. 6)	0. 0001/ 0. 02	1. 9 (1. 4,2. 6)	0. 0001	1. 7 (1. 3,2. 3)	0. 003	0. 9 (0. 62,1. 4)	0. 7
rs7090445 (LD=1. 0 with rs10821936)	ARID5B	1. 1 (1. 0,1. 3)/ 1. 2 (0. 9,1. 6)	0. 1/ 0. 3	1. 4 (1. 1,1. 7)	0. 009	1. 0 (0. 8,1. 2)	0. 9	1. 1 (0. 8,1. 4)	0. 6
rs7157021	CEBPE	0. 8 (0. 7,0. 9)/ 1. 3 (1. 0,1. 7)	0. 02/ 0. 09	0. 8 (0. 6,1. 0)	0. 07	0. 9 (0. 8,1. 2)	0. 6	0. 8 (0. 6,0. 9)	0. 03
rs7088318	BMI1-PIP4K2A	0. 9 (0. 8,1. 0)/ 0. 9 (0. 7,1. 3)	0. 09/ 0. 7	0. 8 (0. 7,1. 1)	0. 1	0. 9 (0. 7, 1. 1)	0. 5	0. 9 (0. 7,1. 1)	0. 4

Conclusions: SNP associations varied across the lifespan, indicating that susceptibility to ALL may differ by age of onset. It is possible that associations are specific to ALL subtypes that differ by age, thus we are currently testing these relationships. The *GATA3* A allele association with hazard of transplant related mortality requires validation. The null associations with known GW significant pediatric SNPs (ie *ARID5B*) in our study may indicate those treated by URD-HCT for ALL have different genetic susceptibilities.

2603W

ENCAPP: elastic-net-based prognosis and biomarker discovery across human cancers. *J. Das[1,2], K. M. Gayvert[3], H. Yu[1,2].* 1) Department of Biological Statistics & Computational Biology, Cornell University, Ithaca, NY; 2) Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY; 3) Tri-Institutional Training Program in Computational Biology and Medicine, New York 10065, NY, USA.

With the explosion of genomic data, there has been significant effort to understand the molecular basis of cancer using informatics approaches. One particularly challenging problem is to determine prognosis for different patients. Conventional clinical approaches rely on overall physiological and histological parameters and have been shown to be inaccurate. Alternatively, traditional genomic approaches use alterations in transcriptomic profiles to predict outcome but assume independence between genes, a major limitation since the proteins encoded by these genes are part of a protein-protein interaction network within the cell. Here, we present ENCAPP, an elastic-net-based approach that uses interaction dynamics, a combination of gene expression data with the topological structure of the reference human protein interaction network, to predict outcome for different human cancers. Our method identifies functional modules that are differentially expressed between patients with good and bad outcomes and uses these to fit an elastic-net-based regression model for cancer prognosis. ENCAPP is successful at outcome prediction across cancer types and subtypes (median AUCs = 0. 79 and 0. 69 for two clinically distinct breast cancer subtypes, 0. 67 and 0. 77 for colon, and ovarian cancers, respectively) and performs significantly better than existing methods (increase in AUC of 0. 07 over the best existing method). It can also predict outcome across breast cancer subtypes: trained on one subtype, it is works reasonably well at outcome determination for a different subtype (median AUC = 0. 65). Finally, ENCAPP can be used for prognosis across related cancer types: trained on colon cancer data, it is successful at predicting rectal cancer outcome (median AUC = 0. 80). Our method is highly robust to incompleteness of the reference protein network. It also works well for alternate definitions of prognosis i. e. , different outcomes such as death or metastasis and different cutoffs for right censoring of the outcome variable. Furthermore, we find that top differentially expressed functional modules used as decision boundaries in outcome prediction are enriched for known cancer genes. Thus, these modules can be leveraged to identify novel prognostic biomarkers that can be probed to generate mechanistic and therapeutic insights.

2604T

Genetic propensity to allergic rhinitis is associated with sex-specific leukemia risk reduction in children. *E. Elhauge[1], J. Wiemels[2], M. Wrensch[2], A. Molinaro[2], M. van der Laan[1], L. F. Barcellos[1], K. Walsh[2].* 1) School of Public Health, Division of Epidemiology, UC Berkeley, CA; 2) Dept. of Neuroepidemiology, UCSF, San Francisco, CA.

Previous epidemiologic studies have observed inconsistent associations between acute lymphoblastic leukemia (ALL [MIM 613065]) risk and measures of atopic disease. To address this, we quantified the effect of genetic propensity to allergic rhinitis (AR [MIM 607154]) on the risk of ALL. An instrumental variable (IV) was constructed utilizing genome-wide SNP data and adaptive statistical methods to predict medically diagnosed AR in an adult population; this AR IV predicts the genetic propensity to AR. The AR IV was then applied to an independent ALL case-control dataset to test whether the AR IV was associated with ALL risk. This study is the first to leverage the size and SNP data available from the Kaiser Genetic Epidemiology Research on Aging (GERA) medical cohort (N=78,000, European-ancestry) to create a predictive and generalizable IV for propensity to AR of higher power than is typically available in a case-control studies. A causal model was built that depended exclusively on the AR IV of each subject and their genetic ancestry; this model meets the assumptions of Mendelian Randomization, preserved analytical power and eliminated multiple-testing correction. To produce the most predictive AR IV, machine-learning, cross-validation methods (ML) were used. ML selected sparse, partial least squares as the most predictive model. ML was used to guide a dimensionality-reduction process from the Axiom genome-wide array to just 983, highly-predictive SNPs. The use of ML blocked model overfitting, while the use of an independent test dataset eliminated any confounding due to residual overfitting. The ALL case-control subjects included cases from The Children's Oncology Group (n=980, European-ancestry) and controls from The Wellcome Trust Case-Control Consortium (n=2,624, European-ancestry), genotyped on the Affymetrix 6. 0 array. Among males, the AR IV was associated with a reduced risk of ALL: RR 0. 43, 95% CI [0. 20, 0. 94], P-value 0. 04; suggesting that inherited propensity to atopic disease is associated with reduced leukemia risk. Among females, the AR IV had a weak protective effect that was not significant: RR 0. 86, 95% CI [0. 38, 1. 96], P-value 0. 73. Ongoing work will attempt to: a) validate this association in additional ALL case-control datasets, b) characterize functional pathways for AR propensity SNPs, c) understand the sexual dimorphism of this effect, and d) reduce the variance and increase the predictive power of the AR propensity estimation.

2605W

Common Genetic Variants in 11q13. 3 and 9q22. 33 Are Associated With Molecular Subgroups of Multiple Myeloma. *S. W. Erickson[1,2], C. Stein[3], O. W. Stephens[3], S. S. Chavan[3], N. Sanathkumar[3], E. Tian[3], J. Epstein[3], G. Morgan[3], B. Barlogie[3], C. J. Heuck[3], A. J Vangsted[4].* 1) Department of Biostatistics, University of Arkansas for Medical Sciences, Little Rock, AR; 2) Currently: Genetic Epidemiology and Omics Research, RTI International, Research Triangle Park, NC; 3) Myeloma Institute, University of Arkansas for Medical Sciences, Little Rock, AR; 4) Department of Hematology, Rigshospitalet, Copenhagen University, Denmark.

Weinhold et al (2013) reported the association of a potentially functional SNP in *CCND1* with increased incidence of the t(11;14) translocation in multiple myeloma (MM). MM can be classified by gene expression profiling (GEP) into seven well-defined molecular subgroups (Zhan et al, 2006), of which four are strongly influenced by specific translocations involving the immunoglobulin heavy chain (IgH) locus. The generation of fusion proteins with functional consequences is one of the biological impacts of IgH-related translocations, and this forms the basis of the translocations and cyclin D expression (TC) classification system of MM. We investigated whether SNPs previously associated with overall risk of MM or t(11;14) are associated with GEP- and TC-based disease subsets. Genotypes at 777,681 SNPs in a discovery set of 713 newly diagnosed MM patients of European ancestry (EA) and 1064 cancer-free controls, plus a replication set of 252 more recent EA MM patients genotyped at selected SNPs, were used for this study. GEP data from CD138-selected plasma cells taken within 3 months prior to treatment were available for all MM patients. The SNP rs603965, located in *CCND1* at 11q13. 3, was significantly associated with the TC-based CCND1 subgroup in the discovery set (OR=1. 82, p=1. 7 x 10-4) and in our replication set (OR=2. 19, p=1. 3 x 10-4), which confirms the finding of Weinhold et al (2013). We found that overall survival of CCND1 patients trended worse among non-carriers of the G allele than either heterozygous or homozygous carriers, although with marginal statistical significance (p=0. 08). The SNP rs73486634 at 9q22. 33, located between and upstream of *FOXE1* and *XPA*, was associated with the proliferative (PR) GEP-based subgroup in the discovery set (OR=5. 06, p=6. 0 x 10-9) and was confirmed in the replication set (OR=3. 16, p=0. 006). The PR subgroup exhibits increased expression of proliferation-associated genes and is characterized by high rates of relapse, but almost always in the absence of IgH translocations. Variants in *FOXE1* and *XPA* have been associated with carcinogenesis, and lung and colorectal cancer. Our findings confirm previous results and identify novel association loci, showing that genetic variation influences the risk for specific subgroups of MM.

2606T

Large scale meta-analysis identifies several new risk loci for development of esophageal adenocarcinoma and Barrett's esophagus in individuals of European descendents. *P. Gharahkhani[1], R. Fitzgerald[2], T. Vaughan[3], I. Tomlinson[4], I. Gockel[5], C. Palles[4], M. Knapp[6], MM. Noethen[7, 8], J. Becker[7, 8], S. MacGregor[1], P. Pharoah[9], DC. Whiteman[10], J. Jankowski[11,12], J. Schumacher[7, 8], Barrett's and Esophageal Adenocarcinoma Consortium (BEACON), wellcome Trust Case Control Consertium.* 1) Statistical Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia; 2) Medical Research Council (MRC) Cancer Cell Unit, Hutchison-MRC Research Centre and University of Cambridge, Cambridge, United Kingdom; 3) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; 4) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 5) Department of Visceral, Transplant, Thoracic and Vascular Surgery, University Hospital of Leipzig, Leipzig, Germany; 6) Institute for Medical Biometry, Informatics, and Epidemiology, University of Bonn, Bonn, Germany; 7) Institute of Human Genetics, University of Bonn, Bonn, Germany; 8) Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany; 9) Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, United Kingdom; 10) Cancer Control, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4029, Australia; 11) University Hospitals Coventry & Warwickshire NHS Trust, Warwickshire, England; 12) Warwick Medical School, University of Warwick, Warwickshire, England.

Esophageal adenocarcinoma (EA) is a fatal cancer with rising incidence in the developed world. Barrett's esophagus (BE), a metaplastic change of esophageal epithelium, is associated with greatly increased risk of EA. Genetic predisposition to BE and EA is incompletely understood. Although genome-wide association studies (GWAS) have identified eight loci associated with development of BE and/or EA, these loci together explain small proportion of phenotypic variance. To investigate additional loci increasing the risk of BE and EA, we performed the largest meta-analysis of GWAS results conducted for BE and EA to date, including 6,177 cases of BE, 4,120 cases of EA, and 17,067 unscreened controls. All participants were of European descent. Since we recently showed a significant genetic overlap between BE and EA, we also performed a combined analysis of BE and EA (10,297 cases of BE+EA combined) to maximize power. Imputation was carried out using the June 2014 release of the 1000 Genomes Phase I haplotypes as the reference panel. Meta-analysis was performed using fixed-effects inverse variance weighting approach. Our analysis of BE+EA identified eight new loci increasing risk of BE and EA in single variant tests and three additional loci in gene-based tests. These include an association in the gene *CFTR* (rs17451754 [G], P=5. 21×10-10, OR=1. 19) as well as two other associations to regions known to affect obesity related traits (*TPPP/CEP72* and *MFHAS1* loci). The *TPPP/CEP72* locus has been also associated with ulcerative colitis in a previous study. Additionally, at the gene *HTR3C* (rs9823696 [A], P=2. 01×10-8, OR=1. 17), we identify for the first time an association signal where the effect is only on EA and not mediated via BE (rs9823696 [A], P= 0. 494, OR= 1. 02 for BE). Functional annotation shows several of the novel loci are either expressed in gastrointestinal tract, change sequence motifs for binding of several proteins and/or are eQTL in adipose or other tissues. These new loci, in addition to the previously identified risk loci, will improve risk profiling and management of high-risk individuals. Further investigation of these new risk loci and downstream biological pathways will help in better understanding the etiology of BE and EA, and paves the way for subsequent development of treatments.

2607W

Non-Small Cell Lung Carcinoma patient survival with ALK polysomy vs. ALK-EML4 rearrangement identified by Fluorescence In Situ Hybridization analysis. *R. Gupta[1], L. Cannizzaro[2], L. Vasovic[3], K. H. Ramesh[2].* 1) Montefiore Medical Center, Bronx, NY; 2) Albert Einstein School of Medicine, Bronx, NY; 3) Cornell University, New York, NY.

Non-small cell lung cancer (NSCLC) is the leading cause of cancer-related mortality world-wide with a 5-year survival rate of 15%. Polysomy, defined as three or more gene copies, is commonly seen in a wide variety of cancers and plays an oncogenic role. A fusion oncogene consisting of an inversion in chromosome 2, anaplastic lymphoma kinase (ALK)-echinoderm microtubule-associated protein-like 4 (EML4), is described in 2-5% of NSCLC and is associated with the progression of disease. ALK polysomy and ALK-EML4 gene rearrangement cases were first identified by Fluorescence In situ Hybridization (FISH) analysis. FISH was performed on formalin-fixed paraffin embedded tumor tissue cases with NSCLC, using the ALK gene break-apart probe (Abbott Molecular, USA, FDA approved). Areas representative of tumor were selected by reviewing pathologists, serial unstained sections were prepared for FISH with deparaffinization and pretreatment using the Vysis ThermoBrite FDA approved protocol. Image analysis was performed using the Metasystem Ikaros Software (Metasystem, Germany). A sample was considered positive if >50% cells were positive; negative if <10% cells were positive; and equivocal if 10-50% of the cells were positive. The aim of our study was to assess the survival difference in NSCLC patients without a history of tobacco use with ALK polysomy or the fusion oncogene. Using the Clinical Looking Glass database at Montefiore Medical Center, we retrospectively identified four cases of ALK-EML4 gene rearrangement and 108 cases of ALK polysomy by FISH analysis from 2011-2014. Amongst the two groups, there were no significant differences in age (p=0. 47) and there was a higher percentage of female patients in the rearrangement group than in the polysomy group (3/4Vs. 54/54). Using log-rank statistical analysis, there were no significant differences in survival from the date of NSCLC diagnosis between the polysomy and rearrangement groups (p=0. 37). In conclusion, the lack of statistical significance in survival between the two groups may suggest that the oncogenicity of polysomy of ALK and the ALK-EML4 gene rearrangement in NSCLC patients works by similar mechanisms. However, the small sample size and single center study preclude any definitive conclusions in the survival differences. With clear knowledge of mortality in the two groups with a larger cohort of patients, molecular targets can be identified for the formulation of drugs that can prolong survival.

2608T

Association of *-31 C/T* polymorphism of the *IL-1*β gene was associated with colorectal cancer patients from Mexican population. *I. A. Gutierrez[1, 2], A. Y. NARANJO[3], V. TRUJILLO[3], M. G. VILLA[3], L. E. FIGUERA[4], A. M. PUEBLA[5], M. P. GALLEGOS[1].* 1) MEDICINA MOLECULAR, INSTITUTO MEXICANO DEL SEGURO SOCIAL, GUADALAJARA, JALISCO, Mexico; 2) DOCTORADO EN GENETICA HUMANA, CENTRO UNIVERSITARIO DE CIENCIAS DE LA SALUD, UNIVERSIDAD DE GUADALAJARA, GUADALAJARA, JALISCO, MEXICO; 3) SERVICIO DE ONCOLOGIA, UMAE, HOSPITAL DE ESPECIALIDADES, CENTRO MEDICO NACIONAL DE OCCIDENTE, INSTITUTO MEXICANO DEL SEGURO SOCIAL; 4) DIVISION DE GENETICA, CIBO, IMSS; 5) LABORATORIO DE INMUNOFARMACOLOGIA, CUCEI, UNIVERSIDAD DE GUADALAJARA.

Background: A quarter of malignant lesions of colorectal cancer (CRC) are rooted in chronic inflammatory processes, however, the exact mechanism is still unknown, but several studies have shown that inflammatory cells modulate colon carcinogenesis through cell stimulation and synthesis stromal mitogenic molecules. Some Studies evidences that role of inflammation in CRC showing that proinflammatory cytokine gene polymorphisms increase the risk of cancer and its precursors. The *IL-1b-31C/T* has been shown to be functionally significant with the *T* allele that being associated with increased production of IL-1 cytokine. Our goal is to evaluate the association of polymorphism *-31C/T* in colorectal cancer patients from Mexican populations. **Methods:** 206 samples (UMAE Hospital Especialidades, CMNO, IMSS) and 124 controls group (health Mexican population) were included in this study. The *-31C/T* polymorphism genotyping was determinate by PCR method. **Results**: The genotype *-31 TT* was associated with colorectal cancer [OR = 2. 3 (95% CI 1. 2 – 4. 4), p = 0. 007]. **Conclusion:** These results suggest that the *-31 TT* genotype of the *IL-1b* polymorphism was associated with colorectal cancer patients from Mexican population.

2609W

A common variant on 2q31. 3 reduces lung cancer risk among light smokers: Transdisciplinary Research in Lung Cancer Consortium. *S. Han[1], I. Cheng[2,3], Y. Han[4], D. Christiani[5], R. Hung[6], H. Bickeböller[7], X. Wu[8], P. Brennan[9], C. Amos[4].* 1) Department of Radiology, Stanford University, Stanford, CA; 2) Epidemiology, Cancer Prevention Institute of California, Fremont, CA; 3) Stanford Cancer Institute, Stanford, CA; 4) Biomedical Data Science, Dartmouth, Hanover, NH; 5) Harvard School of Public Health; Boston, MA; 6) Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada; 7) Department of Genetic Epidemiology, University Medical Centre Göttingen, Göttingen, Germany; 8) Department of Epidemiology, the University of Texas MD Anderson Cancer Center, Houston, TX; 9) International Agency for Research on Cancer, Lyon, France.

Background: Heavy smoking increases the risk of lung cancer (LC) by 50-fold. However, the lifetime risk of LC is between 15-20% even among heavy smokers, raising the question whether there may be protective factors that mitigate the carcinogenic risk of smoking. Previously, we have shown significant gene-smoking interactions on chromosome 15q25. 1, with genetic variants in this region associated with minimal increased risks of LC among never smokers, but substantially increased risks among smokers. Studies contrasting LC risk between heavy and light smokers have not been conducted, and little is known about protective factors that may reduce LC risk for certain smokers. **Methods:** To identify genetic factors that may reduce LC risk by levels of smoking, we conducted a genome-wide case-only analysis to detect gene-smoking interactions for LC, comparing heavy smokers (≥ 30 cigarette pack-years) versus light smokers (< 30 cigarette pack-years). The case-only design provides improved power for detecting gene-environment interactions provided there is no correlation between the genetic factor and the environmental exposure in the underlying population. Genotype data for 4,639 heavy smoker cases and 1,824 light smoker cases were meta-analyzed from 7 studies within the Transdisciplinary Research in Cancer of the Lung (TRICL) consortium. Imputation was conducted based on the 1000 Genomes Project to analyze data on roughly 9 million genetic variants. **Results:** The most significant gene-smoking interaction was found on 2q31. 3 with rs62180069 (*P*=5x10-8; OR=0. 76). This SNP lies between the *SCHLAP1* gene encoding the SWI/SNF complex antagonist associated with prostate cancer 1 and the *UBE2E3* gene. While this variant showed no association with risk of LC among heavy smokers (OR=1. 03; 95% confidence interval (CI): 0. 94-1. 13), it significantly lowered the risk of LC among light smokers (OR=0. 85; CI: 0. 78-0. 93). There was no evidence of gene-smoking correlation among controls (*P*=0. 41) and no evidence of heterogeneity in the associations across studies (*P*=0. 94). **Conclusions:** Our large-scale meta-analysis identified a protective genetic variant for LC that reduces the risk of disease among light smokers. Further studies are needed to characterize the biological mechanism underlying the interaction between rs62180069 and smoking behavior and the functional impact of this variant on LC risk.

2610T

Interactions of Family History with Baseline Risk Factors at 5-year Follow-up in a Colorectal Cancer (CRC) Screening Study in US Veterans. *E. R. Hauser[1,2], L. W. Musselwhite[2], R. B. McNeil[1], D. H. Abbott[1], C. M. Thomas[1], M. C. O'Leary[1], S. F. Morris[1], L. Wagner[1], Z. F. Gellad[1,2], D. G. Weiss[3], D. A. Lieberman[4], D. T. Provenzale[1,2].* 1) Cooperative Studies Program Epidemiology Center, Durham VAMC, Durham, NC; 2) Duke University Medical Center, Durham, NC; 3) Department of Veterans Affairs, Perry Point, MD; 4) Oregon Health & Science University, Portland, OR.

The Veterans Health Administration Cooperative Studies Program Study #380 (CSP380) "Prospective Evaluation of Risk Factors for Large (>1 cm) Colonic Adenomas in Asymptomatic Subjects" is a longitudinal study of 3121 participants enrolled for screening colonoscopy (CP)in 1994-1997. CSP380 also includes a biorepository of blood, serum, DNA, and normal rectal tissue in 815 subjects along with sections of large polyps removed during the baseline CP. Longitudinal pathology specimens are available in individuals with repeat CPs. The VA EMR allows follow-up of medical conditions over the ~20 years since the baseline visit. Risk factors were assessed by detailed survey data collected at the time of study enrollment. Per the CSP380 protocol, follow-up CPs wereperformed at either 5 years or 2 and 5 years. Advanced neoplasia (AN)was defined as tubular adenoma >10mm, adenoma with villous histology, adenoma with high-grade dysplasia, or invasive cancer. Of the 1193 participants with follow-up CPs, 92 hadAN, 392 had small adenomas, and 709 had no neoplasia(NN)detectedup to5-years post baseline. In CSP380 models comparing AN to NN, after adjustment for sex, race, smoking, BMI, alcohol consumption, coronary heart disease, and family history (FH, defined as positive for any 1st degree relative with CRC), baseline risk factors for ANat 5 years included older age, baseline neoplasia, and diabetes (self-report at baseline), which were significantlyassociated with AN at 5 years. Family history (FH)has been implicated in increasing CRC risk in younger individuals and individuals with AN at baseline. Diabetes has a strong genetic component and the most well-replicated type-2 diabetes gene *TCF7L2* displays multiple somatic mutations in analysis of CRC tumors. We hypothesized that FH would interact with baseline age, baseline neoplasia, or diabetes to multiplicatively increase risk of AN at 5 years. In multivariable models, after adjusting for all other baseline risk factors, we observed no evidence of a multiplicative interaction between FH and age (p-value for interaction=0. 65) or baseline neoplasia (p-value for interaction=0. 39). We observed evidence for amultiplicative interaction between FH and diabetes (p-value for interaction=0. 04). Future genomic analysis will include sequencing of *TCF7L2* in neoplastic and normal tissues. The CSP380 dataset provides a unique opportunity to evaluate the lifestyle and genetic aspects of lifetime risk of AN and CRC.

2611W

Identification of Subtype Specific Somatic Mutations in Korean Patients with Acute Myeloid Leukemia. *S. Heo[1,2], Y. Koh[3], J. Kim[4], J. Jung[5], H. Kim[1], S. Yoon[3], J. Park[2].* 1) Ewha Womans University, Seoul, South Korea; 2) Hallym University, Chuncheon, South Korea; 3) Seoul National University Hospital, Seoul, South Korea; 4) National Cancer Center, Seoul, South Korea; 5) Syntekabio Inc. , Seoul, South Korea.

Acute myeloid leukemia (AML) is a biologically and clinically heterogeneous cancer of the bone marrow characterized by rapid growth of abnormal myeloid cells. Recent next-generation sequencing approach is capable of extending AML research to genome-wide scope at a single-base resolution. We performed analysis of somatic mutations for AML followed by subtype-specific analyses (M2 and M3) and survival analyses using the whole-exome sequencing data of 36 Korean patients with *de novo* AML. We identified 61 significantly mutated genes in AML which include six M2 subtype-specific genes, *CEBPA*, *EP400*, *MAGI1*, *MAML2*, *ZFHX3*, and *TRIM49* and seven M3 subtype-specific genes, *ATXN3*, *TDG*, *HCLS1*, *IGSF3*, *CDC27*, *PDE4DIP*, and *CFHR1*. We also demonstrated that the patients with mutations in a tumor suppressor gene,*RUNX1* (HR=2. 25, *p*=0. 007) anda notch pathway gene, *MAML3* (HR=2. 39, *p*=0. 037) showed worse prognosis compared to the patients with no mutationin the total AML patients and the group with M2 subtype respectively. We explored the functional impact of the genes identified in the mutation analysesand highlighted the functions of these genes incell aging (GO:0045766, *p*=2. 78×10-3), myeloid cell differentiation (GO:0030099, *p*=0. 03), and AML signaling pathway (hsa05221, *p*=0. 001), which plays an important role in AML tumourigenesis. Future studies would be worthwhile to evaluate the usefulness of these genes in genetic testing for early diagnosis and prognosis prediction of AML patients.

2612T

Focused analysis of a panel of DNA repair-related genetic polymorphisms reveals *MGMT* as a novel susceptibility gene for papillary thyroid carcinoma in Belarusian children exposed to radiation. *F. Lesueur[1], C. Lonjou[1], F. Damiola[2], M. Moissonnier[3], G. Durand[4], V. Tenet[5], I. Malakhova[6], V. Masyakin[7], F. Le Calvez-Kelm[4], E. Cardis[8], G. Byrnes[9], A. Kesminiene[3].* 1) Inserm U900, Institut Curie, Mines ParisTech, Paris, France; 2) CRCL, CNRS UMR5286, INSERM U1052, Centre Léon Bérard, Lyon, France; 3) Environment and Radiation, IARC, Lyon, France; 4) Genetic Cancer Susceptibility, IARC, Lyon, France; 5) Infections and Cancer Epidemiology, IARC, Lyon, France; 6) Republican Scientific and Practical Center for Medical Technologies, Informatisation, Administration and Management of Health (RSPC MT), Minsk, Belarus; 7) Republican Research Center of Radiation Medicine & Human Ecology, Gomel, Belarus; 8) Centre for Research in Environmental Epidemiology (CREAL), IMIM (Hospital del Mar Research Institute), CIBER Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain; 9) Biostatistics, IARC, Lyon, France.

BACKGROUND. The genetic contribution to papillary thyroid carcinoma (PTC) risk is greater than that of any other cancer but most of the heritable risk of PTC remains unexplained. Moreover inter-individual variability in response to radiation exposure suggests that carriers of susceptibility alleles may modify the risk of PTC in populations exposed to ionising radiation. In a previous study we showed that common variation within the DNA damage recognition gene *ATM* contributes to the risk of PTC in Belarusian children exposed to fallout from the Chernobyl power plant accident. Here we further explored the role of single nucleotide polymorphisms (SNPs) in DNA repair genes acting downstream of *ATM* in the same population. METHODS. Using a pathway-based approach we examined the association of 146 SNPs included in the Illumina Cancer SNP panel array andlocated in 43 DNA repair genes in 82 PTC cases and 325 population controls matched to cases for age at the time of accident, sex, and type of settlement. The studied population corresponds to a sub-group of subjects from the population-based case-control study carried out in the most contaminated areas of Belarus to evaluate the risk of thyroid cancer after exposure to radioactive iodine in childhood. All genotyped subjects are from the Gomel region in Belarus, and were younger than 15 years at the time of the Chernobyl accident. SNP and gene-based associations were evaluated using conditional logistic regressions with PLINK. A combined p-value for the joining effects of all markers within a gene was calculated using the rank truncated product method. Statistical significance was evaluated empirically by permutation. RESULTS. The intronic SNP rs2296675 in *O6-methylguanine-DNA methyltransferase*, *MGMT*, was significantly associated with an increased PTC risk [per minor allele odds ratio (OR) 2. 44 95%CI 1. 47-4. 04, P= 0. 00039]. This association remained significant after adjusting for multiple testing (Pperm=0. 001). Suggestive association was also found for SNPs rs1047768 in *ERCC5*, (Pperm=0. 03), rs17349 in *PCNA* (Pperm=0. 04), and rs1051685 in *XRCC5* (Pperm=0. 07). CONCLUSIONS. Our results support the involvement of several DNA repair genes acting in distinct functional modules in the risk of developing PTC. In particular, the role of *MGMT* involved in the direct repair module is highlighted and investigation of the functional properties of the methyltransferase may further enhance our understanding about the etiology of PTC.

2613W

Longer telomere length estimated by a genetic risk score is associated with increased risk of non-Hodgkin lymphoma. *M. Machiela[1], Q. Lan[1], L. Teras[2], J. Vijai[3], J. Spinelli[4], R. Vermeulen[5], N. Camp[6], S. Slager[7], M. Yeager[8], Z. Wang[8], A. Nieters[9], K. Smedby[10], J. Cerhan[7], S. Wang[11], S. de Sanjose[12], C. Skibola[13], S. Berndt[1], B. Birmann[14], S. Chanock[1], N. Rothman[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) Epidemiology Research Program, American Cancer Society, Atlanta, GA; 3) Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY; 4) Cancer Control Research, BC Cancer Agency, Vancouver, British Columbia, Canada; 5) Institute for Risk Assessment Sciences, Utrecht University, Utrecht, The Netherlands; 6) Department of Internal Medicine, University of Utah School of Medicine, Salt Lake City, UT; 7) Department of Health Sciences Research, Mayo Clinic, Rochester, MN; 8) Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, MD; 9) Center for Chronic Immunodeficiency, University Medical Center Freiburg, Freiburg, Baden-Württemberg, Germany; 10) Department of Medicine, Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden; 11) Division of Cancer Etiology, City of Hope Beckman Research Institute, Duarte, CA; 12) Cancer Epidemiology Research Programme, Catalan Institute of Oncology-IDIBELL, L'Hospitalet de Llobregat, Spain; CIBER en Epidemiología y Salud Pública (CIBERESP), Barcelona, Spain; 13) Department of Epidemiology, School of Public Health and Comprehensive Cancer Center, Birmingham, AL; 14) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

Non-Hodgkin lymphoma (NHL) is a cancer of lymphocytes with variation in incidence, aggressiveness, and treatment based on subtype. Common NHL subtypes include chronic lymphocytic leukemia/small lymphocytic lymphoma (CLL/SLL), diffuse large B-cell lymphoma (DLBCL), follicular lymphoma (FL), and marginal zone lymphoma (MZL). Evidence from previous nested case-control studies in prospective cohorts measuring peripheral white blood cell telomere length by qPCR suggests peripheral leukocyte telomere length is positively associated with NHL risk. However, nested case-control studies might be confounded by very early disease and could miss biologically relevant time periods if cohort follow-up is not long enough. Therefore, we performed an analysis on inferred telomere length and NHL risk in 10,102 NHL cases (CLL/SLL, DLBCL, FL, and MZL subtypes) and 9,562 controls using genetic proxies of telomere length identified by previous telomere length genome-wide association studies. We investigated individual associations with 9 telomere length-associated single nucleotide polymorphisms (SNPs), as well as with a combined genetic risk score that summed up risk alleles across the 9 telomere length-associated SNPs and weighted risk alleles by their previously reported telomere length regression values. Our analysis detected an excess of significant associations (P-value <0. 05) between telomere length-associated SNPs and NHL risk overall and for CLL/SLL and FL subtypes determined by an exact binomial test. The telomere length genetic risk score, which is an unconfounded proxy for telomere length, was positively associated with overall NHL risk (P-value=8. 8×10−5). Subtype-specific analyses indicated CLL/SLL was the main NHL subtype contributing to this association (OR=3. 02, 95% CI=2. 13-4. 30, P=7. 1×10−10), although a marginal association was observed for DLBCL (P-value=0. 06). In addition, the CLL/SLL association was stronger in men (OR=4. 22, 95% CI=2. 69-6. 62, P-value=3. 96×10−10) than women (OR=1. 85, 95% CI=1. 05-3. 27, P-value=0. 03, P-value interaction= 0. 02). Our results suggest that individuals with a genetic background favoring longer telomere length may have an increased risk for CLL/SLL and possibly other NHL subtypes and that the effects are especially pronounced among men.

2614T

Germline variation in inflammation-related pathways and risk of Barrett's esophagus and esophageal adenocarcinoma. *M. M. Madeleine[1,2], M. F. Buas[1], Q. He[1], L. G. Johnson[1], L. Onstad[1], D. M. Levine[3], A. P. Thrift[4], D. C. Whiteman[5], T. L. Vaughan[1,2], BEACON Consortium Investigators.* 1) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Epidemiology, University of Washington, School of Public Health, Seattle, WA; 3) Department of Biostatistics, University of Washington, School of Public Health, Seattle, WA; 4) Department of Medicine, Baylor College of Medicine, Houston, TX; 5) Cancer Control, QIMR Berghofer Medical Research Institute, Brisbane, Australia.

Incidence of esophageal adenocarcinoma (EA) has risen sharply in Western countries over recent decades. Local and systemic chronic inflammation, operating downstream of disease-associated exposures, appear to be important drivers of the development of BE and its progression to EA. Strong risk factors have been identified for EA and its precursor, Barrett's esophagus (BE), including reflux, obesity, and smoking. An understanding of the role of inherited genetic susceptibility has recently been emerging. To explore whether germline genetic variation related to inflammatory processes may influence susceptibility to BE/EA, we used data from a genome-wide association (GWA) study of 2,515 EA cases, 3,295 BE cases, and 3,207 controls. Our analysis included 10,774 single nucleotide polymorphisms (SNPs) in 455 genes assigned to five pathways: cyclooxygenase (COX), cytokine signaling, oxidative stress, MHC, and NFkB. A principal-components-based framework was employed to evaluate pathway-level and gene-level associations with disease risk. We identified a significant signal for variation in the COX pathway in relation to risk of BE (P=0. 0059, FDR q=0. 03), and in subsequent gene-level analyses found an association with the *MGST1* locus (microsomal glutathione-S-transferase 1; P=0. 0005, q=0. 005). Assessment of 36 *MGST1* SNPs identified 14 variants associated with elevated risk of BE (q<0. 05), including a cluster of 3' intronic variants and several 5' SNPs near the transcriptional start site. Eight of these variants were similarly associated with increased risk of EA. This study provides the most comprehensive evaluation to date of inflammation-related germline variation in relation to risk of BE/EA, and suggests that variants in *MGST1* may influence disease susceptibility.

2615W

Association of MDM309 polymorphism of the MDM2 gene was associated with breast cancer patients from Mexican population. *M. G. Marquez[1,2], H. MONTOYA[1], J. SANCHEZ[1], M. P. GALLEGOS[1].* 1) Medicina Molecular, Instituto Mexicano del Seguro Social, Guadalajara, Jalisco, Guadalajara, Jalisco, Mexico; 2) Doctorado en Genetica Humana, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara.

BACKGROUND: The MDM2 gene encodes a phosphoprotein that negatively regulates the activity of TP53. T309G polymorphism in the promoter of MDM2 gene has been reported to be associated with enhanced MDM2 expression and tumor development in breast cancer risk. However, the results have been inconsistent. Our goal is to evaluate the association of T309G polymorphism of MDM2 gene in breast cancer patients from Mexican populations. **Methods:** 510 samples (UMAE Hospital de Gineco-Obstetricia, CMNO, IMSS) and 237 controls group (health Mexican population) were included in this study. The MDM309 polymorphism genotyping was determinate by PCR method. **Results**: The genotype MDM309GG was associated with breast cancer [OR = 1. 49 (95% CI 1. 04 – 2. 1) , p = 0. 028]. **Conclusion:** These results suggest that the MDM309 GG genotype was associated with breast cancer patients from Mexican population.

2616T

Study of DNA sequence and expression variation in miRNAs and target genes to explore risk of two types of oral precancerous lesions and cancer. *R. Roy, B. Roy.* Human Genetics Unit, Indian Statistical Institute, 203,B. T. Road, Kolkata, India.

Background: In the last few years, study of expression deregulation of microRNAs and their target genes has emerged as a promising field in the area of cancer research as these variations may have functional consequences. **Objectives**: To study effect of polymorphisms and expression deregulation of microRNAs and their target genes in two types of oral precancerous lesions and cancer. **Methods**- Variations in 8 microRNAs and 4 miRNA processing genes were studied in 452 oral cancer patients, 299 leukoplakia patients and 451 healthy controls. Cumulative risk of these SNPs along with gene-environment interaction was calculated. Relative expression of 4 microRNAs (*miR-26a, miR-29a, miR-34b and miR-423*) and miRNA target genes were studied in oral normal, lichen planus, leukoplakia and cancer tissues, with sample size of 20 in each group and data was analyzed to explore their role in carcinogenesis. **Result-** Variant genotypes at *miR-34b* reduced risk of both leukoplakia and cancer. Variant genotypes in *miR-196a2, miR-26a-1* and *RAN* increased cancer risk while that in *XPO5* and *GEMIN3* reduced leukoplakia risk. Cumulative risk of leukoplakia and cancer increased by 2-3 times when individual harbored risk alleles at multiple loci. Gene environment interaction analysis showed *tobacco habits* in combination with *Xpo5, miR-423, miR-34b* and *miR 29a* significantly modulated risk of leukoplakia. Normalized expression of *miR-423, miR-29a, miR-34b and miR-26a* were distinctly different in two precancer and cancer tissues. Expressions of some of the target genes of these microRNAs were significantly down regulated. Based on expression profile of deregulated miRNAs and their target genes, individuals could be accurately clustered. Pathway analysis of the deregulated targets gives hits in EGFR signaling, hemostasis and growth factor related pathways. Also interestingly, the negative correlation between microRNAs and their target genes was observed in normal tissues but got disrupted in the disease tissues including cancer leading to a hypothesis that repression potential of microRNAs on several targets get lost in diseased state. **Conclusion**: We found that SNPs in microRNAs and their biogenesis genes indeed conferred risk of leukoplakia and oral cancer. We not only observed differential miRNA expression in cancer and precancer but also that the miRNAs showed different ability in targeting oncogenes and/or tumor suppressor genes.

2617W

Genome-wide analysis for prostate cancer biochemical recurrence identifies novel loci for disease progression at 9p23. *C. G. Tai[1], T. Hoffmann[1,2], N. Emami[3], E. Jorgenson[4], C. P. Quesenberry[4], J. Shan[4], D. Aaronson[4], J. Presti[4], L. A. Habel[4], C. R. Chao[4], N. R. Ghai[4], D. K. Ranatunga[4], C. Schaefer[4], N. J. Risch[2,4], S. K. Van Den Eeden[4,5], J. S. Witte[1,2,5,6].* 1) Department of Epidemiology & Biostatistics, UCSF, San Francisco, CA; 2) Institute for Human Genetics, UCSF, San Francisco, CA; 3) Biological and Medical Informatics Program, UCSF, San Francisco, CA; 4) Kaiser Permanente Division of Research, Oakland, CA; 5) Department of Urology, UCSF, San Francisco, CA; 6) Diller Family Cancer Center , UCSF, San Francisco, CA.

Prostate cancer continues to account for a large proportion of the cancer burden in the United States as the most common non-skin cancer diagnosed among men. Although the latest estimate for 5-year survival for prostate cancer is about 98. 9%, one of the long-standing challenges has been to distinguish indolent disease from aggressive disease. Previous efforts have defined aggressive disease as high Gleason score at diagnosis or prostate cancer mortality within 10 years, but disease progression may be more clinically relevant. Using a clinically comprehensive dataset that allowed for estimation of time to disease progression based on biochemical recurrence, we were able to conduct a genome-wide association analysis for prostate cancer progression. The analysis was conducted in two datasets defined by first course of treatment: in 3,495 men with prostate cancer who received radiation therapy and in 2,248 men with prostate cancer who underwent radical prostatectomy from the Kaiser Permanente Research Program on Genes, Environment, and Health, the ProHealth study, and the California Men's Health study. They were genotyped at over 650,000 SNPs across the genome on four race/ethnicity-specific Affymetrix Axiom arrays, for Europeans, African-Americans, East Asians, and Latinos. Biochemical recurrence after radiation therapy was defined using the Phoenix definition, a rise in prostate specific antigen (PSA) level of at least 2. 0 ng/mL above PSA nadir following external beam radiation therapy with or without androgen deprivation. Biochemical recurrence after surgery was defined as two consecutive measures of PSA at or above 0. 2 ng/mL following surgery. Among those who received radiation therapy, we identified 783 men (52 African-Americans) who experienced biochemical recurrence (BCR) and compared them to 2,712 men (210 African-Americans) who did not. We identified a significant tag SNP (rs35098745) at 9p23 (p=4. 2x10-8) within the African ancestry subpopulation in strong linkage disequilibrium (LD r2 = 1. 0) with rs62540661, which is located in a DNase hypersensitive region specific to a prostate cancer cell line, LNCaP. This finding may suggest mechanisms for prostate cancer progression due to changes in DNAse hypersensitivity, which can affect transcription of nearby genes and regulatory factors.

2618T

Analysis of Rare Variants in Finnish Colorectal Cancer Patients. *T. Tanskanen[1], R. Katainen[1], J. Kondelin[1], U. Hänninen[1], T. Cajuso[1], JP. Mecklin[2], H. Järvinen[3], L. Renkonen-Sinisalo[3], E. Pitkänen[1], K. Palin[1], LA. Aaltonen[1].* 1) Department of Medical Genetics, University of Helsinki, Helsinki, Finland; 2) Department of Surgery, Jyväskylä Central Hospital, University of Eastern Finland, Jyväskylä, Finland; 3) Department of Surgery, Helsinki University Central Hospital, Helsinki, Finland.

Colorectal cancer (CRC) accounts for 10% of the incidence of cancer worldwide. Inherited factors influence CRC risk substantially, but much of the heritability remains explained. In particular, the significance of rare, intermediately penetrant variants is unclear, and such variants may be population-specific. In this study, we analyze protein-coding loss-of-function variants in 345 Finnish unselected CRC cases and 2 437 Finnish controls using exome and whole-genome sequencing. Analysis methods will include single-variant testing and rare variant association methods (SKAT and CMC). To improve statistical power, we sequenced CRC cases with phenotypic characteristics suggesting high genetic risk (familial cancer, age of onset less than 40 years, more than one primary CRC). Also, because Finland is an isolated founder population, allele frequencies ofcertain disease-causing variants may be highdue to genetic bottleneck and drift effects, which may substantially facilitate their discovery. In addition, we will use whole-genome sequencing and SNP genotyping data to impute low-frequency variants in 1 474 independent Finnish CRC cases and 8 645 Finnish controls. SNP genotyping data will also be used to assess allelic imbalance in the respective tumor samples (n = 1 642). In preliminary analyses, joint variant calling was performed for currently availablewhole genomes (218 cases) and exomes (1 453 controls) using separate bioinformatic workflows. Within the coding target regions (32 Mb), the per-sample mean number of SNVs was 20 800 for exomes and 21 500 for whole genomes with ti/tv ratios of 2. 76 and 2. 89, respectively. An average of 1 300 indelsper sample were called in both exomes and whole genomes. Althougha potential source of genomic inflation, the joint analysis of whole-genome and exome sequencing data seems feasible in this setting, and whole genomes offer the advantage of genotype imputation from population-matchedSNP panels.

2619W

Genome-wide Association Study on Secondary Cancer Risk among Patients with Basal Cell Carcinoma (BCC) of the Skin. *W. Wu[1], P. Kraft[2,4], L. Liang[2,4], J. Han[1,3].* 1) Department of Epidemiology, Richard M. Fairbanks School of Public Health, Melvin and Bren Simon Cancer Center, Indiana University, Indianapolis, Indiana; 2) Department of Epidemiology, Harvard School of Public Health; 3) Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital; 4) Department of Biostatistics, Harvard University School of Public Health, Boston, Massachusetts.

Previous studies suggest a positive association between personal history of non-melanoma skin cancer (NMSC) and risk of subsequent cancer at other sites. To identify underlying genetic mechanisms of second primary cancer after basal cell carcinoma, we conducted a genome-wide association study (GWAS) in two large US cohorts, the Nurses' Health Study (NHS) and the Health Professionals' Follow-up Study (HPFS). We conducted the GWAS among 1,428 BCC patients without subsequent second cancer development and 305 BCC patients with other internal cancers in Caucasians. We used 1000 Genomes version 1 phase 3 reference imputation panel. The most promising variant (association P =1. 32* 10–6) was an intronic SNP located at Staphylococcal nuclease and tudor domain containing 1 (SND1) gene. 76,364 SNPs (P value <= 0. 01) were further grouped into gene transcripts within 20 kb in distance and then into Kyoto Encyclopedia of Genes and Genomes pathways, BioCarta pathways for further pathway analysis. Seven pathways were significantly associated with risk of second cancer: Arrhythmogenic right ventricular cardiomyopathy (ARVC) (p = 3. 85 * 10 -5, Bonferroni p = 0. 0069), Pathways in cancer (p = 4. 99 * 10 -5, Bonferroni p = 0. 0089), Calcium signaling pathway (p = 1. 18 * 10 -4, Bonferroni p = 0. 021), Gap junction pathway (p = 1. 58 * 10 -4, Bonferroni p = 0. 028). Functional annotation using data from Genotype-Tissue Expression (GTEx) project eQTL analysis in whole blood identified seven top associated genes LRRC37A, ALDH16A1, CRHR1-IT1, SETD9, SNX16, TNFAIP6 and ZNF429. This is the first GWAS to our knowledge on second cancer risk among BCC patients, and the findings might advance understanding of the etiology of subsequent malignancies among individuals with BCC.

2620T

Whole-exome case-control association and somatic-germline interaction analyses in melanoma. *Y. Yu, H. Hu, A. Deshpande, R. Fowler, X. Wu, P. Scheet, H. Zhao, C. Huff.* Epidemiology, The University of Texas M. D. Anderson Cancer Center, Houston, TX.

While skin cutaneousmelanoma (SKCM) has a strong hereditary component, the majority of familial clustering of SKCM is still unexplained. To evaluate the contribution of rare, protein-coding variation to SCKM risk, we conducted a whole-exome case-control study of individuals with European ancestryinvolving 314 cases from the Cancer Genome Atlas (TCGA) and a population control set of 2,424 individuals from three sources: 1000 Genomes, controls from a Swedish schizophrenia study, and cancer-free controls recruited at MD Anderson. We performed a gene-based association analysis across the exome using VAAST 2. 1, incorporating a new quality control pipeline for next-generation sequencing data to greatly mitigate systematic biases from sequencing-platform artifacts (with final genomic inflation factor, λ equal to 1. 3). Although no gene reached genome-wide significance, our results replicated several genes previously implicated in melanoma risk, including *MC1R* (p<1x10-5), *MITF* (p = 4. 9x10-4), *BRCA2* (p = 2. 8x10-3), *ATM* (p = 6. 8x10-3), *CHEK2* (p =0. 02) and *TP53* (p = 0. 02). Our results also provide suggestive support for *FANCA* (p = 2x10-3), a gene associated with melanoma survival. In addition, we performed an exome-wide somatic-germline interaction analysis of rare, putatively damaging germline variants and somatic loss of heterozygosity (LOH) events in tumors using the tools hapLOH and SGI. We observed significantly fewer LOH events in the tumors of individuals with putatively damaging *MC1R* variants relative to non-carriers (p = 0. 04), supporting a one-hit model and suggesting that *MC1R* susceptibility variants are incompletely dominant at the cellular level in SKCM tumorigenesis. Additionally, we report the top genome-wide signals in our case-control and somatic-germline analyses. In conclusion, our results confirm previously established SKCM-susceptibility loci and provide a foundation for future rare-variant association studies in SKCM.

2621W

The genetics of gene expression in human pancreatic tissues. *M. Zhang[1], W. Xiao[2], J. Hoskins[1], I. Collins[1], J. Jia[1], H. Parikh[1], J. Powell[3], R. Kurtz[4], J. Smith[5], S. Olson[4], J. Shi[1], G. Petersen[6], N. Chatterjee[1], L. Amundadottir[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA; 2) Division of Bioinformatics and Biostatistics, National Center for Toxicological Research, FDA, Jefferson, AR 72079; 3) Division of Computational Bioscience, Center for Information Technology, National Institutes of Health, Bioinformatics and Molecular Analysis Section, Bethesda, MD 20892, USA; 4) Department of Epidemiology and Biostatistics, Memorial Sloan-Kettering Cancer Center, New York, NY, 10065, USA; 5) Division of Gastroenterology & Hepatology, Georgetown University Hospital, Washington, DC 20007, United States; 6) Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota.

Analyses of RNA-seq data have demonstrated its power in investigating the role non-coding germline variation plays in the regulation of gene expression in multiple tissues. Due to the limited availability of such data for pancreatic tissue samples, pancreas specific expression quantitative trait loci (eQTLs) are still largely unknown, while the identification of these regulatory loci are of particular interest for understanding the development of pancreatic diseases. In this study, we analyzed RNA-seq data and 1000 Genome imputed genotype data for 98 normal and 115 tumor derived (PAAD, TCGA) human tissue samples to explore pancreas specific eQTLs as well as to compare tumor-specific and shared eQTLs in pancreas. We tested the association between ~6M SNPs and expression of ~17,000 genes across the whole genome. After adjusting for multiple testing based on a permutation procedure, we identified 359 cis-eQTL genes and 4 trans-eQTL genes in normal pancreatic tissues as well as 198 cis-eQTL genes and 5 trans-eQTL genes in pancreatic tumor tissues at an FDR<0. 05. 73 out of the 359 cis-eQTL genes in normal tissues have cis-eQTLs in the tumors and 79 out of the 198 cis-eQTL genes in tumor have cis-eQTLs in the normal tissues. Cis-eQTLs were significantly enriched in DNase I hypersensitivity sites (2. 3-6. 8 fold) and histone modification marks/transcription factor binding sites (2. 9-14. 7 fold) based on ENCODE DNase-seq and ChIP-seq data for pancreatic cells. Of interest, such enrichments were more prominent in datasets specific for pancreatic cells as compared to all cell types combined. In addition, enrichment of cis-eQTLs in H3K4me3 binding sites (9. 4 fold) was replicated in an independent ChIP-seq dataset for PANC-1 (9. 5 fold) and h-TERT pancreatic cell lines (8. 8 fold). Cis-eQTLs were more likely to be located closer to the transcription start sites (TSSs) compared to the non-significant SNPs (8-fold enrichment for eQTLs located within 1kb from TSSs). Besides the understanding of gene regulatory mechanisms in pancreas, the identification of pancreas eQTLs aids in the interpretation of genome-wide association studies (GWAS) of pancreatic cancer.

2622T

Mendelian randomization to assess the causal role of dietary calcium on colorectal cancer risk. *N. Zubair[1], M. Du[1], J. Gong[1], P. A. Newcomb[1], L. Hsu[1], A. P. Thrift[2], J. A. Baron[3], H. Brenner[4], G. Casey[5], A. T. Chan[6], S. J. Chanock[7], M. Cotterchio[8], M. Hofmeister[4], L. Le Marchand[9], J. D. Potter[1], M. L. Slattery[10], E. White[1], P. T. Campbell[11], J. Chang-Claude[4], U. Peters[1].* 1) Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Baylor College of Medicine, Houston, TX; 3) University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, NC; 4) German Cancer Research Center (DKFZ), Heidelberg, Germany; 5) University of Southern California, Keck School of Medicine, Los Angeles, CA; 6) Massachusetts General Hospital and Harvard Medical School, Boston, MA; 7) National Cancer Institute, National Institutes of Health, Bethesda, MD; 8) Cancer Care Ontario, Toronto, ON; 9) University of Hawaii Cancer Center, Honolulu, HI; 10) University of Utah Health Sciences Center, Salt Lake City, UT; 11) American Cancer Society, Atlanta, GA.

Background: While some observational studies have reported an inverse relationship between dietary calcium intake and colorectal cancer risk, other studies have reported no association. Inherent limitations of bias and confounding in observational studies may explain these inconsistent results. We therefore used Mendelian randomization (MR) and instrumental variable methods to reassess the role of dietary calcium intake on colorectal cancer risk. **Methods:**Participants included 10 912 individuals with confirmed colorectal adenocarcinomas and 10 908 population-based controls from the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO) and the Colon Cancer Family Registry (CCFR). We used the rs4988235 variant in *MCM6* as an instrumental variable for dietary calcium; this genetic marker is used in genetic tests of lactose intolerance. We used a two-stage approach to estimate the causal association of dietary calcium intake with colorectal cancer risk. First, we fit a multivariable linear regression model to predict dietary calcium. In the second-stage, using a multivariable logistic regression model, we examined the association between genetically predicted dietary calcium and colorectal cancer risk. In each stage we adjusted for age, sex, study, and the first 3 principal components of genetic ancestry. **Results:**In the first-stage (F statistic = 141. 5), each copy of the T allele of rs4988235 was associated with a 60. 66 mg/day (95% CI= 50. 19-71. 13) increase in dietary calcium. In the second stage, we found no association, between genetically predicted dietary calcium and colorectal cancer risk. **Conclusions:**In 21,820 individuals, we did not detect a causal role for genetically predicted dietary calcium intake in colorectal cancer risk. Despite our sizable study population, this null association is difficult to interpret given that the observed association between dietary calcium intake and CRC risk is not strong. Our data suggest that even larger scale MR analyses will be required to corroborate these findings.

2623W

Variants in *ELL2* influencing immunoglobulin levels associate with multiple myeloma. *B. Swaminathan[1], G. Thorleifsson[2], M. Jöud[1,3], M. Ali[1], E. Johnsson[1], R. Ajore[1], P. Sulem[2], B-M. Halvarsson[1], G. Eyjolfsson[4], V. Haraldsdottir[5], C. Hultman[6], E. Ingelsson[7], S. Kristinsson[8], A. Kähler[6], S. Lenhoff[9], U-H. Mellqvist[10], S. Nelander[11], A. Vangsted[12], U. Vogel[13], A. Waage[14], H. Nahi[15], D. Gudbjartsson[2], T. Rafnar[2], I. Turesson[9], U. Gullberg[1], K. Stefánsson[2], M. Hansson[1,9], U. Thorsteinsdóttir[2], B. Nilsson[1,3,16].* 1) Department of Laboratory Medicine, Lund University, Lund, Sweden; 2) deCODE genetics, Sturlugata 8, IS-101 Reykjavik, Iceland; 3) Clinical Immunology and Transfusion Medicine, Laboratory Medicine, Office of Medical Services, Akutgatan 8, SE-221 85, Lund, Sweden; 4) The Laboratory in Mjodd, IS-109 Reykjavik, Iceland; 5) Department of Hematology, Landspitali, The National University Hospital of Iceland, IS-101 Reykjavik, Iceland; 6) Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, SE-171 77, Stockholm, Sweden; 7) Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, SE-751 85 Uppsala, Sweden; 8) Faculty of Medicine, University of Iceland, IS-101 Reykjavik, Iceland; 9) Hematology Clinic, Skåne University Hospital, SE-221 85 Lund, Sweden; 10) Section of Hematology, Sahlgrenska University Hospital, SE-413 45 Gothenburg, Sweden; 11) Department of Immunology, Pathology and Genetics, Uppsala University, Rudbeck Laboratory, SE-751 05 Uppsala, Sweden; 12) Department of Haematology, University Hospital of Copenhagen at Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark; 13) National Research Centre for the Working Environment, Lersø Parkallé 105, DK-2100 Copenhagen, Denmark; 14) Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology, Box 8905, N-7491 Trondheim, Norway; 15) Center for Hematology and Regenerative Medicine, Karolinska Institutet, SE-171 77 Stockholm, Sweden; 16) Broad Institute, 7 Cambridge Center, Cambridge, MA 02142, USA.

We report a novel risk locus for multiple myeloma (MM) at ELL2 (rs56219066T; odds ratio (OR)=1. 25; P=9. 6×10-10). This gene encodes a stoichiometrically limiting component of the super-elongation complex that drives secretory-specific immunoglobulin mRNA production and transcriptional regulation in plasma cells. We found that the MM risk allele harbors a Thr298Ala missense variant in an ELL2 domain required for transcription elongation. Consistent with a hypomorphic effect, we found that the MM risk allele also associates with reduced levels of immunoglobulin A and G in healthy subjects (P=8. 6×10-9 and P=6. 4×10-3, respectively) and, potentially, with an increased risk of bacterial meningitis (OR=1. 30; P=0. 0024). .

2624T

Genome Wide Association Study of Estrogen Receptor Positive Breast Cancer In Ashkenazi Jews Using A Pooled Genotyping Approach. *J. Vijai[1], C. Manschreck[2], J. Przybylo[3], D. Villano[1], T. Thomas[1], A. Maria[1], N. Hansen[4], R. Sakr[1], T. Kirchhoff[5], S. Domchek[6], K. Nathanson[6], M. Robson[1], T. King[1], K. Offit[1].* 1) Medicine, Memorial Sloan Kettering Cancer Center, New York, NY; 2) School of Medicine and Biomedical Sciences, University of Buffalo, New York; 3) Stanford University, Palo Alto, California; 4) Kaiser Permanente , San Diego, CA; 5) NYU Cancer Institute, New York University School of Medicine, New York, NY; 6) Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA.

Genome wide association studies (GWAS) have identified numerous genomic loci associated with increased breast cancer [MIM 114480] risk. While a small set of high-penetrance genes account for 25-30% of inherited risk, other identified genetic loci are low-penetrance, common single nucleotide polymorphisms (SNPs). GWAS studies typically require large number of cases and controls to have sufficient power to detect further low-risk SNPs. Individual genotyping of cases and control DNA is necessary in the presence of population stratification and heterogeneity to find meaningful association. Here, we studied the Ashkenazi Jewish (AJ) founder population as a test case to discover novel DNA variations that could explain association with estrogen receptor positive (ER+) breast cancer using a pooled design. In Stage 1, we created precisely quantitated equimolar pools of individual DNA and genotyped these pools on Illumina Human OmniExpress array. In total, 2 case pools and 3 control pools were sourced from 758 and 1,114 unique samples, respectively. To reduce inaccuracies in estimating population allele frequencies, pooling errors and array batch variance were minimized by separately constructing each pool in triplicate and hybridizing each pooled replicate thrice. After accounting for replicate variance, we ranked our SNPs based on the pooled test-statistic for association. The 32 SNPs with the highest association in our Stage 1 dataset were added to a Sequenom panel, which was used in Stage 2 to individually genotype the same validation cohort of 1,204 cases and 1071 controls. Several of these 32 SNPs were not previously associated with cancer incidence, and most were validated in our Stage 2 analysis at a nominal significance of p=0. 05, while the top variant (rs1115245) approached genome-wide significance (p=5. 47x10-6; OR=0. 735). Case and control allele frequencies estimated in the Stage 1 pooled data were replicated in Stage 2 (Rcase = 0. 957, Rcontrol = 0. 962), a considerable improvement in the ability to estimate population allele frequencies over previous pooled GWAS. Our study represents a technical advancement in the accuracy and scale of pooled GWAS, and suggests novel SNPs potentially associated with ER positive breast cancer in Jews. This confirms the method's efficacy and encourages further use of pooled genotyping or sequencing in the discovery stage of GWAS. The total cost of this GWAS was 7. 5X cheaper than an individually genotyped GWA study.

2625W

Haplotype analysis of common variants in the *BRCA1* and *BRCA2* genes in Turkey. *B. Saglam Ada[1], K. Bilecen[1,2], K. Yararbas[3], T. Hatipoglu[4], M. Say[1], C. Tuncer[3], E. Kavak[5], A. Tukun[1,6].* 1) Genetic Diagnosis Center, Duzen Laboratories Group, Ankara, Turkey; 2) Konya Food and Agriculture University, Department Of Moleculer Biology and Genetic Engineering,Konya,Turkey; 3) Genetic Diagnosis Center, Duzen Laboratories Group, Istanbul, Turkey; 4) Ankara University Biotechnology Institute,Ankara, Turkey; 5) Genomize Inc; 6) Department of Medical Genetics, Faculty of Medicine, Ankara University , Ankara, Turkey.

IntroductionBreast cancer is one of the most common cancers seen in women. About one in eight women will develop breast cancer in her lifetime. Around 10% of overall breast cancer cases have family history. Carriers of *BRCA1* or *BRCA2* germline mutations have been estimated to have a 65-85% lifetime risk of developing breast/ovarian cancer. To take advantage of personalized cancer prevention strategies and for early diagnosis patients and family members at higher risk of breast/ovarian cancer are routinely tested to see whether they are BRCA1/2 carriers. In addition to pathogenicalterationsof BRCA1/2 genes, several polymorphisms have alsobeen reported. Many of them are shown to have protective or possible causal effect on breast cancer. Interestingly, several studies showed that some of these polymorphisms tend to occur together despite of their controversial effects. In this study, *BRCA1/2* single nucleotide polymorphisms (SNPs) were examined for the identification of haplotypes and to evaluate the prevalence of these haplotypes among sporadic early-onset breast cancer patients, familial breast cancer patients, and healthy high risk females in multiple-ethnic regions of Turkey. **Method**Peripheral blood DNA from 271 subjects were investigated by next-generation sequencing using *Illumina MiSeq* between 2014 January and 2015 May. Genomize-Seq and SeqPilot softwares were used for analysis of NGS data. Mutations in the coding gene sequence were screened and all observed mutations have been checked for their pathogenicity at Human Gene Mutation Database. Statistical evaluation of all SNPs and mutations in study group were carried out using by using in-house developed Python scripts. **Results**c. 4837A>G, c. 3113A>G, c. 2311T>C, c. 4308T>C, c. 4485-63C>G, c. -19-115T>C, c. 2082C>T, c. 3548A>G in BRCA1 were determined as a haplotype in the population. All SNP's allele frequency is 33,76%. Either the heterozygosity or homozygosity was identical among all samples. Homozygote haplotype ratio was 6,64%, heterozygote ratio was 27,12% in all cases. c. 426-89T>C, c. 1365A>G, c. 2971A>G, c. 7435+53C>T, c. 865A>C, c. 425+67A>C in BRCA2 were determined as a haplotype in population of Turkey. All SNP's allele frequency is 7,01%. All of them are homozygote or heterozygote in each case. Homozygote haplotype ratio was 0,73%, heterozygote ratio was 6,27% in all cases. Our study indicates a high incidence of one haplotype in BRCA1 gene and one haplotype in BRCA2 gene in our study population.

2626T

Predicting breast and ovarian cancer risks for *BRCA1* and *BRCA2* mutation carriers using common genetic variants. *K. Kuchenbaecker[1], J. Simard[2], K. Offit[3], F. Couch[4,5], D. Easton[1], G. Chenevix-Trench[6], A. Antoniou[1] On behalf of the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (CIMBA).* 1) Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK; 2) Centre Hospitalier Universitaire de Québec (CHUQ) Research Center, Laval University, Quebec City, Quebec, Canada; 3) Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, New York, USA; 4) Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA; 5) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA; 6) Cancer Division, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia.

Women who carry a pathogenic mutation in the *BRCA1* or *BRCA2* gene are at increased breast (BC) and ovarian cancer (OC) risk. Genome-wide association studies have identified >100 common genetic variants that are associated with BC or OC risks. Several of these variants are also individually associated with risk of BC or OC for *BRCA1* and *BRCA2* mutation carriers. However, no study has evaluated the combined effects of all the known common genetic variants on BC or OC risk for *BRCA1/2* mutation carriers. We constructed Polygenic Risk Scores (PRS) derived from population-based studies to systematically assess the combined effects of all known common BC and OC susceptibility variants on the risks for *BRCA1/2* mutation carriers. We used data for 15,252female-*BRCA1* and 8,211 female*BRCA2*carriers who were genotyped on the iCOGS custom array. Each PRS was formed by the sum of the number of risk alleles across the variants weighted by their log-Odds Ratio estimate from population-based studies. We investigated 3 PRS for BC: overall, estrogen receptor (ER) positive, and ER-negative; and one PRS for OC. The association of each PRS with BC or OC risk was evaluated within a retrospective cohort framework with time to diagnosis as the outcome and estimated the Hazard Ratios (HR) per standard deviation increase in the PRS. All PRS were significantly associated with cancer risks for *BRCA1/2* carriers. The PRS for ER-negative BC displayed the strongest association with BC risk in *BRCA1* carriers (HR=1. 29 [1. 25-1. 33], p=8x10-64). In *BRCA2* carriers, the strongest association was seen for the overall BC PRS (HR=1. 26 [1. 21-1. 31], p=3x10-27). The OC PRS was strongly associated with OC risk for both *BRCA1* (HR=1. 26 [1. 20-1. 31], p=1x10-23) and *BRCA2* carriers (HR=1. 51 [1. 39-1. 64], p=9x10-22). These relative risks translate to large differences in absolute risks for carriers: e. g. , the 10% of *BRCA2* carriers with the lowest OC PRS had a ≤8% lifetime risk compared to ≥19% for the 10% of *BRCA2* carriers with the highest PRS. Our findings imply that there is a large overlap in susceptibility loci in the general population and in mutation carriers once stratified by tumor subtype. Tumor subtype-specific PRS derived from population-based studies could be used to predict individualized cancer risks for *BRCA1/2* mutation carriers and should be incorporated into risk prediction models to inform decisions on the cancer risk management for *BRCA1/2* mutation carriers.

2627W

Quantitative analysis of aberrant *BRCA1* and *BRCA2* RNA splicing using targeted RNA-Seq. *M. Blok[1], R. Brandão[2], K. Mensaert[3], D. Tserpelis[1], T. De Meyer[3], R. Szklarczyk[1].* 1) Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, Netherlands; 2) Maastricht Science Programme, Maastricht University, Kapoenstraat 2, 6211 KW Maastricht, the Netherlands; 3) Department of Mathematical Modeling, Statistics and Bioinformatics, Ghent University, Coupure Links 653, B9000, Ghent, Belgium.

Targeted RNA-Seq appears to be a promising method to sequence transcripts in depth and to study alternative splicing in detail for specific genes of interest. We investigated if this technology can be used to detect putative pathogenic changes in RNA splicing caused by genetic variants in the hereditary breast and ovarian cancer genes *BRCA1* and *BRCA2*. We used RNA isolated from IL2/PHA stimulated peripheral blood leukocytes collected from patients with a genetic variant in either *BRCA1* (n=4: c. 213-12A>G; c. 693G>A; c. 4987-3C>G; c. 5277+1G>A) or *BRCA2* (n=4: c. 516+1G>T; c. 6935A>T; c. 7806-1G>T; c. 9117G>A). Except for c. 7806-1G>T, these variants are well characterized for their effect on splicing based on previously published data from RT-PCR studies. These effects include *de novo* exon skipping and intron retentions, and relative changes in the abundance of normal isoforms compared to healthy controls. We performed targeted RNA-Seq according to a protocol from Agilent Technologies (CA, USA) on poly A+ enriched mRNA, using custom designed SureSelect baits (double tiling over the complete locus including UTRs), to enrich for *BRCA1/BRCA2* transcripts. Sequencing was performed on the Illumina HiSeq platform. We developed an in-house data analysis pipeline to detect and prioritise the most significant alternative splicing events in comparison to the other samples. All known relevant alternative splicing events were detected as top prioritised events in an extensive list of other non-significantly different events, except increased exon 12 skipping levels for *BRCA2* c. 6935A>T. The potential sources for the discrepancy between RT-PCR and RNA-Seq will be discussed in more detail. For *BRCA2* c. 7806-1G>T, we report the use of a cryptic splice acceptor site in exon 17 in stead of the theoretically expected complete exon 17 skipping, which causes a pathogenic translation frame shift. Our results confirm that targeted RNA-Seq can serve as a generic, high throughput method to test *BRCA1* and *BRCA2* genetic variants for their putative effect on RNA-splicing. It enables detection of changes in (cryptic) RNA splicing caused by deep intronic variants or missense variants in splice-enhancer/silencer sites which are currently not routinely included in genetic DNA screening or their effect is poorly predicted, respectively. Our RNA-Seq pipeline can also easily be adapted to other genes of interest by the use of other enrichment baits.

2628T

Targeted RNA sequencing with cBROCA to identify and characterize variants altering transcription in breast cancer genes. *S. Casadei[1], S. Gulsuner[1], M. Lee[1], A. Thornton[1], J. Mandell[1], M. -C. King[1, 2], T. Walsh[1].* 1) Dept. Medical Genetics, University of Washington, Seattle, WA; 2) Genome Sciences, University of Washington, Seattle, WA.

A major challenge of clinical genetic testing is to identify DNA variants that alter gene splicing. In addition to appearing at canonical splice sites, such variants may lie deep within exons or introns, where they may alter splice enhancers or introduce new splice acceptors or donors. To identify such variants anywhere in breast cancer genes, we developed a targeted RNA sequencing approach, which we named cBROCA, followed by validation by RT-PCR and cloning. cBROCA entails isolating RNA from patient blood or lymphoblasts, generating cDNA, then hybridizing to our BROCA gene panel. BROCA targets the entire loci, not only exons, of known and candidate breast cancer genes. After capture and barcoding, cBROCA samples are multiplexed and sequenced on our HiSeq2500. Genomic DNA from the same patients is hybridized to the same BROCA panels and run in parallel. The goals are simultaneously to identify all genomic variants of each subject and to determine if any are associated with altered transcription. Two challenges encountered in characterizing transcript effects are the low abundance of mutant transcripts due to nonsense-mediated decay (NMD) and normal expression of multiple transcripts of the same gene even from wildtype sequences. To address NMD, we treated EBV-transformed lymphocytes with 500ug/ml puromycin for 6 hours prior to harvesting, so as to retain mutant transcripts. To address the problem of noise introduced by expression of multiple transcripts from wildtype sequences, we normalized to gene expression levels of multiple controls. Independently, we also evaluate variants for allelic imbalance at transcribed SNPs. As a pilot study, we tested cDNA from 49 individuals with variants previously confirmed experimentally to alter transcription by whole or partial exon skipping, inclusion of new exons, intron retention, or a combination thereof (complex events). RNA was extracted from patients' blood, lymphoblasts, or both. Altered transcript expression was accurately identified by cBROCA for all 49 events. Compared to genome-wide RNA-Seq, targeted RNA sequencing allows very high coverage and multiplexed sequencing. Our data suggests that identification and characterization of transcript effects can be performed in this high-throughput fashion using cBROCA and that gene targeting can robustly detect low abundance transcripts using RNA from patients' blood or lymphoblast cell lines.

2629W

Impact of SNVs on gene expression and alternative splicing in Diffuse Intrinsic Pontine Gliomas. *A. K. Ramani[1], P. Buczkowicz[3,4,5], M. Yu[3,4,5], P. Rakopoulos[3,4,5], Y. Jiang[1], C. Hawkins[3,4,5], M. Brudno[1,2].* 1) Centre for computational medicine, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Computer Science, University of Toronto, Toronto, Ontario, Canada; 3) Arthur and Sonia Labatt Brain Tumour Research Centre, The Hospital for Sick Children, Toronto, Ontario, Canada; 4) Department of Laboratory Medicine and Pathobiology, Faculty of Medicine, University of Toronto, Toronto, Ontario, Canada; 5) Division of Pathology, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada.

The genomic alterations associated with cancers are numerous and varied involving both isolated and large-scale complex genomic rearrangements. Diffuse Intrinsic Pontine Glioma (DIPG) is a rare and fatal form of pediatric high-grade gliomas arising in the brainstem. Most pediatric gliomas exhibit complex genomic signatures with alterations in copy number, SNVs and structural rearrangements(Jones and Baker 2014). Here we present our analysis of the impact of SNVs on gene expression and alternative splicing in DIPGs. We first identified SNVs from three parallel sequencing platforms (whole genome, exome and transcriptome) and select correlated variants across the three platforms. We then study the cis-regulatory impact of these variants on the expression profiles and alternative splicing of downstream genes. By comparing these results in tumour and normal samples we have identified a set of SNVs that show significant impact in DIPGs.

2630T

Modulation of GATA3 binding to an enhancer at the *ESR1* locus (6q25. 1) underlies the association with breast cancer risk. *S. D. Bailey[1,2], K. Desai[3], P. Mazrooei[1,2], M. Lupien[1,2,4].* 1) Princess Margaret Cancer Centre, University Health Network, Toronto, ON; 2) Department of Medical Biophysics, University of Toronto, Toronto, ON; 3) Department of Genetics, Norris Cotton Cancer Center, Dartmouth Medical School, Lebanon, NH; 4) Ontario Institute for Cancer Research, Toronto, ON.

The identification of an association between variation at the estrogen receptor alpha (*ESR1*) locus and breast cancer (BCa) serves as a 'proof of principle' for BCa genome-wide association studies (GWAS). Although, *ESR1* has been implicated as the target gene, the causal variant(s) underlying the association have yet to be elucidated. Determining the variant(s) responsible for the association and how it promotes BCa risk may provide novel insight into BCa progression. Multiple risk-associated single nucleotide variants (raSNVs), originating from samples of different ethnic backgrounds (European & East Asian), have been identified at the *ESR1* locus. Using an integrative genomics approach that combines chromatin openness, histone modifications and transcription factor binding, while leveraging the patterns of linkage disequilibrium (LD) among the different ethnic groups, we identify a functional SNV that is shared between ethnicities and in strong LD with the original raSNVs. The variant allele of the SNV decreases the affinity of GATA3 for the chromatin by ~8-fold and is located within an active enhancer element specific to BCa cells. The functional enhancer physically interacts with the *ESR1* promoter, through a long-range chromatin interaction and the modulation of the enhancer by the SNV results in the allele-specific expression of *ESR1*. Although additional functional variants may exist at the *ESR1* locus, we demonstrate that a single SNV that acts by disrupting GATA3 binding can account for the multiple GWAS signals observed at the *ESR1* locus.

2631W

SMART-ddPCR: a novel method for accurate assessment of preferential allelic imbalance in tumor DNA. *A. J. de Smith[1], K. M. Walsh[2], H. M. Hansen[2], A. Endicott[1], J. K. Wiencke[2], C. Metayer[3], J. L. Wiemels[1,2].* 1) Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA; 2) Division of Neuroepidemiology, Department of Neurological Surgery, University of California San Francisco, San Francisco, CA; 3) School of Public Health, University of California Berkeley, Berkeley, CA.

Genome-wide association studies have identified numerous SNPs associated with cancer risk. Tumor selection of SNP risk alleles, or preferential allelic imbalance (PAI), has been demonstrated in some cancer types. However, the extent to which heritable genetic variants affect tumor evolution, as opposed to tumor initiation, is underexamined. We developed a novel methodology termed Somatic Mutation Allelic Ratio Test using Droplet Digital PCR (SMART-ddPCR) for accurate assessment of tumor PAI. Using SMART-ddPCR, we investigated allelic imbalance (AI) at genomic loci associated with childhood ALL, including: *CDKN2A* (missense variant rs3731249 and tag SNP rs3731217) and *IKZF1* (rs4132601), genes frequently lost in ALL tumors; and *CEBPE* (rs2239633), *ARID5B* (rs7089424), *PIP4K2A* (rs10764338), and *GATA3* (rs3824662), all located on chromosomes gained in hyperdiploid ALL. Using constitutive DNA from SNP heterozygotes to establish thresholds, we measured allelic copy number in tumor DNA from 19-142 heterozygotes per SNP locus. We found significant tumor PAI for *CDKN2A* SNP rs3731249, with 14 subjects with preferential retention of the risk allele and only 3 subjects preferentially retaining the protective allele (p=0. 006). There was a trend towards tumor PAI favoring the risk allele for the *CDKN2A* tagging SNP (p=0. 17) and the *IKZF1* SNP (p=0. 23). We did not find significant tumor PAI in hyperdiploid ALL subjects heterozygous for SNPs in *CEBPE*, *ARID5B*, *PIP4K2A*, or *GATA3*. Using a genomic copy number control assay, we investigated somatic copy number alterations (SCNA) underlying AI at *CDKN2A* and *IKZF1*, revealing a complex range of alterations including homozygous and hemizygous deletions and copy-neutral loss of heterozygosity, with varying clonality. Copy number estimates from ddPCR showed high agreement with those from multiplex ligation-dependent probe amplification assays. We present SMART-ddPCR as a highly accurate method for investigation of tumor PAI and assessment of somatic alterations underlying AI, and demonstrate a direct interaction between heritable and somatic *CDKN2A* variation underlying leukemogenesis. Furthermore, analysis of the overlap of recurrent SCNAs from The Cancer Genome Atlas with cancer-associated SNPs from the NHGRI GWAS Catalog revealed 16/140 recurrent SCNA loci overlapping heritable variants associated with a matching cancer type, including at *TERT*, *MYC*, and *EGFR*. These loci represent candidate PAI regions warranting further investigation.

2632T

Annotation of functional impact of missense mutations in BRCA1. *P. Gaudet, I. Cusin, M. Zahn, A. Bairoch.* SIB Swiss Institute of Bioinformatics, Geneva, Switzerland.

Characterization of the phenotypic effect of mutations provides evidence on which variants of unknown significance (VUS) can be evaluated. We have annotated the phenotypes caused by missense mutations in BRCA1 associated with increase susceptibility to breast and ovarian cancers. Using the information derived from 87 publications, the functional impact of 488 variants was captured, resulting in 1300 different annotations. Each annotation is supported by detailed experimental evidences. Well characterized and assessed functions of BRCA1 includes its ubiquitin-protein ligase activity and its role in DNA repair, as well as its transcriptional regulation activity, response to DNA damage, and UBE2D1, BARD1 and BRIP1 binding. These data provide the most comprehensive resource on phenotypes of BRCA1 variants. We are expanding this work to other disease-causing genes, including BRCA2, as well as genes implicated in Lynch syndrome (MSH2, MSH6, MLH1).

2633W

Extending the prediction of pathogenicity for genetic variants in the DNA double-stranded break repair pathway. *J. Loke[1], K upadhyay[1], M. Groden[1], A. Pearlman[1], S. Klugman[2], H. Ostrer[1].* 1) Pathology, Albert Einstein Sch Med, Bronx, NY; 2) Montefiore Med Ctr, Bronx NY.

Introduction. An average woman's lifetime risk for developing breast cancer is 11%. Among these women, 5-10% come from families at high risk for breast or other cancers and 20% have a mutation in the genes, *BRCA1* and *2.* Others have mutations in genes that, along with *BRCA1* and *2,* play a role in the repair of DNA double-stranded breaks (DSB). Previously, we showed functional effects of mutations in DSB repair genes based on the nuclear localization of BRCA1 in response to radiomimetic agents, changes in the binding of cofactors (BARD1, BRCA2, FANCD2 and PALB2) and phosphorylation of p53. Here, we demonstrate that these effects are observed for mutations in other DSB genes. Functional changes can be used to identify pathogenic mutations in these genes, especially those with negative *BRCA1* and *2* genetic tests. **Methods and Methods.** Lymphoblastoid cells (LCLs) were collected from individuals with four different sets of variants, including known *BRCA1* pathogenic mutations, benign variants (normal), *ATM, BRCA2*, *FANCC*, *FANCG FANCD2, and NBN* mutations and a cohort of 16 subjects with negative *BRCA1* and *2* genetic tests. Flow variant assays (FVAs) were used to determine whether pathogenic variants in DSB repair genes disrupt the nuclear localization of BRCA1, the phosphorylation of p53 and the nuclear p53 regulation by MDM2 in response to the combination of DNA crosslinking and breakage drugs, diepoxybutane, mitomycin C, and bleomycin. All LCLs had DNA sequencing that was analyzed for mutations. **Results.** Mutations in *BRCA1* and *ATM, BRCA2*, *FANCC*, *FANCG*, *FANCD2, and NBN* (phenocopies) decreased nuclear localization of BRCA1 in response to the drug combination (p=3. 9 x10-81 for *BRCA1* and 3. 4 X10-93 for phenocopies), reduced phosphorylation of p53 (p=3. 6 X10-81 for *BRCA1* and 7. 5x10-90 for phenocopies) and increased nuclear MDM2 (p=1. 2 X10-27 for *BRCA1* and 2. 2 X10-44 for phenocopies). Unsupervised analysis of all of these assays demonstrated two apparent clusters, high-risk *BRCA1* and phenocopy mutations and normal sequenced controls. Three subjects with negative *BRCA1* and *2* genetic tests and fully sequenced genomes fell into the high-risk cluster. **Conclusion.** FVAs distinguish mutations in *BRCA1* and phenocopy genes in the DSB repair pathway from benign variants. When applied to a small, high-risk, *BRCA1* and *2* negative genetic test cohort, ~19% had positive results. FVAs of circulating cells may represent a fast and effective tool for assessing breast cancer risks.

2634T

Functional characterization of *BARD1* missense variants. *A. Toland[1,2], C. Lee[3], T. Banerjee[3], J. Gillespie[1], L. Starita[4], J. Parvin[3].* 1) Molecular Virology, Immunology & Medical Genetics, Ohio State University, Columbus, OH; 2) Divison of Human Genetics, Ohio State University, Columbus, OH; 3) Biomedical Informatics, Ohio State University. , Columbus, OH; 4) Genome Sciences, University of Washington, Seattle, WA.

Clinical testing of *BRCA1* and *BRCA2* for susceptibility to hereditary breast and ovarian cancer (HBOC) has been available for 20 years. Recently, due to identification of new HBOC genes, companies that offer BRCA1/2 testing have expanded the number of genes being tested for breast cancer predisposition. Truncating and nonsense mutations in BARD1 are associated with an increased risk of breast cancer. The exact penetrance and lifetime risk of mutations in these genes is not well characterized. Nonetheless, genetic testing of *BARD1* is being offered clinically, and many missense variants of uncertain clinical significance have been identified. BARD1 and BRCA1 proteins heterodimerize, and BARD1 function is critical for BRCA1 function in homology directed repair. Because we have previously shown that functional assays, such as those that measure homology directed repair, have a high sensitivity and specificity for prediction of known *BRCA1* missense mutations, we hypothesized that this assay could have utility in classifying missense variants in *BARD1*. To test this, we assayed the effect of 29 *BARD1* missense variants on BARD1 function in homology-directed repair. Unlike *BRCA1*, we did not have any "known" pathogenic missense variants in *BARD1*. To this end, we used a "synthetic" mutation, L44R, that had been reported to disrupt BARD1/BRCA1 binding. We also included known neutral variants that had high population frequencies. From our assay we identified four variants, L44R, C53W, C71Y and G623E that were defective in DNA repair and one variant that showed intermediate repair capabilities. Two of these variants, C53W and C71Y, may result in loss of function by making the BARD1 protein insoluble. All variants with allele frequencies greater than 1% in dbSNP behaved functionally as wildtype BARD1. These results do not rule out potential effects on splicing. It will be important to integrate this functional data with additional segregation or evolutionary conservation data to provide further support for these four variants as being pathogenic. In summary, the HDR assay may be a robust tool for characterization of variants in genes important in homologous recombination.

2635W

Germline mutations in *ETV6* confer risk of acute lymphoblastic leukemia and thrombocytopenia. *S. Topka[1, 8], V. Joseph[1, 8], MF. Walsh[2], L. Jacobs[1], A. Maria[1], D. Villano[1], P. Gaddam[1], G. Wu[2], RB. McGee[2], E. Quinn[2], H. Inaba[2], C. Hartford[2], CH. Pui[2], P. Steinherz[1], M. Edmonson[2], M. Zhang[3], K. Schrader[5], A. Lincoln[1], P. Stepensky[4], S. Lipkin[6], Y. Goldgur[7], M. Harit[4], ZK. Stadler[1], C. Mullighan[2], M. Weintraub[4], A. Shimamura[3, 9], J. Zhang[2], JR. Downing[2], KE. Nichols[2], K. Offit[1, 6, 8].* 1) Department of Medicine, Memorial Sloan Kettering Cancer Center (MSKCC), New York, NY; 2) St Jude Children's Research Hospital (SJCRH), Memphis, TN; 3) Fred Hutchinson Cancer Research Center and University of Washington, Seattle, WA; 4) Pediatric Hematology/Oncology Department, Hadassah-Hebrew University Medical Center, Jerusalem, Israel; 5) University of British Columbia, Vancouver, B. C; 6) Weill Cornell Medical College, New York, NY; 7) Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY; 8) Cancer Biology and Genetics Program, Memorial Sloan Kettering Cancer Center, New York, NY; 9) Seattle Children's Hospital, Seattle, WA.

Somatic mutations affecting *ETV6* often occur in acute lymphoblastic leukemia (ALL), the most common childhood malignancy. The genetic factors that predispose to ALL remain poorly understood. Here, we identify a novel germline *ETV6* p. L349P mutation in a kindred affected by thrombocytopenia and ALL. A second *ETV6* p. N385fs mutation was identified in an unrelated kindred characterized by thrombocytopenia, ALL and secondary myelodysplasia/acute myeloid leukemia. Leukemic cells from the proband in the second kindred showed deletion of wild type *ETV6* with retention of the *ETV6* p. N385fs. Enforced expression of the *ETV6* mutants revealed normal transcript and protein levels, but impaired nuclear localization. Accordingly, these mutants exhibited significantly reduced ability to regulate the transcription of *ETV6* target genes. Our findings highlight a novel role for *ETV6* in leukemia predisposition.

2636T

Systematic prioritization of druggable mutations across 16 cancer types using a structural genomics-based approach. *J. Zhao[1], F. Cheng[1], Y. Wang[1], Z. Zhao[1,2,3,4].* 1) Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN 37203, USA; 2) Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN 37232, USA; 3) Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, TN 37232, USA; 4) Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN 37232, USA.

The progress of several large-scale cancer genome sequencing projects such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium projects has accumulated a large number of somatic mutations. However, how to address the biological impacts of these mutations to distinguish damaging 'driver' from neural 'passenger' and prioritize druggable mutations remain important but challenging problems in cancer genomics research. Fortunately, the rapid advancements of structural genomics technologies (e. g. X-ray) have made a wealth of protein structure data available, providing an opportunity to interrogate somatic mutations in the context of protein functional features (i. e. , protein-ligand binding sites) to determine their pathophysiological roles in cancer and prioritize druggable mutations that may have the potential to mediate drug binding at the atom resolution. In this study, we proposed a structural genomics-based method SGDriver that incorporates the protein structural information into somatic missense mutations to detect significantly mutated protein-ligand binding sites which we assume are likely to be druggable targets. Applying SGDriver to 746,631 missense mutations from 4,997 tumors across 16 cancer types from TCGA, we detected 14,471 potential druggable mutations in 2,091 proteins (including 1,516 recurrently mutated proteins) across 3,558 cancer genomes, and further identified 298 proteins bearing significantly mutated ligand binding sites (adjusted p-value < 0. 05). we also implemented drug-target network analysis suggesting 98 known and 126 repurposed druggable anticancer targets (e. g. *SPOP* and *NR3C1*). Furthermore, our integrated analysis indicated that 13% of patients harbored known druggable mutations and could benefit from current targeted therapy, and this fraction would increase to 31% after considering drug repositioning strategy. This study demonstrates a novel approach for interpreting somatic mutations and prioritizing druggable mutations for precision cancer medicine research.

2637W

An intronic indel variant confers melanoma risk through *PARP1* expression regulation. *J. Choi[1], M. Makowski[1,2], T. Zhang[1], M. Law[3], C. Sutherland[4], W. Kim[1], M. Kovacs[1], H. Parikh[1], L. Aoude[3], M. Gartside[3], J. Trent[5], L. Hurley[4], M. Vermeulen[2], S. Macgregor[3], N. Hayward[3], M. Xu[1], K. Brown[1].* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD; 2) Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, The Netherlands; 3) Queensland Institute of Medical Research, Brisbane, QLD, Australia; 4) College of Pharmacy, University of Arizona, Tucson, AZ; 5) Translational Genomics Research Institute, Phoenix, AZ.

Recent genome wide association studies identified several new loci for melanoma susceptibility including chr1q42. 1 locus encompassing Poly [ADP-ribose] polymerase 1 (*PARP1*). As an effort to identify effecter genes and functional risk variants from this locus we performed expression quantitative trait loci (eQTL) analysis in 62 melanoma cell lines. Among the genes in 1Mb area eQTL was identified for *PARP1* and subsequently validated by qPCR, where increased *PARP1* levels are significantly correlated with the risk allele (p=0. 03, copy number adjusted). Further allele discrimination qPCR of *PARP1* transcripts in 14 melanoma cell lines heterozygous and of neutral copy for the lead SNP indicated significantly higher proportion for the risk allele (p=. 0001). Same allelic imbalance was also observed in 51 heterozygous melanomas with neutral copy numbers from The Cancer Genome Atlas (p=. 028). To identify functional risk variants mediating these effects we annotated thislocus using six melanoma relevant cell types available from ENCODE and Roadmap database. Based on recent fine mapping data suggesting single-SNP model for this locus we prioritized high LD variants for nomination. Among 65 SNPs of high LD with the lead SNP or imputed best SNP (r2>0. 6 using 1000 Genomes phase3), four exhibited strong evidence as potential transcriptional enhancers. Electro Mobility Shift Assays and luciferase assays on these four variants demonstrated that one of them, a six-base pair indel (-/GGGCCC) in the first intron, displayed strong allelic functionality. Consistent with the expression data melanoma-associated deletion allele results in higher luciferase activity while protective insertion allele binds a group of proteins with higher affinity. Subsequent comparative mass-spectrometry for these insertion-binding proteins identified a striking collection of Guanine-quadruplex binding proteins including RECQL as potential inhibitors of *PARP1* expression. Consistent with this hypothesis RECQL knockdown leads to increased *PARP1* expression in melanoma cell lines carrying insertion allele. Over-expression of RECQL also results in more pronounced allelic difference in luciferase activities indicating that RECQL contributes to allelic *PARP1* expression. These data demonstrate that increased *PARP1* expression is correlated with melanoma risk and an indel variant mediates differential *PARP1* expression possibly through secondary DNA structure binding proteins including RECQL.

2638T

Breast cancer pedigree exome sequencing reveals inherited Rad52 truncation mutation implicated in breast cancer susceptibility. *H. A. Costa[1], M. Sikora[2], K. Hastak[1], J. M. Ford[1], L. C. Laurent[3], C. D. Bustamante[1].* 1) Stanford University School of Medicine, Department of Genetics, Stanford, CA, 94305, USA; 2) Natural History Museum of Denmark, Centre for GeoGenetics, 1350, Copenhagen, Denmark; 3) University of California, San Diego, Department of Reproductive Medicine, La Jolla, CA, 92037, USA.

BRCA1/2 mutations have been the most exhaustively elucidated mechanism for inherited breast cancer cause, however only an estimated 25% of cases are atributable to mutations in these two genes. We have sequenced the exomes of a non-BRCA breast cancer pedigree and have identified a shared truncation mutation in the Rad52 protein. We have performed cellular based assays to functionally characterize this truncation mutation and demonstrate it results in the mis-localization of Rad52 during double strand break repair. This observation gives credence to the growing body of work regarding the complex involvement of Rad52 in breast cancer etiology and progression, and suggests that genetic background can modulate the effect of common or perceived ''benign" mutations among different populations.

2639W

RNA functional studies for the classification of germline variants of uncertain significance that may impair splicing. *M. B. Warf[1], E. Goosen[1], D. Mancini-DiNardo[1], J. Willmott[2], Y. Qian[1], M. Landon[2], K. Rushton[2], K. R. Bowles[1], B. B. Roa[1].* 1) Myriad Genetics, Inc. , Salt Lake City, UT. , USA; 2) Myriad Genetic Laboratories, Inc. , Salt Lake City, UT, USA.

Background: RNA splicing is the process by which non-coding intronic regions of a gene are removed after transcription. Genetic variants within introns may impair splicing, resulting in an RNA molecule that does not encode a functional protein. Depending on the sequence context, certain positions within an intron or exon cannot tolerate changes without functional splicing consequences, while other positions can tolerate a variety of genetic variation without impairing splicing. In these studies we establish an IRB approved research process to enroll and collect samples from selected patients after clinical testing has been performed. These patients carry certain variants of uncertain significance (VUS) that may alter splicing. Using the collected sample, we generate and interpret functional RNA splicing data to aid in the reclassification the variant. **Methods:** RNA is extracted from blood samples acquired from a patient(s) carrying a variant of interest. cDNA is synthesized and PCR is performed, amplifying regions within the gene to be assessed. Splicing patterns are visualized on an agarose gel, and all splicing products are identified via sequencing. The wild-type splicing pattern is confirmed in age/gender matched blood derived control samples, and preferably also within RNA extracted from normal tissue (i. e. breast, ovarian, or colon, etc). Further analysis is performed to determine whether the allele carrying the variants of interest can produce any wild-type splice product. **Results:** We analyzed multiple variants within genes in which pathogenic variants cause an increased cancer risk. In one case, *MLH1* c. 306G>T (which is the last base of an exon), the data indicated the variant fully impaired proper splicing of the mutant allele, and the variant was reclassified as Likely Pathogenic. In another case, *CDH1* c. 715G>T (which creates a cryptic splice acceptor), the data showed the variant only partially impaired proper splicing of the mutant allele, and a wild-type splice product was still made from the mutant allele; this variant remained a VUS. **Discussion:** These studies demonstrate that precise and specific functional RNA studies are helpful to differentiate variants that fully impair splicing and those that only partially impair splicing. Variants that fully impair splicing can be reclassified, while variants with intermediate splicing defects may require additional data to determine if the splicing defect is sufficient to cause increased cancer risk.

2640T

Breast cancer risk variants at 11q13 alter the activity of transcriptional enhancers of novel long noncoding RNAs, *CUPID1* and *CUPID2,* implicated in DNA repair. *J. D. French[1], J. A. Betts[1], M. Clark[2], A. Wiegmans[1], M. Moradi Marjaneh[1], B. Gloss[2], M. Ghoussaini[3], G. Chenevix-Trench[1], D. F. Easton[3], A. M. Dunning[3], T. Mercer[2], M. E. Dinger[2], S. L. Edwards[1].* 1) QIMR Berghofer Medical Resesarch Institute, Brisbane, Queensland (QLD), Australia; 2) Garvan Institute of Medical Research, Sydney, Australia; 3) Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge CB1 8RN, UK.

One of the strongest breast cancer associations identified via GWAS is with SNP rs614367, which lies in a 350kb intergenic region on 11q13. As part of the iCOGs project, we have previously fine-mapped this locus and identified three independent sets of correlated, highly trait associated variants (iCHAVs) associated with estrogen receptor positive (ER+) breast cancer and showed that the strongest iCHAV1 lies in a transcriptional enhancer of *CCND1* [MIM168461], located 125kb away (French et al, 2013). In the current study, we have performed targeted RNA CaptureSeq using capture arrays covering the 11q13 intergenic region and identified two novel lncRNAs, *CUPID1* and *CUPID2,* transcribed from the region. RNA CaptureSeq from a range of tissues demonstrated that *CUPID1* and *CUPID2* expression is highly tissue specific, restricted to breast and kidney. Both lncRNAs are estrogen regulated and overexpressed in ER+ cell lines and tumors. Using chromosome conformation capture (3C) and reporter assays we show that, in addition to the *CCND1* promoter, iCHAV1 acts as a strong enhancer on the *CUPID1* and *CUPID2* promoters. In a breast cancer cell line heterozygous for the iCHAV1 haplotype, we demonstrate by allele-specific 3C that the protective alleles preferentially bind CTCF and participate in chromatin interactions with the *CUPID1* and *CUPID2* bidirectional promoter. Knockdown of *CUPID1* and *CUPID2* decreased *CCND1* and reduced RAD51 foci formation after treatment with ionizing radiation suggesting the both lncRNAs are implicated in HR-mediated DNA repair. In summary, we suggest that in addition to reducing the expression of *CCND1*, breast cancer risk variants at 11q13 reduce the expression of *CUPID1* and *CUPID2* causing impaired DNA damage repair. .

2641W

Multi-modal report summary for AML: Combining NGS and Cytogenetics Data to Enable Therapeutic Decision Making. *C. E. Carmack, T. Natarajan, C. J. Miller, B. Parikh, I. Maurer, M. J. Glynias.* GenomOncology, Cleveland, OH.

IntroductionAcute myeloid leukemia (AML) is a clonal disorder of the blood forming cells characterized by accumulation of immature blast cells in the bone marrow and peripheral blood. It is a heterogeneous disease, and as such, its molecular footprint varies. It is therefore important to test for key genetic biomarkers for both prognostic and predictive (treatment) purposes. Patients can benefit greatly utilizing genetic markers for risk stratification. Given the nature of the various genetic markers, several different molecular methodologies are used to evaluate the patient. Evaluation of mutational status of molecular markers like NPM1, CEBPA, KIT and FLT3 has been shown to help refine the prognosis. For example, if a patient is found to have a normal karyotype and a favorable fusion, he/she may qualify for conventional therapy. However, the physician may decide to further test the tumor for FLT3 ITD, and which, if present, would change the prognosis to unfavorable risk. Now the physician can instead decide to consider the patient for transplant, or clinical trial, as the patient is least likely to benefit from conventional therapy. Karyotype analysis, fluorescence in situ hybridization (FISH) and next generation sequencing (NGS) are all used to evaluate AML patients. By employing all of the molecular data, a better diagnostic can be made. **Methods/Results**Clinical molecular results from patient . bam and . vcf files were filtered through over 200 treatment rules contained in the GenomOncology (GO) Clinical Workbench. Of these 150 were based on SNV only; 45 came from fusion rules; and 14 based on amplification rules. Patient results along with clinical interpretation were reported in a lab configured pathology report. **Conclusions**Currently, DNA-Seq data sets (e. g. Illumina and Ion Torrent Hotspot Panels) are rich sources of clinically relevant information which include SNVs, indels, etc. Even more complex data sets are on the horizon (e. g. RNA-Seq, copy-number, LOH, etc.). The ultimate utility of multi-modal clinical tests lies in being able to provide a comprehensive view of complex genomic alterations that work in tandem to drive tumor growth, response to therapy and disease course. Software that pulls complex data sets together into a single record along with providing clinical rules-based medicine prioritization, helps communicate all the useful genetic information and value contained in precision medicine testing like NGS, and microarrays.

2642T

The correlation between the expression level of has-miR-338-3p with the rate of breast cancer metastasis to lymph nodes. *A. Abak[1], E. Sakhinia[2], S. Amini[2].* 1) Tabriz University of Medical Sciences, Tabriz, East Azerbaijan, Iran; 2) Biotechnology Research Center,Tabriz University of Medical Sciences,Tabriz,Iran.

Breast canceris the most common female malignancies in the world which seriously impacts the female health. In recent years, various studies have been reported to determine the relevance of miRNAs to human cancer. microRNAs (miRNAs) have been demonstrated to be important gene regulators with critical roles in diverse biological processes, including tumorigenesis. miR-338-3p is a recently discovered miRNA and is involved in cell differentiation. Previous studies suggested microRNA-338-3p (miR-338-3p) was down-regulated and play tumor suppressor roles in gastric cancer, colorectal carcinoma ,lung cancer and hepatocellular carcinoma. Knocking downmiR-338-3p in breast cancer cells led to mesenchymal-like changes. MiR-338-3p influenced the expression of the EMT-associated proteins by upregulating the epithelial marker E-cadherin and downregulating the mesenchymal markers, N-cadherin, fibronectin, and vimentin. However, the role ofmiR-338-3p in breast cancer tissues is still unclear. The aim of this study was to investigate the expression level of miR-338-3p in breast cancer and its relationship with the clinicopathological features of patients. Total RNA was first isolated from 40 pairs of breast cancer tissues and adjacent non-tumor tissues. Afterwards cDNAs were synthesized and the expression level of microRNA-338-3p was quantified by real time PCR. The Correlation between the expression level of microRNA-338-3p and clinicopathological features was studied and the capability of microRNA-338-3p to function as a breast cancer tumor marker was also explored. The expression of microRNA-338-3p was significantly down-regulated in breast cancer tissues in comparison with those in the adjacent non-tumorous tissues, and the value was negatively related to advanced tumor, node, metastasis stage and local invasion. The expression of microRNA-338-3p in breast cancer cells transfected with pre-microRNA-338-3p was significantly increased. Furthermore, over-expression of microRNA-338-3p showed obviously suppressed invasion and migration ability in breast cancer cells. Therefore, we conclude thatmiR-338-3p acts as a novel tumor suppressor gene in breast cancer. miR-338-3pcan decrease migratory, invasive, proliferative and apoptotic behaviors, as well as breast cancer EMT. Therefore, targeting with themiR-338-3p might serve as a novel therapeutic application to treat breast cancer patients.

2643W

Acquired Telomere Shortening in Women Diagnosed with Breast Cancer is Associated with Chemotherapy and Depressive Symptoms. *A. Alhareeri[1], D. Lyon[2], M. Dozmorov[3], R. K. Elswick[3,5], TP. York[1], D. Lynch Kelly[2], C. Jackson-Cook[1,4,6].* 1) Department of Human and Molecular Genetics, Virginia Commonwealth University School of Medicine, Richmond, VA; 2) Biovehavioral Nursing Science, University of Florida School of Nursing, Gainesville, FL; 3) Department of Biostatistics, Virginia Commonwealth University School of Medicine, Richmond, VA; 4) Department of Pathology, Virginia Commonwealth University School of Medicine, Richmond, VA; 5) Department of Family and Community Health Nursing, Virginia Commonwealth University School of Nursing, Richmond, VA; 6) Massey Cancer Center, Virginia Commonwealth University, Richmond, VA.

Breast cancer (BC) is one of the most commonly diagnosed malignancies in females. The 5-year survival rate for women with early stage BC is about 90%; however, the quality of life for many of these women may be adversely affected due to treatment/cancer-related side effects, including a constellation of symptoms (pain, depression, fatigue, stress, sleep disturbance and cognitive dysfunction) collectively termed psychoneurological symptoms (PNS). The primary aim of this study is to test the hypothesis that these PNS are correlated with alterations in telomere length (TL) that may lead to acquired changes in the methylation status of juxtaposed genes. To date, we have longitudinally studied 64 women (ages 23-71) with early stage breast cancer (I to IIIA) at 2 time points: prior to chemotherapy (baseline) and 4 weeks following the first chemotherapy treatment (mid-treatment). Measures quantified include TL (using both a monochrome multiplex qPCR assay and a semi-quantitative chromosome-specific FISH assay); genome-wide methylation (Illumina 450K); and depression (using HADS). Mean TLs (based on T/S ratios; qPCR) at mid-treatment were not significantly different from those at baseline (p=0. 493), but chromosome-specific shortening at mid-treatment was observed for 1p; 9q; 12q; 18p; 18q; 21p; and 22p (p=0. 019 to 0. 048). Also, a non-random distribution of significant differentially methylated sites was observed for regions juxtaposed to distal 1p (p<2. 2e-16), 9q (p<2. 2e-16) and 12q (p=6. 314e-12), the latter of which are rich in CpGs. Analyses of TL with depressive symptoms showed significant negative correlations (higher depressive symptom scores with shorter telomeres) for 35 of the 46 telomeres (1p, 2p, 2q, 3p, 3q, 4p, 4q, 5p, 5q, 6p, 6q, 7q, 8p, 8q, 9q, 10p, 11q, 12p, 12q, 13p, 13q, 14q, 15p, 15q, 16p, 17q, 18p, 18q, 19p, 19q, 20p, 20q, 21p, Xp and Xq) (p=0. 001 to 0. 041). Model fitting showed chemotherapy to be predictive of TL changes for a subset of telomeres (p=0. 017 to 0. 049). Collectively, the results of this study show that telomere shortening for a subset of chromosomes is associated with chemotherapy, methylation alterations and depression. Knowledge gained from this study provides a plausible biological basis, at least in part, for the development/persistence of PNS following chemotherapy and offers hope for the future development of therapeutic interventions that could alleviate undesirable side effects in patients receiving treatment for BC.

2644T

Molecular basis of the resistance to X-rays of chondrosarcomas. *J. Aury-Landas[1,2], N. GIRARD[1,2], M. BARREAU[1,2], E. LHUISSIER[1,2], K. BOUMEDIENE[1,2], C. BAUGE[1,2].* 1) Normandie Université, France; 2) UNICAEN, EA4652 MILPAT, Caen, France.

Chondrosarcomas are malignant tumors of bone that produce cartilage matrix. Primary chondrosarcoma is the second most frequently primary malignant tumor of bone after osteosarcoma, and represents about 25% of bone sarcomas. Chondrosarcoma is considered as resistant to both chemotherapy and radiation, making surgical resection the only curative treatment. However, mechanisms of resistance are not well understood. This project aims to investigate resistance mechanism to radiotherapy. We compared the response to X-rays radiation of five different cell lines derived from human chondrosarcomas. They can be classified according to their sensibility to X-rays within 3 groups: X-rays induced both apoptosis and senescence in the most sensitive group, whereas they induce only apoptosis or senescence for the intermediate group. No death could be detected for the most resistant group. To understand the molecular basis of theses different responses to X-rays, we performed whole-exome sequencing on the five cell lines. On average, 60 000 variants were identified per exome, including 1700 new variants. After filtration, 245 to 476 coding or splice variants per cell line were predicted as somatic variants with deleterious functional impact on the protein. Interestingly, we found ten commonly mutated genes in the cell lines in which apoptosis is observed after X-rays radiation. Their putative role in radioresistance is under investigation. In conclusion, we show that chondrosarcoma cell lines respond differently to X-ray. In addition, our study is the first one which extensively characterizes commonly used human chondrosarcoma cell lines by exome sequencing. Our preliminary results provide essential genetic information on radioresistance mechanisms through the identification of genes potentially involved in the response to radiation.

2645W

Significance of maspin protein expression in prognosis of gallbladder cancer. *K. Baghel[1], HR. Kazmi[1], S. Raj[3], A. Chandra[1], RN. Srivastava[2].* 1) Surgical Gastroenterology , King George's Medical University, Lucknow, Uttar Pradesh, India; 2) Orthopaedics Surgery, King George's Medical University, Lucknow, Uttar Pradesh, India; 3) M S Ramaiah Medical College, Bangalore.

Introduction: Gallbladder cancer (GBC) is highly fatal gastrointestinal malignancy with poor prognosis in northern India. Prognostic significance of Maspin has reported in various cancers but has not been evaluated for GBC. We investigated maspin protein expression in normal, cholelithiasis and GBC patients along with its prognostic importance in cancer patients. **Methodology:** We investigated maspin protein expression in gallbladder tissue of subjects with cholelithiasis (n=36), GBC (n=46) and controls (n = 25). Clinicopathological parameters and prognosis of patients with GBC were correlated with the expression of maspin protein by ELISA. Kaplan- Meier method was performed for survival analysis. **Results:** Significant increased (P<0. 0001) expression of maspin protein was observed in GBC as compared to cholelithiasis, whereas no expression was found in normal tissues. Increased maspin expression was significantly associated with higher tumor stage (stage III vs. stage II; P <0. 0001) and cellular differentiation (poor and moderate vs. well differentiated; P<0. 0001) in GBC. Significant correlation (Pearson's coefficient(r)=-0. 617;P<0. 0001) was observed between maspin protein expression and survival of cancer patients after surgery. Overall survival was significantly shorter (P=<0. 0001; hazard ratio [HR] for death= 2. 84; 95% Confidence interval [CI] = 0. 09865 to 0. 3683) in patients of GBC with maspin expression > 169. 56 pg/mg (median survival; 10 months) as compared to those with expression < 169. 56 pg/mg (median survival; 16 months). Statistically significant difference (P=0. 03) in overall survival was observed for stage II patients with maspin expression > 169. 56 pg/mg (median survival; 11 months) as compared to those with expression < 169. 56 pg/mg (median survival; 18 months). Similarly, for stage III, there was significant difference (P=0. 04) in overall survival between patients with maspin expression > 169. 56 pg/mg (median survival; 9. 5 months) and with expression < 169. 56 pg/mg (median survival; 13 months). The HR for stage II and III disease was 3. 45 (95% CI = 0. 09247 to 0. 9088) and 3. 21 (95% CI = 0. 1010 to 0. 9598) respectively. **Conclusion:** Up-regulation of maspin mRNA may play an important role in malignant progression and is correlated with a poor prognosis. Maspin expression can be useful as a clinical marker in the management of this disease.

2646T

Association of *CDH1* gene polymorphisms -472delA and -160C>A with diffuse and intestinal gastric cancer in Mexican population. *A. R. Bustos-Carpinteyro[1,2], N. Delgado-Figueroa[2], S. Luna-Ernesto[3], M. T. Magaña[2], J. Y. Sánchez-López[2].* 1) Instituto de Genética Humana, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) División de Genética. Centro de Investigación BioMédica de Occidente, Instituto Mexicano del Seguro Social. Guadalajara, Jal. México; 3) Servicio de Endoscopía, Hospital de Especialidades, Centro Médico Nacional de Occidente, Instituto Mexicano del Seguro Social. Guadalajara, Jal. México.

Introduction. Gastric cancer (GC) constitutes the third cause of cancer death in Mexico and worldwide. Due to its histological characteristics, GC can be classified as diffuse and intestinal. It has been observed that diffuse GC has a reduced expression of the E-Cadherin cell adhesion protein, encoded by the *CDH1* gene. A reduction on *CDH1* gene transcription has been reported whenever the mutated alleles -472delA (rs5030625) and -160C>A (rs16260) are present. The aim of this study was to determine the association of *CDH1* -472delA and -160C>A polymorphisms with diffuse and/or intestinal gastric cancer in Mexican patients, as well as to determine the frequencies of these polymorphisms in Mexican healthy population. **Methods.** Four groups of genomic DNA samples were analyzed: diffuse gastric cancer (n=24), intestinal gastric cancer (n=23), controls (n=48), and healthy blood donors (n=93). Genotyping of SNPs was performed by PCR followed by RFLPs through *Ban*II (-472delA) and *Hinc*II (-160C>A) enzymes. The restriction fragments were observed on 6% polyacrylamide gels. The data were analyzed through SPSSv11 and Arlequin 3. 1 statistic programs. **Results.** The frequency of the mutated allele of -472delA SNP was 0. 326, 0. 318 and 0. 284 in the diffuse-gastric-cancer, intestinal-gastric-cancer and control groups, respectively. For the -160C>A SNP, the frequencies of the mutated allele were 0. 174, 0. 318 and 0. 313 for each respective group. Comparison of the genotypic and allele frequencies of both polymorphisms among the three groups shows no significant differences (p>0. 05). Three haplotypes were observed: 1) delA-C: 52. 3%, 42. 5% and 39. 8%, for diffuse-gastric-cancer, intestinal-gastric- cancer and control groups, respectively; 2) delA-A: 18. 2%, 25% and 31. 8%; and 3) A-C: 29. 5%, 32. 5% and 28. 4%, for each group respectively. The haplotypes frequencies were similar among the studied groups. Both polymorphisms showed linkage equilibrium (r2<0. 33, p>0. 05). Hardy-Weinberg Equilibrium was observed for -472delA and -160C>A SNPs in healthy Mexican population. **Conclusions.** The -472delA and -160C>A SNPs are not associated with gastric cancer in patients from western Mexico. However, further studies are required in order to increase our understanding of the role that these polymorphisms play in the development of this disease.

2647W

High Grade Glioblastoma Patients have Increased Levels of Systemic Chromosomal Instability, detected by High Resolution Flow Cytometry in Circulating Reticulocytes. *M. Camargo[1], LM. Barrera-Arenas[1], LD. Ortiz[2].* 1) University Research Center - Biology, University of Antioquia, Medellin, Antioquia, Colombia; 2) Cancerology Institute, Clinica Las Americas, Medellin, Colombia.

The automatic detection of genetic damage by advanced flow cytometry of micronucleated (MN) reticulocytes, has increased the speed and sensitivity of *in vivo* detection of cytogenetic damage. This has been possible through the refinement of the original three color fluorescence labeling technique (Dertinger et al. , 2011) that includes antihuman-CD71-FITC for the immature fraction of reticulocytes, anti-human-CD61-PE to exclude platelets, and PI for nuclear DNA. We have applied this methodology with the aim of exploring chromosomal instability in circulating cells in a population of cancer patients with newly diagnosed high grade glioblastomas (WHO IV), measuring the frequency of MNCD71+ cells immediately before therapy. The study included 50 patients and 25 healthy controls, paired by gender and age matched by decils. Basic confounding factors that may affect the backgrounds levels of genotoxicity were controlled. Samples were analyzed with a flow cytometer FACS CANTO II (BD Biosciences). To our knowledge this is the largest group of patients with glioblastomas in whom, this has been studied. In healthy controls the mean frequency of MNCD71+ was 0. 098 ± 0. 019%. (CI: 0. 058-0. 139%, 95%) while in patients was 3. 060 ± 0. 625 (CI:1. 802-4. 318, 95%). These observations suggest an association between elevated levels of micronucleated reticulocytes and the risk of this particular type of aggressive tumor. They also lso support the intriguing observation made by others about a possible systemic effect of cytogenetic instability associated with a localized tumor. In fact, a number of studies have found elevated levels of MN formation or increased cytogenetic abnormalities in the peripheral blood lymphocytes of cancer patients prior to chemotherapy or radiotherapy. Despite the fact that, to our knowledge, this is the largest group of patients with glioblastomas in whom this has been studied, our findings should be replicated in largest cohorts of patients, carefully looking for parameters such as hidden confounding factors, tumor-derived DNA in circulating blood, oxidative stress in circulating blood and/or in blood cells, among others, before proposing that the MNCD71+ endpoint represents a valuable biomarker for glioblastomas. FUNDING by Sostenibility Grant (01740), University of Antioquia. Note: Authors (1) contributed equally to this work.

2648T

The Prevalence of CYP2D6 Gene Polymorphisms among Filipinos and Their Use as Biomarkers for Cancer Risk among Those with Lung Cancer. *E. Cutiongco de la Paz[1], C. Ngelangel[2], C. Padilla[1], A. Wang[3], J. Nevado[1], R. Ceniza[3], L. Simpao[3], L. Beratio[3], E. Dominguez[3], A. Albay[3], C. Silao[1], F. Rocamora[1], R. Luna[1], The Philippine Cancer Study Group.* 1) Institute of Human Genetics, National Institutes of Health, University of the Philippines Manila; 2) Institute of Clinical Epidemiology, National Institutes of Health, University of the Philippines Manila; 3) Section of Pulmonology, Department of Medicine, College of Medicine, Philippine General Hospital, University of the Philippines Manila.

The highly polymorphic nature of the CYP2D6 gene and its central role in the metabolism of commonly used drugs make it an ideal candidate for pharmacogenetic screening. However, the relevance of such screening depends primarily on the presence of significant variation within a target population. This study aims to determine the prevalence of CYP2D6 polymorphisms among Filipinos. It also aims to determine if such variants are associated with the occurrence of lung cancer, as tobacco-derived procarcinogens are metabolized by CYP2D6 to their carcinogenic derivatives. Utilizing a candidate gene approach, 47 single nucleotide polymorphisms (SNPs) of the CYP2D6 gene were genotyped from DNA samples of 115 cases with lung cancer and age- and sex-matched 115 controls. The samples were genotyped using the Illumina GoldenGate Genotyping assay. Results show that 18 out of 47 polymorphisms have significant genotypic variability (>1% for at least 2 genotypes). No variant is associated with lung cancer. However, rs1135840, rs16947 and rs28360521, were found to be highly variable among Filipinos and these findings may have clinical implications. Both rs1135840 and rs16947 are associated with cancer; rs16947 with timolol-induced bradycardia; and rs28360521 with lose-dose aspirin-induced lower gastrointestinal bleed. This study demonstrates that CYP2D6 polymorphisms are present among Filipinos, which, although not found to be associated with lung cancer, can be useful biomarkers for future pharmacogenetic studies.

2649W

Germline variant analysis and ancestry inference of 2,153 high-depth whole genomes of patients afflicted with diverse cancers. *F. M. De La Vega[1,2], D. Y. Wu. [1], S. Shringarpure[2], S. M. Waszak[3], S. Iakhnin[3], T. Shmaya[1], G. Wojcik[2], C. Gignoux[2], J. Korbel[3], C. D. Bustamante[2], The PCAWG Network.* 1) Annai Systems, Inc. , Burlingame, CA; 2) Stanford University, Stanford, CA; 3) EMBL, Heidelberg, Germany.

Numerous studies have shown a relationship between ancestry, penetrance of disease variants, and effectiveness of therapies. Unfortunately, ethnicity is often not accurate or present in the metadata of cancer research cohorts and thus is ignored in analysis. We aim to explore the relationship between ancestry, cancer etiology and outcomes analyzing 2,153 samples from the TCGA and ICGC projects for which whole genome sequencing data of the germline is available. Samples were sequenced with Illumina technology and realigned to the extended GRCh37 (hs37d5) reference as part the ongoing ICGC PanCancer Analysis of Whole Genomes Project. We performed variant calling using the rtgVariant software, yielding high sensitivity (98. 6%) and non-reference genotype concordance (99. 7%) with SNP genotyping arrays. About 47% of variants are novel as compared with dbSNP and 1000 Genomes Project (v3) data. We developed a maximum likelihood method that uses a previously selected set of 4,235 ancestry-informative-SNPs and their frequency across populations to assign continental ancestry to each donor. For those individuals that appear recently admixed we apply the RFmix software, a fast discriminative modeling approach, to identify ancestral chromosomal segments along the genome. Our results show that most donors are from European or Asian descent in concordance with their source. However, some population diversity is found in particular in the US cohorts. For example, in the US Breast Cancer cohort about a third of donors are from African ancestry, which is highly relevant given the higher prevalence of the triple-negative type of the disease in this population. Other ancestries, such as South-East Asian and Native American, are present although to a much lesser extent. These results highlight the need to recruit other populations for genomic studies to benefit these groups from the advances in personalized medicine. We identified loss of functions (LoF) variants across cancer and other medically relevant genes and stratified by ancestry. A linear relationship exists between cohort size and rate of LoF variant discovery in cancer genes. Interestingly, 58% of cancer genes are found to be mutated in more than two tumor types, suggesting pleiotropic effects. Our variant resource identifies rare human gene knock-outs, enhance imputation panels and the assessment of variant pathogenicity for clinical applications, and is helping to clarify the role of ancestry in cancer outcomes.

2650T

Comparing Heterogeneity across Biomarkers in Cancer Stem Cells and Side Populations in Breast Cancer Cell Lines. *D. Dhawan.* Institute of Life Sciences, Ahmedabad University, Ahmedabad, Gujarat, India, India.

Cancerous epithelial cells are confined to a primary site by the continued expression of adhesion molecules and the intact basal lamina. However, as the cancer progresses some cells are believed to undergo an epithelial-mesenchymal transition (EMT) event, leading to increased motility, invasion and, ultimately, metastasis of the cells from the primary tumour to secondary sites within the body. These disseminated cancer cells need the ability to self-renew, as stem cells do, in order to establish and maintain a heterogeneous metastatic tumour mass. Identification of the specific subpopulation of cancer stem cells amenable to the process of metastasis is highly desirable. In this study, we have isolated and characterized cancer stem cells from luminal and basal breast cancer cell lines (MDA-MB-231, MDA-MB-453, MDA-MB-468, MCF7 and T47D) on the basis of cell surface markers CD44 and CD24; as well as Side Populations (SP) using Hoechst 33342 dye efflux. The isolated populations were analysed for epithelial and mesenchymal markers like E-cadherin, N-cadherin, Sfrp1 and Vimentin by Western blotting and Immunocytochemistry. MDA-MB-231 cell lines contain a major population of CD44+CD24- cells whereas MCF7, T47D and MDA-MB-231 cell lines show a side population. We observed higher expression of N-cadherin in MCF-7 SP cells as compared to MCF-7NSP (Non-side population) cells suggesting that the SP cells are mesenchymal like cells and hence express increased N-cadherin with stem cell-like properties. There was an expression of Sfrp1 in the MCF7- NSP cells as compared to no expression in MCF7-SP cells, which suggests that the Wnt pathway is expressed in the MCF7-SP cells. The mesenchymal marker Vimentin was expressed only in MDA-MB-231 cells. Hence, understanding the breast cancer heterogeneity would enable a better understanding of the disease progression and therapeutic targeting.

2651W

Expression Levels of Matrix Metalloproteinase-9 and P-Element Induced Wimpy Testis Like-2 in Prostate Cancer: A Case-Control Study. *M. Dianatpour[1,2], N. Pouyanfar[1], A. Amin Sharifi[3], A. Monabbati[4].* 1) Medical Genetics, Shiraz university of medical sciences, Shiraz, Fars, Iran; 2) Transgenic technology research center, Shiraz university of medical sciences, Shiraz, Iran; 3) Urology, Shiraz university of medical sciences, Shiraz, Fars, Iran; 4) Pathology, Shiraz university of medical sciences, Shiraz, Fars, Iran.

Prostate cancer is the second most prevalent cancer in human. Also this is the most common malignancy and the sixth important cause of death in men worldwide. The most routine diagnostic test for prostate cancer is PSA test which is associated with some limitations like too many false positive results. This study intends to investigate the role of matrix metalloproteinase-9 (*MMP9*) and P-element induced wimpy testis like-2 (*PIWIL2*) expression levels as different biomarkers in prostate cancer biopsy specimens. Seventy formalin-fixed paraffin embedded samples (35 normal and 35 cancerous cases) were selected. Expression levels of *PIWIL2* and *MMP9* genes were evaluated, using Real-time PCR. *MMP9* and *PIWIL2* expression levels in cancerous tissues were significantly higher than the adjacent normal tissues (P< 0. 05). The survival analysis showed a significant correlation between expression level of *PIWIL2* and survival rate (P<0. 05), but such correlation was not observed in case of *MMP9* (P>0. 05). Higher levels of *MMP9* and *PIWIL2* expression were strongly related to Gleason score and age, using Pearson's correlation coefficient test, however, this kind of association was not evident between PSA and expression levels in the genes of interest. The expression level of *PIWIL2* had a significant correlation with metastasis rate, but this relationship was not seen about *MMP9* expression level. Our results confirmed the validity of *PIWIL2* expression as a valuable prognostic biomarker for early diagnosis of prostate cancer.

2652T

L-1 activity and expression in Normal esophagus, Barrett's esophagus, esophageal adenoarcinoma, and Esophageal Squamous cell carcinoma. *T. T. Doucet[1,2], N. Rodić[3], R. Sharma[3], I. Darbari[1], G. Abril[1], J. T. Choi[1], J. Y. Ahn[1], Y. Cheng[4], R. A. Anders[3], K. H. Burns[1,3], S. J. Meltzer[4], H. H. Kazazian Jr.[1].* 1) Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD. 21205 USA; 2) Pre-doctoral training program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA; 3) Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America Baltimore, MD 21287; 4) Division of Gastroenterology and Hepatology, Department of Medicine and Department of Oncology, Johns Hopkins University, Baltimore, Maryland 21205.

Because retrotransposition has been observed in many gastrointestinal epithelial cancer types, we focused on L1 mobilization as a source of instability in cancer. We hypothesized that L1 retrotransposition is active in esophageal squamous cell carcinoma (ESCC), esophageal adenocarcinoma (EAC), and EAC's precursor Barrett's esophagus (BE). To test our hypothesis, we evaluated 5 patients with benign BE, 5 patients with BE and concomitant esophageal adenocarcinoma (EAC), 10 additional patients with EAC, and 9 patients with ESCC to determine the level of L1 activity in these diseases. Following L1-seq, we confirmed 160 somatic insertions by PCR in 17 of 29 individuals. We observed clonal amplification of several insertions which appeared to originate in normal esophagus (NE) or BE and were later clonally expanded in BE, in EAC, or in ESCC. Additionally, we observed evidence of clonality within the EAC cases: specifically, 22 of 25 EAC-only insertions were present identically in distinct regions available from the same tumor, suggesting they occurred in the founding tumor cell of these lesions. Our data show that somatic retrotransposition occurs early in many patients with BE and EAC, and indicate that early events occurring in histologically normal esophageal cells may be clonally expanded in esophageal adenocarcinogenesis. Additionally, we evaluated L1 ORF1 protein expression in 9 of the carcinoma cases for which formalin-fixed paraffin embedded tissue was available. Using immunohistochemistry, we detected expression of ORF1p in all tumors evaluated. Interestingly, we also observed dim ORF1p expression in the normal esophagus of all 4 patients for whom additional blocks of normal esophagus containing squamous epithelium was available. To determine if ORF1p expression is a hallmark of unaffected tissues, we obtained both skin and esophageal biopsies from two unaffected individuals. In both biopsies, ORF1p expression was evident in the squamous cell epithelium. ORF1p may be expressed in many normal epithelial tissues which could account for the high incidence of somatic retrotransposition events in epithelial cancers. Thus, our data show that L1 is weakly expressed in normal esophagus and retrotransposition can occur in normal tissue during the development of esophageal adenocarcinoma. Due to the pervasive activity of retrotransposons in epithelial cancer, it is likely somatic insertions play a role in tumorigenesis.

2653W

Caveolin-1: A Potential Biomarker of Aggressive Triple-Negative Breast Cancer in African American Women. *J. Getz[1,2], D. Teoh[1,2], S. Nasser[1], C. Legendre[1], W. Tembe[1], V. Yellapantula[1], M. E. Ahearn[1], C. Gomez[3], M. Jorda[3], S. M. Wong[1], M. Pegram[4], J. Carpten[1], L. Baumbach-Reardon[1].* 1) Translational Genomics Research Inst. , Phoenix, AZ; 2) Arizona State University. , Phoenix, AZ; 3) University of Miami Medical School, Miami, FL; 4) Stanford University School of Medicine, Stanford, CA.

In the U. S. , breast cancer (BC) incidence between African American (AA) and Caucasian (CA) women are similar, however, AA women have a significantly higher mortality rate. In addition, AA women often present with tumors at a younger age, with a higher grade and at a later stage and are more likely to be diagnosed with the highly aggressive triple-negative breast cancer (TNBC) subtype. Even within the TNBC subtype, AA women have a worse clinical outcome compared to CA. Although multiple socio-economic and lifestyle factors may contribute to these observed health disparities, it is essential that the molecular characteristics and underlying biological differences between CA and AA TNBC are identified. In this study, gene expression profiling was performed on archived FFPE samples, obtained from CA and AA women diagnosed with early stage TNBC. Initial analysis revealed a pattern of differential expression in the AA cohort compared to CA. Further molecular characterization of the data showed that the AA cohort segregated into 3-TNBC molecular subtypes; Basal-like (BL2), Immunomodulatory (IM) and Mesenchymal (M). Gene expression analyses resulted in 190 differentially expressed genes between the AA and CA cohorts. Pathway enrichment analysis identified Cytoskeletal Remodeling, Cell Adhesion, WNT, Tight Junctions, and Immune Response in the AA TNBC-Cohort and significantly upregulated genes associated with the Wnt/β-catenin pathway. These results were validated using qRTPCR on an independent cohort of FFPE samples from AA and CA women with early stage TNBC, and identified Caveolin-1 (CAV1) as being signficantly expressed in the AA-TNBC cohort. Furthermore, CAV1 was shown to be highly expressed in a cell line panel of TNBC, in particular, those of the mesenchymal and basal-like molecular subtype. Finally, a loss of function study by siRNA revealed that silencing of CAV1 resulted in a significant decrease in proliferation in each of the TNBC cell lines. These observations suggest that CAV1 expression may contribute to the more aggressive phenotype observed in AA women diagnosed with TNBC.

2654T

Chromosomal imbalances detected by array CGH from thirty-one Chinese non-smoker adenocarcinomas of the lung. *Y. Gu[1,2], L. Peng[1], X. Wang[2], W. Li[2], S. Li[2], S. Hua[1].* 1) Department of Reparatory and Critical Care Medicine, The First Hospital of Jilin University, Chang Chun City, Ji Lin Province, P. R. China; 2) Department of Pediatrics, College of Medicine, The University of Oklahoma Heath Sciences Center, Oklahoma City, OK 73104, US.

Smoking tobacco is the number one risk factor for lung cancer. However, people who never smoked can also develop lung cancer. It is unclear whether chromosomal changes in these two groups of patients with smoking or non-smoking history have the same or different genetic profile at the cytogenetic level. In this pilot study, we have collected 31 tumor samples from patients diagnosed with lung adenocarcinoma in clinical phase I-IIa who claimed they had never smoked before. They were 23 female patients with a mean age of 57. 4 years old and 8 male patients with a mean age of 62. 4 years old. Samples of DNA were isolated from the tumor tissues and aCGH was performed using Agilent's array platform Sureprint 2x400K. Approximately 81% of the samples showed a gain of whole chromosome 5 or partial duplication of the short arm of chromosome 5 and 77% of the samples showed a gain of whole chromosome 7 or a duplication of short arm of chromosome 7. Duplication of the long arm of chromosome 8 (61%), deletion of the short arm of chromosome 8 (52%), deletion of the short arm of chromosome 9 (41%), duplication of the long arm of chromosome 14 (48%) and duplication of short arm of 16 (48%) were also detected. Compared to common deleted or duplicated regions of adenocarcinoma of the lung published previously, these findings indeed had unique patterns. Whether chromosomal imbalances found in our cohort of patient samples were contributed by the fact that our samples were from non-smoker adenocarcinomas of the lung or not, more patients were needed to draw a definitive conclusion.

2655W

The landscape of microsatellite instability in cancer exomes. *R. J. Hause[1], E. H. Turner[2], M. Beightol[2], C. C. Pritchard[2], J. Shendure[1], S. J. Salipante[2].* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Laboratory Medicine, University of Washington, Seattle, WA.

Microsatellites, 2-5 base pair repetitive sequences present throughout the human genome, can abnormally shorten or lengthen because of defects in the DNA mismatch repair (MMR) system, resulting in a "microsatellite instability" (MSI) phenotype. MSI is a key prognostic and diagnostic tumor phenotype that has been well studied by conventional methods. However, both the genomic landscape of MSI events and differences in MSI among cancer types remain poorly illuminated. We here present a comprehensive analysis of the landscape of MSI in cancer exomes. We catalogued MSI events at over 500,000 incidentally sequenced microsatellite loci across 4,224 cancer exomes spanning 18 different cancer types from The Cancer Genome Atlas. We constructed a global classifier for MSI that achieved 93. 75% sensitivity and 98. 5% specificity compared to gold-standard MSI calls based on the revised Bethesda guidelines. We observed that MSI-low (MSI-L) samples did not display significant differences from MS-stable (MSS) samples in the number of MSI events and support discontinuation of the use of MSI-L as a distinct classification. Comparative examination of MSI revealed both cancer-specific and core loci associated with global MSI, such as a frameshift mutation in an 8-bp polyadenine tract in exon 10 of the tumor suppressor *ACVR2A* observed in 56. 6% of MSI-high (MSI-H) cancers. Lastly, we investigated the relationships between MSI and mutations in MMR genes, gene regulatory features, and clinical covariates. Our results provide a comprehensive view of MSI in cancer exomes, highlighting both conserved and cancer-specific MSI properties and identifying candidate genes underlying predisposition to global MSI. Future work will attempt to functionally validate these candidates as causally influencing global MSI.

2656T

Cell-free urinary microRNA quantification as non-invasive biomarker in patients with bladder urothelial carcinoma. *A. Horinek[1,2], M. Korabecna[1,2], A. Brisuda[2], S. Pospisilova[1], E. Pazourkova[1,3], V. Soukup[4], J. Hrbacek[2], O. Capoun[4], I. Svobodova[1], J. Mares[5], T. Hanus[4], M. Babjuk[2].* 1) Institute of Biology and Medical Genetics, 1st Faculty of Medicine, Charles University and General Faculty Hospital, Prague 2, Czech Republic; 2) Department of Urology, 2nd Faculty of Medicine, Charles University and University Hospital Motol in Prague, Czech Republic; 3) Department of Nephrology, 1st Faculty of Medicine, Charles University and General Faculty Hospital in Prague, Czech Republic; 4) Department of Urology, 1st Faculty of Medicine, Charles University and General Faculty Hospital in Prague, Czech Republic; 5) Department of Biology and Medical Genetics, Charles University in Prague, 2nd Faculty of Medicine, Czech Republic.

Background: Bladder carcinoma is one of the most common urooncological diseases. It is the sixth oncological disease and it makes 4% of newly diagnosed oncological illnesses in the Czech Republic. The early noninvasive detection contributes to decreased mortality rate in this disease. **Aim:**MicroRNAs (miRNAs) are small non-coding RNA molecules that function in the regulation of gene expression. The aim of the study was to detect candidate miRNAs in urine whose expression could reflect the onset and progression of bladdder cancer and that may be used as a potential diagnostic marker for early noninvasive diagnostics and for prediction of the severity of bladder cancer, as well. **Patients and methods:**Urine samples were collected in the period 10/2013 – 4/2015 and 61 samples were included in this study (15 healthy controls and 46 patients in different stages of bladder cancer (pTa - pT4)). The samples with positive culture were not included. In the first part of our study 381 miRNAs from urine were analyzed on TaqMan Human MicroRNA Array A (TLDA) by Real-time PCR method. In the second part TaqMan MicroRNA Assays of selected miRNAs from the TLDA platform were used to confirm the results. Relative quantification and statistical analyses were performed with Expression Suite v1. 0. 3 and with qBase+ v2. 4. Relative quantification was calculated using 3 reference genes (miR-191, miR-28-3p, miR-200b). Expressions of miRNAs were compared between the patients and controls using the Mann-Whitney test. **Results:**In the first part we identified 58 miRNAs whose expression was significantly different ($p<0.05$) between controls and patients. According the results of both methods we found miRNAs with statistically significant differences: 9 with $p<0.01$ (let-7c, miR-125b, miR-204, miR-425, miR-532-3p, miR-99a, miR-16, miR-30b and miR-93) and 1 with $p<0.05$ (miR-199a-3p). Others were not confirmed in the second part. 6 miRNAs (let-7c, miR-125b, miR-204, miR-532-3p, miR-99a, miR-30b) were 4-10 times more expressed in controls than in patients and 4 were 2-4 more expressed in patients (miR-425 miR-16 miR-93, 199a-3p). **Conclusion:**It has been demonstrated that the expression of six candidate miRNAs with significant differences in expression in urine of patients and controls could be promising diagnostic and prognostic markers for noninvasive diagnostics of bladder cancer. **Acknowledgements:**Supported by the Internal Grant Agency of the Ministry of Health of the Czech Republic no. NT12417.

2657W

Using Rare Mutation Analysis Using Digital PCR on QuantStudio® 3D to Validate Ion AmliSeq™ Next Generation Sequencing for Cancer Research. *M. Laig[1], A. Justino[2], R. Petraroli[1], L. Lac[1], D. Linh Vo[1], K. Varma[1], J. C. Machado[2], J. L. Costa[2].* 1) Thermo Fisher Scientific, South San Francisco, CA; 2) IPATIMUP, Porto, Portugal.

Introduction: Next generation sequencing (NGS) is becoming an increasingly important analysis tool in cancer research. In this study, we set out to confirm its sensitivity and accuracy detecting low frequency mutations using rare mutation analysis by digital PCR as an independent technology. **Methods:**Eleven cell free DNA samples isolated from plasma were typed using the Ion AmliSeq™ panel and the Ion Proton PGM. The samples were then analyzed by digital PCR in a blinded laboratory study to confirm the presence of the mutations detected. Samples represented a mixture of independent samples as well as the same sample taken from the same individual at different time points. Mutations were identified and quantified by digital PCR using Rare Mutation Assays (Applied Biosystems) for the following mutations: EGFR T790M (COSMIC ID 6240), L858R (COSMIC ID 6224), and exon 19 deletions E746_A750delELREA (COSMIC ID 6223) and L747_P753>S (COSMIC ID 12370), and KRAS mutations G12V (COSMIC ID 520) and G12D (COSMIC ID 521). Experiments were run according to the Rare Mutation Assays protocol. **Results:**Rare mutation analysis by digital PCR on QuantStudio® 3D confirmed the low mutation frequency detected by NGS. Both technologies showed matching results for all eleven samples tested at mutation rates below 1%. Both technologies also detected expected mutation rates based on the time points. **Conclusions:**We confirmed the sensitivity and accuracy of NGS for low frequency mutation detection. Rare mutation analysis by digital PCR on QuantStudio® 3D is an excellent application to validate NGS results for cancer research. For Research Use Only. Not for use in diagnostic procedures.

2658T

Targeted RNA Sequencing for simultaneous expression profiling and detection of gene rearrangements in FFPE biopsies. *K. Maddula, P. Roche, D. Thompson.* HTG Molecular, Tucson, AZ.

Recent developments in targeted cancer therapies have highlighted a need for quick, accurate gene mutation and rearrangement assays that work well with small biopsy specimens. Routine screening for rearrangements is complicated by the size of biopsy samples (fine needle aspirates or bronchoscopic biopsies) and the low prevalence of rearrangements. Ideally, fusion detection would be done in conjunction with routine pathology, using the same sample. We have developed HTG EdgeSeq chemistry and system, a coupling of nuclease protection with next generation sequencing (NGS) - mediated quantification. Library preparation occurs in two simple steps: nuclease protection, followed by limited PCR cycles to prepare libraries for NGS. No RNA extraction from FFPE is necessary. After sequencing, probe sequences are aligned to the sequencing reads, making data processing and fusion detection simple and very quick. To demonstrate utility of this assay for simultaneously detecting rearrangements and tumor subtyping markers, we are developing a HTG EdgeSeq lung fusion assay, a combination lung fusion and lung cancer subtyping assay. FISH is currently the gold-standard method for detecting ALK and RET rearrangements. We show here three examples from a small methods-comparison study. In the first two examples, the HTG EdgeSeq lung fusion assay and FISH assay both provided positive results (one sample is RET-positive, one ALK-positive). Interestingly, the third example failed two test attempts by an accredited ALK FISH testing lab due to poor sample quality; however, this sample tested positive for ALK rearrangement using the HTG EdgeSeq lung fusion assay. We confirmed the HTG EdgeSeq assay results by qPCR after isolating RNA from the FFPE block. This is significant because many clinical specimens do not contain enough tissue to test via qPCR, resulting in a meaningful ALK rearrangement potentially going undetected. In summary, the HTG EdgeSeq lung fusion assay is a promising technique for detecting gene rearrangements in small amounts of fixed tissue.

2659W

Genetic risk assessment for lung squamous cell carcinoma using exome sequencing in Korean men. *J. W. Park[1], E. P. Hong[1], Y. Kim[2], J. Jung[3], K. S. Park[4], K. Park[2], H. Kim[5].* 1) Dept Medical Genetics, Hallym Univ Coll Med, Chuncheon, Gangwon-Do, Republic of Korea; 2) Div Hematology, Samsung Medical Center, Seoul, Republic of Korea; 3) Syntekabio Inc, Seoul, Republic of Korea; 4) Dept Internal Medicine, Soul Natn'l Univ Coll Med, Seoul, Republic of Korea; 5) Dept Cell Biology, Ewha Womans Univ Coll Med, Seoul, Republic of Korea.

Lung squamous cell carcinoma (LSCC), the most common type of non-small cell lung cancer, is a leading cause of cancer deaths worldwide. Previous candidate gene and genome-wide association studies identified multiple single-nucleotide polymorphisms (SNPs) associated with lung cancer, each of which has a small to moderate effect. We performed an exome-wide association study to identify germline mutations by comparing100 Korean men diagnosed with LSCC to 131 age matched diabetic men without cancer. We developed LSCC risk prediction models composed of the genetic variants confirmed their functions by pathway analysis. Of 107,758 exonic variants, 158 nonsynonymous SNVs (nsSNVs) and seven insertions/deletions located in 102 candidate genes showed an exome-wide significance ($p < 4.6\times10\text{-}7$). A SNV (rs141472183, S209C) and a frameshift deletion (rs45576436) located in the genes, *HNRNPCL2* (OR=129, p=1. 4×10-41) and *HLA-A* (OR=191. 4, p=1. 4×10-35), showed the most significant association with LSCC. Our findings suggest that the genes, *HLA-A*, *MUC5B*, *TAS2R46*, and *TAS2R43,* play key roles in the primary tumor-cell growth of the epithelial surfaces and the dysfunctions of the pulmonary system. A genetic model composed of four nsSNVs and one deletion located in the genes that are associated with lung function showed 99. 9% predictabilityfor LSCC risk. Our findings will provide a foundation for future researches on early genetic diagnosis and risk prediction of LSCC susceptibility.

2660T

Cancer Genomic resources, opportunities and needs in the Latin American region. *S. Perdomo[1], J. Olivier[2], F. Vaca[3, 4].* 1) Instituto de Investigación en nutrición, Genética y metabolismo IINGM , Universidad El Bosque, Bogotá, Colombia; 2) Instituto de Ciencias Básicas y Medicina Experimental. Instituto Universitario Hospital Italiano de Buenos Aires Ciudad Autónoma de Buenos Aires, Argentina; 3) Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México Tlalnepantla, Estado de México, Mexico; 4) Instituto Nacional de Cancerología, Estado de México, Mexico.

In 2012, approximately 1,005,255 new cancer cases and 550,164deaths from cancer occurred in both sexes in central and South American region, by 2025 and increment of nearly 30% of new cases and 35% of deaths from cancer are predicted. One of the priorities of cancer control within the region is to reduce avoidable deaths from cancer by improving early detection and personalized treatment. The development of cancer genomics and their integration to cancer care has shown a great improvement in cancer control worldwide; however, all these key advances have been mainly concentrated in highly developed nations and little is known about the capacities and needs of cancer genomics in the Latin-American (LA) context. In order to evaluate the capacity and development of cancer genomic applications in the LA we collected available information for all countries in central, South America and Cuba. Data reviewed included: number of cancer research institutions, number of NGS platforms, research groups working in cancer genetics, publications on cancer genetics for the last 10 years, educational programs on genomics and related health policies. In average 15 Research groups per country were registered as conducting cancer-genetics related projects. In the las 10 years, 206 publications on cancer genetics including reviews were led by (1st, 2nd or corresponding) authors affiliated to LA institutions and only publications in the last 2 years included cancer genomic analysis. Most publications related to breast and gastric cancer and hematological cancers were underrepresented in general. In the past 5 years physical resources have grown markedly by the acquisition of nearly 150 NGS platforms in 9 countries, the majority installed in medical related services and universities. Participation in large genomic consortia has only included two countries Mexico (as partner) and Brazil (as leading). Educational programs in genomics are scarce, almost exclusive of graduate programs and few are applied to cancer. Despite the recent advances in introducing cancer genomics knowledge and application in LA, the region lacks development of integrated genomic research projects, improved use of platforms dedicated to cancer, educational programs and health policies that might focus on the most frequent cancers and could impact cancer care.

2661W

Evaluation of four putative genetic risk factors affecting breast cancer susceptibility in patients with sporadic breast cancer. *M. Saeidnejad[1], A. Khorrami[2], E. Abedini[1], M. Mohaddes[2], M. Ziadi[2].* 1) Azad Islamic university of Tabriz, Tabriz, Iran; 2) Tabriz University of Medical Sciences.

Breast cancer is a complex multifactorial disease. Along with environmental factors genetic variations play a significant role in breast cancer Etiology. Single nucleotide polymorphisms (SNP) were highlighted as the key variations leading to genetic differences in breast cancer susceptibility between individuals. The methylene tetra hydro folate reductase enzyme (MTHFR), regarding its key role in DNA synthesis and DNA methylation, and hereditary hemochromatosis gene (HFE), coding iron-regulatory protein, have been associated with neoplastic disease development. The current study represents the analysis of A1298C, C677T (MTHFR) and H63D, C282Y (HFE) polymorphisms possible association with breast cancer risk in Iranian Azeri Turk women with sporadic breast cancer. A total of 80 cases and 100 controls participated in the study. Genotyping the markers in MTHFR gene were done using ARMS-PCR technique and in HFE gene using PCR-RFLP technique. The chi-square test with Yates' correction and Fisher's exact test were used for statistical analysis. Among the investigated polymorphisms, frequency of *HFE* H63D showed significant difference between patients and healthy volunteers (P = 0. 04), while allele and genotype distribution of A1298C, C677T and C282Y were similar in all groups. Our results suggest that *HFE* H63D mutation is associated with higher risk of sporadic breast cancer in studied population.

2662T

Contribution of molecular changes to Uveal Melanoma tumorigenesis in a Brazilian cohort. *H. Sarubi[1], F. Melo[1], E. Friedman[2], L. Rodrigues[3], M. Pedrosa[4], N. Amaral[5], A. Carmo[5], C. Gomes[5], R. Gomez[6], L. De Marco[1].* 1) Surgery, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; 2) The Susanne Levy Gertner Oncogenetics Unit, Chaim Sheba Medical Center, Tel-Hashomer, Israel; 3) Department of Basic Sciences, Universidade Federal de Juiz de Fora, Brazil; 4) Department of Pathology, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; 5) Department of Pathology, Biological Sciences Institute, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; 6) Department of Oral Surgery and Pathology, School of Dentistry, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

Introduction: Uveal melanoma (UM) is the most common primary cancer of the eye in adults. Despite the availability of high effective treatments for eradicating the primary tumor, up to half of affected individuals later develop metastatic disease that is almost always fatal within a few months. The discovery of several common driver mutations has oppened the door to rational targeted therapies . BRCA2 has a predominant tumor suppressor function thought to be the repair of DNA double-strand breaks through homologous recombination. BRCA2 mutation carriers have elevated risk of breast and ovarian cancers and its mutation has also been implicated in uveal melanoma etiology, but the impact that molecular alterations in this gene have on uveal tumourigenesis is still unclear. The purpose of this study was to investigate BRCA2 contribution to the disease and to correlate these findings with clinical and histopathological parameters, such as P53 expression profile, since the tumour suppressor TP53 is frequently mutated in BRCA2 cancers and studies have shown a genetic interaction between them resulting in an increased constitutive DNA damage. **Methods:** 32 formalin-fixed paraffin-embedded tissues from enucleated eyes of UM patients were immunostained for BRCA2 and P53 and sequenced for Uveal Melanoma signature genes, such as GNA11, GNAQ and SF3B1. **Results:** Generally, immunostaining for p53 was weak (suggesting a lack of p53 mutations), but 05 of the 32 patients (15,62%) showed a markedly p53 positivity which has to be correlated with clinic and histopathologic parameters. Strong cytoplasmic staining was observed for BRCA2, which can be due to truncated forms of the protein, that do not translocate to the nucleus, and can be involved in the pathogenesis of the disease. **Conclusion:** Our data shows that despite BRCA2 and p53 are mutated infrequently in uveal melanoma, as shown in recent studies, their respective pathways may be functionally inactivated in the disease. Therefore, further studies are needed to investigate and help to define the real contribution of these pathways to tumorigenesis in Uveal Melanoma patients.

2663W

Mitochondrial DNA variation as a biomarker for the development of radiation-induced lung toxicity. *H. Smeets[1,2], A. Voets[1,2], C. Oberije[3], G. Nalbantov[3], A. Stassen[1], A. Hendrickx[1], K. Vandekasteele[5], K. Deruyck[5], H. Thierens[5], Y. Lievens[4], C. Herskind[6], P. Lambin[2,3].* 1) Dept Clin Genet, Maastricht UMC, Maastricht, Netherlands; 2) Maastricht University, GROW School for Oncology and Developmental Biology, Maastricht, The Netherlands; 3) MAASTRO Clinic, Data center, Maastricht, The Netherlands; 4) Ghent University Hospital, Radiotherapy, Ghent, Belgium; 5) Ghent University, Basic Medical Sciences, Ghent, Belgium; 6) Universitàtsmedizin Mannheim, Labor für Zelluläre und Molekulare Radioonkologie, Mannheim, Germany.

Purpose/Objective: Radiation-induced lung toxicity (RILT) varies significantly between patients at similar doses to the lung and can seriously affect the quality of life. The identification of prognostic biomarkers for radiation-induced toxicity is crucial for personalized RT: to select patients for proton therapy or dose escalation. We hypothesized variation in the mitochondrial genome (mtDNA) is a biomarker for RILT, since mitochondria and RT have several processes in common, among which reactive oxygen species (ROS) production. **Materials and Methods**: Blood DNA was isolated from 372 (training set, Maastricht) and 68 (test set, Ghent) lung cancer patients. After exclusion of patients that had surgery, had other tumors within 5 years prior to lung cancer, received a palliative dose or for which baseline dyspnea score was unknown, 277 and 53 patients were remaining for the training and test set respectively. Baseline dyspnea (at the start of RT) and maximal dyspnea 3-6 months after RT were scored according to the CTCAE version 3. 0 criteria. The endpoint of analysis was dyspnea grade>=2 after RT. Additionally, DNA was obtained from fibroblasts of 21 breast cancer patients for which the toxicity grade after radiotherapy was known (LENT/SOMA criteria). mtDNA was resequenced using mitochips and homoplasmic deviations from the revised Cambridge reference sequence were recorded. Variants were classified into 7 functional categories based on their theoretical effect on OXPHOS function. **Results**: Using the 7 functional categories as input features, logistic regression analysis corrected for baseline dyspnea score resulted in an AUC of 0. 78 for the training set, which was significantly better than the current international gold standard Mean Lung Dose (AUC 0. 57; $p < 0.001$). The AUC for the test set was 0. 66 but the power of the validation was limited due to the small sample size. Validation in a second external test set is ongoing. Additionally, using mtDNA variation data we were able to classify breast cancer patients in the correct toxicity group with 80% accuracy. **Conclusions**: Our data showed that mtDNA variation is a valuable biomarker for RILT. Furthermore, we have preliminary data in breast cancer patients that the predictive effect of mtDNA might be applicable to radiation toxicity in general.

2664T

Metastatic Signature Profiles in Triple Negative Breast Cancer. *K. Upadhyay[1], A. Pearlman[1], MT. Rahman[1], J. Loke[1], S. Fineberg[2], H. Ostrer[1].* 1) Pathology, Albert Einstein College of Medicine, Bronx, NY; 2) Pathology, Montefiore Medical Center, Bronx, NY.

Breast cancer is the most commonly diagnosed female cancer and the second leading cause of cancer deaths among women in the United States. Among the 3 clinical breast cancer subsets defined by immunohistochemistry, ER+ (65%-75%), HER2+ (15%-20%) and Triple Negative (TN 15%), TN have the highest rate of metastases. Previously, we developed a metastatic potential score (MPS) based on copy number alterations in a specific repertoire of 366 genes in metastatic prostate cancers and observed that this signature is predictive of metastastic risk in primary prostate cancers from surgical cohort (area under the receiever-operator charasteristic curve –AUC ~80%)(Pearlman et. al. , Journal of Probability and Statistics, vol. 2012, Article ID 873570, 19 pages, 2012). To examine the utility of the MPS in triple negaive breast cancer(TNBC), we performed array-CGH on 41 surgical specimen (28 metastatic-prone primaries - mBCs, 13 indolent primaries - iBCs) on Affymetirx Oncoscan V2 arrays. The median of MPS was lower for iBCs (0. 71) than mBCs (1. 00, p-value = 0. 01). A univariate logistic regression model of MPS predicted an odds ratio of 22. 59 (p-value = 0. 02) with an AUC of 75%. A univariate Cox regression model predicted a hazard ratio of 3. 8 (p-value = 0. 02) with Concordance-index 0. 63. This study demonstrates that like prostate cancer, TNBC is a disease of genomic instability, that amplification of a similar repertoire of genes is involved in disease progression, and that MPS is predictive of metastatic risk.

2665W

Association of Wnt signaling pathway gene variants in gallbladder cancer susceptibility, therapeutic response and survival. *A. Yadav[1], B. Mittal[1], A. Gupta[1], N. Rastogi[1], S. Agarwal[1], A. Kumar[1], V. Kumar[2].* 1) Sanjay Gandhi Post Graduate Institute of medical Sciences, Lucknow, India; 2) KGMC, Lucknow, India.

Background: Gallbladder cancer (GBC) is the most common malignancy of the biliary tract with adverse prognosis and poor survival. Wnt signals controls human development and homeostasis by cell-cell communication. Disregulations caused by germline genetic variants in Wnt pathway genes are cognate in development and expansion of cancers. This study analyzed association of Wnt signaling pathway genes (GLI-1, SFRP2, SFRP4, DKK2, DKK3, WISP3, APC, B-Catenin, AXIN-2) variants in GBC genetic predisposition, treatment outcomes and survival. **Materials and Methods:** The study included 610 GBC patients and 250 controls. Out of 610, 200 patients were followed-up for treatment response and survival. Toxicity grading was done in 200 patients (CTCAE V 3. 0) and tumor response (RECIST 1. 1) was recorded in 140 patients undergoing non-adjuvant chemotherapy (NACT). Genotyping was done by using PCR-RFLP, ARMS-PCR and Taqman assays. Statistical analysis was done by binary logistic regression, SNPstats, CART and GMDR. Survival was assessed by Kaplan Meier survival curve and multivariate Cox-proportional hazard model. **Results:** Single locus analysis showed statistically significant association of GLI-1 rs222826C/G [p-value=0. 001], SFRP4 rs1802073G/T [p-value= 0. 004], AXIN-2 rs4791171C/T [p-value= 0. 001], AXIN-2 rs2240308G/A [p-value=0. 003], B-catenin rs4135385A/G [p-value=0. 037], APC rs4595552 A/T [p-value=0. 021] with increased risk of GBC. Haplotypes of APC Trs459552Trs11954856 [p-value = 0. 0450] and AXIN-2 Trs4791171Ars2240308 [p-value = 0. 039] significantly associated with increased risk of GBC susceptibility. GMDR predicted GLI-1 rs2228226, APC rs11954856 [p-value=0. 0054] as significant model with GBC susceptibility. For treatment response and gastrointestinal toxicity GLI-1 rs2228226, AXIN-2 rs4791171 [p-value= 0. 0107, 0. 001] predicted as significant model. CART analysis represents APC rs4595552 (AT), APC rs4595552 (TT) higher risk genotype for GBC susceptibility and GLI-1 rs2228226 (GG), DKK3 rs3206824 (CC) with poor prognosis. Cox-proportional hazard model showed GLI-1 rs2228226 CG/GG genotype higher mortality and hazard ratio in post-operative and locally advanced GBC cases. **Conclusion:** GLI-1 rs2228226 emerged as major genetic variant in WNT signaling pathway influencing susceptibility, prognosis and survival in gallbladder cancer. Financial support from DBT, DST and CSIR, India.

2666T

Understanding Lung Adenocarcinoma Morphology and Prognosis by Integrating Omics and Histopathology. *K. Yu[1,2], D. L. Rubin[1], M. P. Snyder[2].* 1) Biomedical Informatics, Stanford University, Stanford, CA; 2) Department of Genetics, Stanford University, Stanford, CA.

Lung cancer causes more than 1. 4 million deaths per year worldwide, and adenocarcinoma accounts for 40% of lung cancer. For decades, histopathology evaluation has been the definitive diagnostic method for the disease. However, the molecular mechanisms underpinning histopathological patterns are not fully understood and prognosis prediction through histopathological sections is limited and subjective. In this study, we aimed to identify the transcriptomic and proteomic activities associated with histopathology grades and establish models for prognosis prediction. We analyzed transcriptomic and proteomic profiles of 480 lung adenocarcinoma patients from The Cancer Genome Atlas, and extracted histopathology annotations from their pathology reports. We divided the data into distinct training and test sets, built random forests models with training set data to predict the dichotomized histopathology grades (low grade versus moderate to high grade), and evaluated their areas under receiver operating characteristics curves (AUC) on the test set. We also built a least absolute shrinkage and selection operator (LASSO)-Cox proportional hazards model to predict survival outcomes. Enrichment analyses of the genes were conducted by querying the Database for Annotation, Visualization and Integrated Discovery (DAVID). Leveraging transcriptomic information, we successfully predicted histology grades of the tumor samples with AUC more than 0. 85 on the independent test set with 15 gene expression features. We identified important biological pathways enriched in the most predictive features, such as mitosis and ubiquitin-like proteins conjugation. Incorporating histopathology and omics variables, our model with three features classified patients into two distinct prognostic groups, with statistically significant difference in their survival time in the test set (P=0. 02). Gene ontology terms enriched in the survival-associated genes included extracellular region and cellular defense response. We identified the biological processes associated with pathology grades of lung adenocarcinoma, and accurately predicted patients' prognosis by omics and histopathology features. The analytical approaches described here reveal the underlying molecular mechanisms of cancer, and help predict prognosis. Our approaches can be applied to other types of tumors and will greatly contribute to establishing personalized cancer treatments.

2667W

Single gene (SG) vs. Multi-gene panel (MGP) testing for *TP53* germline mutations in Li Fraumeni syndrome (LFS). *H. Q. Rana[1], R. Gelman[1], J. Thompson[2], R. McFarland[2], H. LaDuca[2], E. Dalton[2], V. Speare[2], J. S. Dolinsky[2], E. Chao[2,3], J. E. Garber[1].* 1) Dana-Farber Cancer Institute, Boston, MA; 2) Ambry Genetics, Aliso Viejo, CA; 3) University of California, Irvine, Irvine CA.

Background:LFS is a rare highly penetrant hereditary cancer syndrome associated with pathogenic germline mutations in the*TP53*gene. Traditionally, testing of*TP53*has been limited to individuals and families who meet well-established criteria. With the advent of MGP tests,*TP53*analysis has expanded beyond these selected groups. **Methods:** *TP53*testing via SG and 8 different MGP by Ambry Genetics for 25,182 patients were collected. Retrospective review of personal and family cancer histories of those with a pathogenic or likely pathogenic*TP53*alteration (*TP53*+) was conducted, with solicitation of more complete family cancer histories ongoing. Cancer histories were examined to determine whether any National Comprehensive Cancer Network (NCCN) testing criteria were met including Classic criteria, Chompret criteria, and breast cancer (BC) diagnosis <36 years. **Results:**Of 25,182 individuals, 187 (0. 74%were *TP53*+; these results came from SG (118 of 2956, 3. 99%) and from MGP (69 of 22,226, 0. 31%). Of all those tested with adequate data, 95% having SG had a personal cancer history versus 82% of those having MGP. Among 102 *TP53*+ by SG providing adequate family history data, 73% (95% CI 63%-81%) met Classic or Chompret criteria for LFS, compared to 30% (95% CI 19%-47%) of 66 *TP53*+ by MGP (p=0. 0000001); adding in the cases meeting meeting the BC<age 36 criterion means 85% of those*TP53*+from SG and 53% of those*TP53*+on MGP met at least 1 of the 3 NCCN testing criteria. The personal BC<age 36 criteria was the only 1 of the 3 criteria met for 29% of 48*TP53*+on 4 MGP tests related to women's cancers versus 6% of 18*TP53*+on the other 4 MGP tests (p=0. 05). **Conclusion**: This is the largest cohort of*TP53*mutation carriers reported from one testing laboratory to date. MPG testing enables the identification of*TP53*mutations in individuals who would not otherwise have been tested by established LFS testing criteria. Further study is needed to determine whether patients and kindreds ascertained by SG versus MGP have truly different or similar LFS manifestations. Ultimately, these findings may alter testing criteria for *TP53* and counseling of those with a molecular diagnosis of LFS.

2668T

Biopsy-free comprehensive genomic profiling of over 5,000 cancer patients using a CLIA-certified commercial cell-free DNA next-generation sequencing test. *S. Mortimer[1], O. Zill[2], J. Vowles[1], R. Lopez[2], D. Delubac[6], K. Dilger[1], R. Mokhtari[2], W. Chen[2], S. Bakhtiari[2], C. Jackson[2], T. Vo[2], B. Kermani[2], K. Banks[3], R. Nagy[3], A. Baca[4], R. Lanman[3], H. Eltoukhy[5], A. Talasaz[5].* 1) Technology Development, Guardant Health Inc, Redwood City, CA; 2) Software and Informatics, Guardant Health Inc, Redwood City, CA; 3) Medical Affairs, Guardant Health Inc, Redwood City, CA; 4) Clinical Lab, Guardant Health Inc, Redwood City, CA; 5) Administration, Guardant Health Inc, Redwood City, CA; 6) Process Engineering, Guardant Health Inc, Redwood City, CA.

Analysis of cell-free circulating tumor DNA by next-generation sequencing (NGS) allows non-invasive real-time profiling of actionable genomic alterations. Current NGS cancer diagnostics based on tissue biopsies provide an incomplete picture of genomic alterations, particularly in metastatic cancers, due to tumor heterogeneity and continuous evolution of disease. Moreover, the widespread use of cell-free DNA (cfDNA) sequencing has been limited by frequent false positives present in standard NGS assays, especially given that the tumor-derived fraction of cfDNA is often below 1%. The Guardant360® test has been clinically validated to detect SNVs, indels, focal gene amplifications, and gene fusions in a 68-gene panel (138 kb), and was designed to include all NCCN-recommended genomic targets for matched therapies. Notably, this is the first cfDNA NGS assay to detect fusions without prior knowledge of both fusion partners. We demonstrate that our Digital SequencingTM methodology through the detection of single molecules achieves high sensitivity (>85%) and ultra-high specificity (>99. 99%). Variants identified in cfDNA with the Guardant360 test were verified by several outside CLIA-certified laboratories using various methodologies. Across 56 patients with a total of 86 variants (SNVs, CNVs, indels, and fusions), 90% of variants found in tissue were also detected in cfDNA. We present an analysis of cfDNA genomic alterations identified in over 5,000 cancer patients across more than 50 cancer types. 74% of patients were positive for at least one class of genetic variant, with a median of 3. 5 SNVs per patient, and median mutant allele fraction of 0. 4%. Of these patients, 15% had mutations indicating an approved therapy and 50% had mutations indicating an off-label therapy. On average we detect nearly one somatic variant of uncertain significance (VUS) per cancer patient. Although the test is optimized for late-stage/metastatic disease, it is also capable of earlier stage detection. We demonstrate detection of somatic SNVs, indels, and gene fusions in cfDNA at mutant allele fractions below 0. 05%. The present work shows the strong clinical impact of Guardant360 test for analysis of cfDNA, thereby allowing researchers and clinicians to comprehensively and non-invasively monitor the genomic profile of cancer throughout the body. Guardant360 technology empowers oncologists to make informed treatment decisions, especially when repeat tissue biopsy is not a feasible option.

2669W

Genomic variations in plasma cell free DNA differentiate early stage lung cancers from normal controls. *S. Xia[1,2], CC. Huang[3], M. Ie[1], RL. Dittmar[1], MJ. Du[1], TZ. Yuan[1], YC. Guo[1], Y. Wang[1], XX. Wang[3], S. Tsai[4], S. Suster[1], AC. Mackinnon[1], L. Wang[1].* 1) Department of Pathology and MCW Cancer Center, Medical College of Wisconsin, Milwaukee, WI 53226; 2) Department of Oncology, Tongji Hospital of Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, 430030; 3) Department of Biostatistics, University of Wisconsin, Milwaukee, WI 53201; 4) Department of Surgery, Medical College of Wisconsin, Milwaukee, WI 53226.

Cell free tumor DNA (cfDNA) circulating in blood has a great potential as biomarker for cancer clinical management. The objective of this study is to evaluate if cfDNA in blood plasma is detectable in early stage lung cancer patients. We extracted cfDNAs and tumor tissue DNAs from 8 lung adenocarcinoma patients. We also extracted cfDNA from 8 normal controls. To evaluate copy number variations (CNV) and identify potential mutations, we performed low pass whole genome sequencing and targeted sequencing of 50 cancer genes. To accurately reflect the tumor-associated genomic abnormality burden in plasma, we developed a new scoring algorithm, Plasma Genomic Abnormality (PGA) score, by summarizing absolute log2 ratios in most variable genomic regions. We performed digital PCR and allele-specific PCR to validate mutations detected by targeted sequencing. The median yield of cfDNA in 400ul plasma was 4. 9ng (range 2. 25-26. 98ng) in patients and 2. 32ng (range 1. 30-2. 81ng) in controls (p=0. 003). The whole genome sequencing generated approximately 20 million mappable sequence reads per subject and 5303 read counts per 1Mb genomic region. Log2 ratio-based CNV analysis showed significant chromosomal abnormality in cancer tissue DNAs and subtle but detectable differences in cfDNAs between patients and controls. Genomic abnormality analysis showed that median PGA score was 9. 28 (7. 38-11. 08) in the 8 controls and 19. 50 (5. 89-64. 47) in the 8 patients (p=0. 01). Targeted deep sequencing in tumor tissues derived from the 8 patients identified 14 mutations involved in 12 different genes. The PCR-based assay confirmed 3 of 6 selected mutations in cfDNAs. These results demonstrated that the PGA score and cfDNA mutational analysis could be useful tool for the early detection of lung cancer. These blood-based genomic and genetic assays are noninvasive and may sensitively distinguish early stage disease when combined with other existing screening strategies.

2670T

Prevalence of *BRCA1* and *BRCA2* common mutations among Algerian patients with breast and/or ovarian cancer. *F. Cherbal[1], H. Gaceb[1], C. Mehemmai[1], R. Bakour[1], K. Boualga[2], H. Mahfouf[3].* 1) Unit of Genetics, LMCB, Faculty of Biological Sciences, USTHB, Algiers, Algeria; 2) Radiation Therapy Services, Anticancer Center of Blida, Blida, Algeria; 3) Public Hospital Academic Medical Oncology Services, School of Medicine, University of Algiers, Rouiba, Algeria.

Background: Breast cancer is currently the leading cause of cancer morbidity and mortality among Algerian women. The present work aimed to screen for the most common mutations in *BRCA1* and *BRCA2* genes in breast and/or ovarian cancer patients. To date, the common mutations in *BRCA1* and *BRCA2* genes screened in this study, have been previously detected in studies in Algerian patients with family history of breast and/or ovarian cancer. **Methods**: Patients were enrolled from 2011 to 2015 in several public hospitals covering most geographical areas of Algeria. Patients were selected on early age onset (< 50 years) and family history of breast and/or ovarian cancer. In addition, in this study, all women (age <50) with triple negative breast cancer (TNBC) were considered eligible. 192 breast and/ or ovarian cancer patients were screened for the common *BRCA1* mutations c. 83_84delTG and c. 798_799delTT. The *BRCA1* mutation c. 181T>G and the common *BRCA2* mutations c. 1310_1313delAAGA and c. 7235_7236insG have been screened in 96 breast and/or ovarian cancer. The approach used is based in PCR-direct sequencing. **Results**: The *BRCA1* mutation c. 83_84delTG detected previously in two unrelated Algerian breast cancer families has been identified here in three triple negative breast cancer patients, and in a patient with a bilateral ovarian cancer, respectively. This mutation has been detected with a frequency of 2% (4/192). The four patients had a strong hereditary breast and/ or ovarian cancer history. The rare unclassified variant c. 122A>G/p. His41Arg located in the Ring Finger of the *BRCA1* exon 3 has been detected here for the first time in young Algerian TNBC patient. Moreover, the pathogenic *BRCA1* mutation c. 181T>G/p. Cys61Gly located in exon 5, has been detected in a young bilateral breast cancer patient with a strong breast cancer family history. Interestingly, the *BRCA1* mutation c. 798_799delTT detected previously in 4 unrelated families from Algeria, has been identified here in a young triple negative breast cancer patient with a family history of prostate and breast cancer. The common *BRCA2* mutations c. 1310_1313delAAGA and c. 7235_7236insG have not been detected in this study. **Conclusions**: Based on our currents results, we recommend using the TNBC immunophenotype as criterion to screen for *BRCA1* germline mutations in patients with early onset breast cancer. Screening for common mutations in *BRCA* genes in Algerian population, may facilitate genetic counseling and testing.

2671W

Next-generation sequencing of the *BRCA1* and *BRCA2* genes for the genetic diagnostics of hereditary breast and/or ovarian cancer. *D. Trujillano.* Centogene AG, Berlin, Germany.

Genetic testing for hereditary breast and/or ovarian cancer (HBOC) mostly relies on laborious molecular tools that use Sanger sequencing to scan for mutations in the *BRCA1* and *BRCA2* genes. We have explored a more efficient strategy based on next-generation sequencing (NGS) of the *BRCA1* and *BRCA2* genes in 210 (HBOC) patients. We first validated this approach in a cohort of 115 samples with previously known *BRCA1* and *BRCA2* mutations and polymorphisms. Genomic DNA was amplified using the Ion AmpliSeq™ *BRCA1* and *BRCA2* panel. The DNA Libraries were pooled, barcoded and sequenced using an Ion Torrent PGM sequencer. The combination of robust bioinformatics tools allowed us to detect all previously known pathogenic mutations and polymorphisms in the 115 samples, without detecting spurious pathogenic calls. The assay achieved a sensitivity of 100% (95% CI: 99. 71% to 100%), with a specificity of detecting non-variant sites from the reference sequence of 99. 99% (95% CI: 99. 99% to 100%), a positive predictive value of 91. 17% (95% CI: 89. 72% to 92. 62%), and a negative predictive value of 100% (95% CI: 100% to 100%). We then used the same assay in a discovery cohort of 95 uncharacterized HBOC patients for *BRCA1* and *BRCA2*. In addition, we describe the allelic frequencies across 210 HBOC patients of 74 unique definitely and likely pathogenic, and uncertain *BRCA1* and *BRCA2* variants, some of them not previously annotated in the public databases. Targeted NGS is ready to substitute classical molecular methods to perform genetic testing on the *BRCA1* and *BRCA2* genes, and provides a greater opportunity for more comprehensive testing for at-risk patients.

2672T

Identification of a circulating miRNA signature for noninvasive detection of early colonic polyps through optimized miRNA-seq of blood plasma. *A. Hardigan[1,2], B. Roberts[1], M. Kirby[1], C. M. Wilcox[2], R. Myers[1].* 1) HudsonAlpha Institute for Biotechnology , Huntsville, AL; 2) University of Alabama at Birmingham, Birmingham, AL.

Colorectal cancer is the third most prevalent cause of cancer morbidity and mortality in the United States. Currently, screening for colorectal cancer involves receiving an annual colonoscopy for those over age 50 and even younger in at-risk populations. Although effective, colonoscopy screening is invasive and its associated costs represent a significant economic burden to the health care system. Development of a less invasive and highly sensitive molecular assay capable of detecting precancerous adenoma formation would be highly informative and provide not only a mechanism for early detection and treatment of colon cancer, but also reduce the personal and economic burden of unnecessary colonoscopies. The discovery of stable miRNA species in blood and other easily accessible human biological fluids has led to their investigation as potential biomarkers for a variety of diseases such as cancer, neurodegenerative disease, and cardiovascular disease. Next-generation sequencing (NGS) of small RNAs in these fluids is a powerful method for comprehensive identification and quantification of miRNAs species. However, due to methodological challenges related to the small size of miRNA species and their relatively low abundance in human plasma, the identification of reproducible and specific miRNA biomarkers has been difficult. Here, we have optimized miRNA-sequencing library generation from human plasma as well as developed statistical quality control metrics to ensure accuracy and reproducibility of the miRNA-seq libraries. In order to determine a circulating miRNA signature capable of distinguishing between patients with early, precancerous adenomas and those without polyps, we applied these optimized methods to generate miRNA libraries from plasma taken from over 200 patients who underwent colonoscopy screening. Through these analyses we identified a candidate panel of miRNAs that is capable of discriminating between adenoma-burdened and normal patients. The methodological improvements we have made and our candidate circulating miRNA signature provide a foundation for future advances in both colon cancer detection and treatment as well as the detection of other diseases by means of circulating miRNAs.

2673W

PIK3CA mutational analysis platform utilizing the Labcyte Echo® Liquid Handler to reduce cost, sample amount and increase throughput to broadly assess acquired mutational status. *J. Lesnick, L. Orren.* Labcyte Inc, Sunnyvale, CA.

The phosphoinositide-3 kinase (PI3K)/AKT signaling network is one of the two most commonly mutated pathways identified in human cancers and the most frequently mutated pathway in breast cancer. Somatic alterations including PIK3CA mutations are the most common genetic alteration of this pathway; 80% or more occur within the helical (E542K, E545K) and kinase (H1047R) domains of p110. Such mutations confer increased catalytic activity for the generation of the second messenger phosphatidylinositol (3,4,5)- triphosphate (PIP3) and downstream pathway activation to induce cell proliferation, survival and potential tumorigenesis. These mutations can also confer resistance to therapeutic intervention. Identifying acquired mutations, not present in normal patient tissue, in tumor derived material, including CTC's and DTC's, is important to diagnostic predisposition testing for cancer, prognosis and can ultimately inform targeted therapy. Utilizing the Labcyte Echo liquid handler, together with real-time PCR mutation analysis, we have demonstrated a highly efficient, low-cost assay platform to detect PIK3CA somatic mutations. The mutations identified include the helical E542K, E545K and kinase domain H1047R pathway activating mutations. The Echo 555 liquid handler was used to dispense sub-microliter volumes of reagents from the EntroGen PIK3CA real-time PCR mutation analysis kit for subsequent analysis on the Roche LightCycler® 480 real-time PCR instrument. The methods detailed here reduce the total volume of reaction from 30µL to 3µL resulting in a lower cost assay platform. The high analytical sensitivity, limit of detection, of the assay platform enables detection of mutant allelic frequency below 10% over wild-type background. These results were obtained following assay miniaturization that used significantly less input of precious, limited sample DNA. The use of the Echo system provides a highly efficient workflow that minimizes the risk of operator handling errors and faster reaction process time. The use of a high density 384-well reaction plate increases throughput further by 4-fold over the standard 96-well format. The Labcyte Echo liquid handler enables highly efficient, low cost, precise and accurate mutational analysis for PIK3CA mutation profiling. This fast, low-cost and robust method can identify mutations of importance toward predisposition testing of cancer, prognosis, targeted therapy as well as identification of inherited disease.

2674T

The upregulation mechanisms of leptin-induced ADAMTS-1, involved in angiogenesis. *K. O. Yaykasli[1], O. F. HATIPOGLU[2], M. DOGAN[3], Y. SAHIN[4], E. YAYKASLI[5], E. GUNDUZ[2].* 1) Department of Medical Biology, Kahramanmaras University Medical Faculty, Kahramanmaras, Turkey; 2) Department of Medical Genetics, Turgut Ozal University Medical Faculty, Ankara, Turkey; 3) Department of Medical Genetics, Duzce University Medical Faculty, Duzce, Turkey; 4) Department of Medical Genetics, Necip Fazil City Hospital, Kahramanmaras, Turkey; 5) Department of Medical Biology and Genetics, Duzce University Institute of Health Science, Duzce, Turkey.

Objective: ADAMTS-1 belonging a disintegrin and metalloproteinase with trombospondin motifs (ADAMTs) protease family. Besides the aggrecanolytic activity, ADAMTS-1 is involved in angiogenesis. Angiogenesis is a normal physiological process and described as a formation of new blood vessels. However, aberrant regulation of angiogenesis caused several pathological processes including cancer. Even, the etiology of cancer has been not understood yet, the obesity was assumed one of the main factors for several cancers like chondrosarcoma. Recently, a novel adipokine called leptin secreted from adipose tissue was identified. Then, leptin become target to develop new therapy against to obesity associated diseases like cancer. The aim of this study was to investigate the signaling pathways involved in ADAMTS-1 upregulation induced by leptin in human chondrocytes. **Material and Method:** Confluent human articular chondrocytes were cultured with serum-free medium for 24 hours. Then the cells were pretreated with SB203580, SP600125, PD98059, QNZ and LY294002 reagents for inhibition of p38, JNK, MEK1, NF-κB and PI3-kinase pathways respectively for 30 min. Subsequently, the cells were treated with leptin at 1000ng/mL for 48 h. At the end of the incubation, total RNA was extracted using TriPure reagent, and 2 µg of total RNA was reverse-transcribed. The ADAMTS-1, and β-actin genes expression were analyzed by real-time polymerase chain reaction. **Results and Conclusion:** The current study was investigated that leptin increased the ADAMTS-1 gene expression level by mitogen-activated protein kinases (MAPKs) signaling pathways. Vascular endothelial growth factor (VEGF) has angiogenic property, and leptin caused VEGF upregulation has been identified recently. Similarly, the upregulation of ADAMTS-1 gene expression level by leptin and VEGF was demonstrated. Even, the roles of ADAMTS-1 in angiogenesis are under debate, it was clarified that ADAMTS-1 involved in angiogenesis. So, ADAMTS-1 has pivotal importance for obesity related cancer like chondrosarcoma. In this study, the signaling pathways of ADAMTS-1 upregulation caused by leptin were investigated. The elucidation of these pathways may give new idea to develeop new approach for chondrosarcoma therapy. However, further studies are needed.

Acknowledgements: This study was supported by TUBITAK (The Scientific and Technical Research Council of Turkey), (Project Number: SBAG/111S218).

2675W

The contribution of high and moderate penetrance breast cancer susceptibility genes to familial breast cancer risk in *BRCAx* families. *K. N. Maxwell[1], S. N. Hart[2], J. Vijai[3], K. Schrader[4], T. P. Slavin[5], T. Thomas[3], R. Moore[2], C. Hu[2], B. Wubbenhorst[6], S. M. Domchek[1], M. E. Robson[7], P. Radice[8], P. Peterlongo[8], J. Ford[9], J. Garber[10], C. Szabo[11], S. Neuhausen[12], J. Weitzel[5], K. Offit[7], K. L. Nathanson[6], F. J. Couch[2].* 1) Department of Medicine, Division of Hematology-Oncology, University of Pennsylvania, Philadelphia, PA; 2) Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN; 3) Clinical Genetics Research Lab, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 4) Hereditary Cancer Program, British Columbia Cancer Agency, Vancouver, BC; 5) Department of Medical Oncology, Division of Clinical Cancer Genetics, City of Hope, Duarte, CA; 6) Department of Medicine Division of Translational Medicine and Genetics, University of Pennsylvania, PA; 7) Clinical Genetics Service, Department of Medicine, Memorial Sloan-Kettering Cancer Center, New York, NY; 8) IFOM, the FIRC Institute of Molecular Oncology, and Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; 9) Division of Oncology, Stanford University School of Medicine, Stanford, CA; 10) Center for Cancer Genetics and Prevention, Dana Farber Cancer Institute, Boston, MA; 11) National Institutes of Health, Bethesda, MD; 12) Beckman Research Institute of City of Hope, Duarte, CA.

BRCA1 and *BRCA2* (BRCA) testing is uninformative in approximately 80% of families with clinical features of inherited susceptibility to breast cancer. Early identification of the individuals at risk in *BRCAx* families may lead to enhanced screening and prevention strategies, and potentially improved overall survival, as has been seen for carriers of *BRCA1/2* mutations. A proportion of these families are likely explained by mutations in other high penetrance and moderate penetrance breast cancer susceptibility genes. We performed targeted panel sequencing in 2331 *BRCA1/2* negative individuals with high-risk familial breast cancer, defined as a proband and at least two first or second degree relatives with breast cancer under age 70 or an individual with both breast and ovarian cancer. Data were analyzed for identification of all variants in 11 known high and moderate penetrance breast cancer susceptibility genes (*CDH1, PTEN, STK11*, *ATM, BARD1, BRIP1, CHEK2, MRE11A, NBN, PALB2, RAD50*). Overall, pathogenic mutations were identified in 179 of the 2331 *BRCA1/2* negative individuals (7. 8%). The most commonly mutated breast cancer susceptibility gene was *CHEK2*, which accounted for 54% of the mutation carriers and was found in 4. 2% of individuals overall. Other more commonly identified moderate penetrance genes included *ATM* (13% of mutation carriers, 1. 0% of individuals overall) and *PALB2* (9% of mutation carriers, 0. 7% of individuals overall). Outside of *TP53* (reported in a separate abstract), high penetrance gene mutation carriers were exceedingly rare, with identification of one *CDH1* mutation carrier (0. 04% of individuals), two *PTEN* mutation carriers (0. 09%) and no *STK11* mutation carriers. Our data show that significant genetic heterogeneity exists in *BRCAx* families; however, the majority of the individuals tested do not have mutations by multiplex panel testing of the assayed genes. Large-scale collaborative efforts are required to attain sufficient power to understand how to appropriately apply these results clinically in cancer risk evaluation and to identify the genetic risk factors in negative families.

2676T

Using expression data to define patient specific predictors for survival outcomes in lung adenocarcinoma. *G. Cai[1], F. Xiao[2], C. Cheng[1], Y. Li[3], C. I. Amos[3], M. L. Whitfield[1].* 1) Department of Genetics, Geisel School of Medicine at Dartmouth, Hanover, NH; 2) Department of Biostatistics, Yale University School of Public Health, New Haven, CT; 3) Department of Biomedical Data Science, Dartmouth College, Hanover, NH.

Background: Disease prevalence and drug responses vary widely among populations because of heterogeneous genetic makeups. Tumor and paired tumor-adjacent normal tissues are often collected and measured to remove noise from this heterogeneity. Methods that take into account patient-specific factors such as tumor-normal pair and ethnicity are required in studying the transcriptome expression in cancer. Here, our objective was to develop a novel mixed model that would combine these variables in analyses of paired tumor-normal studies. **Methods:** A mixed effect linear model was designed to detect differential expression (DE) genes between tumor and normal tissues with random effects of patient and ethnicity. Further, we selected diagnosis and prognosis biomarkers using the L1 penalized regression technique. **Results:** The method we developed was applied in lung adenoma datasets. In both Caucasians and Asians, we found cell cycle checkpoint pathways were down-regulated in tumors, whereas metabolic-related pathways like *PPARG* pathway and acetate metabolism pathways differed between two ethnic groups. We also demonstrated the potential for disease and prognosis prediction by gene expression markers. We identified a 10-gene biomarker panel for both diagnosis and prognosis of lung cancer, which are applicable for both Caucasian and Asian lung adenoma patients. Compared to low risk groups, high risk groups showed significantly shorter overall survival time (HR=2. 70, *p*-value=1. 37E-06 when training algorithm using data from Caucasian patients; HR=3. 25, *p*-value=0. 002 when testing the prediction performance using data from Asian patients). **Conclusions:** We developed a novel statistical method that can detect DE genes between tumor and normal tissues with random effects of patient and ethnicity. These results provide a valid strategy and basis for minimally identifying biomarkers from high-throughput transcriptome profiling data for other tumor types or different diseases.

2677W

Deubiquitinating enzyme Usp15 Regulates cell cycle progression by deubiquitination of spliceosomal protein. *T. DAS[1,2], JK. Park[1], JY. Park[1], E. Kim[1], YS. Yoo[1], Eunice. Kim[1], EJ. Song[1,2].* 1) Korea Institute of Science and Technology, seoul, Seoul, South Korea, South Korea; 2) University of Science & Technology (UST), Daejon, Korea.

Splicing of pre-mRNA is catalyzed by dynamic splicing machineries, spliceosome. Reversible post-translational modifications such as ubiquitination contribute the spliceosome dynamics for efficient pre-mRNA splicing, but little is known about the underlying mechanism. Here, we showed that Prp31, a component of the U4 snRNP, is modified with K63-linked ubiquitin chains by Prp19 complex and is deubiquitinated by Usp15 and Sart3 which translocates Usp15 from cytoplasm to nucleus thereby mediate to target for its substrate. Ubiquitination and deubiquitination status of Prp31 regulates its affinity to U5 snRNP component, Prp8. It regulates catalytic activity of spliceosome for efficient splicing by stabilization of U4/U6. U5 tri-snRNP. In addition, depletion of Usp15 interferes with the proper mRNA splicing of cell cycle related genes, for example Bub1. Usp15 regulate spliceosomal activity by inhibiting the modification of spliceosomal proteins with ubiquitin chains. Loss of Usp15 interferes with the accumulation of correctly spliced mRNAs. Therefore, our findings indicate that Usp15 is involved in cell cycle progression by regulating the modification of spliceosomal proteins.

2678T

Down-regulation of *hTERT* in glioblastoma cell lines using siRNA. *G. K Chetan, Ch. Lavanya, M. K Sibin, M. M Srinivas Bharath, Bhat. I Dhananjaya, K. V. L Narasinga Rao.* National Institute of Mental Health and Neurosciences, Bangalore, Karnataka, India.

Purpose: Glioblastoma accounts for more than half of all malignant glioma cases that are associated with high morbidity and mortality. Presence of increased telomerase activity is a common feature of all cancers. *hTERT* gene expression increases with the grade of glioma and it is considered as a good candidate for the targeted therapy. Both RNA and catalytic subunits are targeted for blocking the activity of *hTERT*. Artificial transcription factors, drugs, antisense oligonucleotides coupled with radiation have been used to regulate *hTERT* with varied success rate. The present study shows the effective down-regulation of telomerase activity using RNA interference approach in glioblastoma cell lines. **Methods:** We used Grade IV LN-18 glioblastoma cell line for this study. siRNA1 was used to target exon4 and siRNA2 was used to target exon3. Level of expression of *hTERT* mRNA was detected by qRT-PCR. MTT assay was performed to determine cell viability after transfection at different time intervals. Expression of *hTERT* protein was determined by western blotting. One-way ANOVA was performed to compare the mean values of control, *hTERT* siRNA treated and negative siRNA treated groups. p<0. 05 was considered statistically significant for all the tests. **Results:** The *hTERT* was transiently transfected in to LN-18 cell line. We compared the results of *hTERT* specific siRNA treated groups with control and scrambled siRNA. We observed a significant decrease in the percentage of cell viability in LN-18 cell line transfected with siRNA1 (P<0. 01) and siRNA2 (P<0. 05) as detected by MTT assay. We also observed that, siRNA1 is more effective than siRNA2 with more than 50% inhibition of cell proliferation in all time points. The level of *hTERT* mRNA expression with siRNA1and siRNA2 treated group was ~ 60%and ~ 40%respectively with a P<0. 05. The level of *hTERT* protein was significantly lower in siRNA1 compared to siRNA2 groups. The siRNA1 against *hTERT* showed increased suppression of *hTERT* at mRNA and protein levels than siRNA 2. We conclude that, choosing the right siRNA for effective *hTERT* down regulation is important and siRNA approach is a promising technique, which can be used in animal models and Humans as a therapeutic target.

2679W

The role of periostin in regulating of early tumorgenesis in oral squamous cell lines carcinomas. *G. E-H. Gawish[1, 3], H. A Al Omer[2, 3].* 1) Medical Genetics, College of Medicine, Al Imam Muhammad Ibn Saud University, riyadh, Saudi Arabia; 2) Prince Abdulrahman Advanced Dental Institute (PAADI), Ministry of Defense and Aviation, Riyadh, Saudi Arabia; 3) Oral Biological Medical Science, University of British Columbia, Canada.

Oral squamous cell carcinomas (OSCCs) are the most frequent malignancy of the oral cavity. Tissue microarray analysis of OSCC found the upregulation of *POSTN* gene (periostin, osteoblast specific factor) expression in OSCC compared with normal tissues. It was reported that periostin contributes to malignancies mainly by preventing apoptosis and promoting angiogenesis, invasion, and metastasis. The roles of periostin in regulating cell proliferation and in genomic instability of cancer cells during tumorigenesis still require further investigation. The aims of this study was to indicate the oncogenic activity effect of periostin in OSCC, contribution of periostin in promoted OSCC proliferation and calculate the DNA index of ploidy and anploidy channels. SCC4, SCC9, SCC15, and SCC25 were transfected with periostin. For unbiased, siRNA and transfected Ca9-22 cells were used. Expression of periostin was analyzed then cells were stained by annexin V for apoptosis detection and PI for proliferation and DNA ploidy then stained cells were analyzed using FACScaliber. This study found that the oral squamous cell carcinoma cell lines transfected with periostin, exhibited significantly increased proliferation, compared with non-transfected. Periostin-overexpressing cells stained with annexin V and PI showed significantly ($p<0.05$) and ($p<0.005$) respectively. We have recorded that the number of hyperdiploid passages was 33% and 67% were diploid. Hyperdiploid passages with DNA index ≥ 1.16 and ≤ 1.6 were12 passages and hyperdiploid with DNA index >1.6 (11passages). We conclude that periostin play an important role in the tumorigenesis of OSCC by deregulation of the cell cycle, escaping from apoptosis, and the potential for unlimited replication. It encourages genetic alterations by enhancing chromosomal instabilities.

2680T

Molecular cytogenetic characterization of the malignant primitive neuroectodermal tumor cell line SK-PN-DW. *N. Du[1,2], W. Bao[1], K. Zhang[1], X. Lu[2], H. Pang[2], X. Wang[2], Y. Gu[2,3], G. Liu[2,4], J. Lu[2,5], Y. Kim[2], Y. Ren[2,6], F. Wang[1], S. Li[2].* 1) Department of Infectious Diseases, the First Hospital of Jilin University, Changchun, Jilin, P. R. China; 2) Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma city, OK; 3) Department of Pulmonology, the First Hospital of Jilin University, Changchun, Jilin, P. R. China; 4) Department of Gastroenterology, the First Hospital of Jilin University, Changchun, Jilin, P. R. China; 5) Department of Hematology, the First Hospital of Jilin University, Changchun, Jilin, P. R. China; 6) Department of Hematology, The First Affiliated Hospital of China Medical University, Shenyang, Liaoning, P. R. China.

Background: The SK-PN-DW cell line was established in 1979 and it is commercially available from the ATCC. This cell line has been used as an in vitro model in functional and therapeutic studies for malignant primitive neuroectodermal tumor (PNET). However, genetic alterations at the cytogenetic level have not been well established. **Results:** Utilizing a combination of multiple techniques (routine G-banded chromosome analysis, fluorescence in situ hybridization (FISH), and oligonucleotide array comparative genomic hybridization (array CGH) assays, we have successfully characterized all chromosomal changes of this cell line. The G-banded karyotype showed that the cell line had a modal number of chromosomes ranging from 40 to 41, all the cells had a loss of the Y chromosome and chromosomes of 11, 13, and 18. Some cells had a loss of chromosome 10. The cells analyzed also had chromosomal structural changes including an unbalanced translocation between chromosomes 1 and 7, the translocation between chromosomes 11 and 22 at breakpoints of 11q24 and 22q12 which is a classical translocation associated with Ewing sarcoma, a derivative chromosome due to the whole arm translocation between chromosomes 16 and 17 at likely breakpoints of 16p10 and 17q10, and possible rearrangement of the short arm of chromosome 18. A variable number of double minutes in each metaphase cell was also observed. The array CGH assay showed genomic-wide chromosomal imbalances in this cell line and precisely placed chromosomal breakpoints on unbalanced, rearranged chromosomes. This assay also disclosed subtle chromosomal changes as well as the chromosomal origin of the double minutes. The FISH assay, with multiple combinations of FISH probes, confirmed routine cytogenetic analysis and array CGH findings. **Conclusions:** Accurate determination of those genetic changes will provide guidance in making genotypic and phenotypic correlations and for establishing authenticity of the cell line when it is applied in biological studies.

2681W

Detection of Subtelomeric CNVs in Colorectal Cancer using MLPA. *M. Rocha[1,2], F. A. R Madia[2], E. A. Zanardo[2], T. V. M. Costa[2], A. T. Dias[2], F. Balsamo[1], S. D. F Borato[1], S. H. C. Horta[1], B. A. Bianco[1], D. M. Christofolini[1], C. P. Barbosa[1], J. Waisberg[1], L. D. Kulikowski[1,2].* 1) Laboratory of Genetic, Faculdade de Medicina do ABC (FMABC), Santo André-SP, Brazil; 2) Laboratory of Cytogenomic (Lim 3), Faculdade de Medicina da USP (FMUSP), São Paulo-SP, Brazil.

Colorectal cancer (CRC) is a health problem that is increasing in importance worldwide but its etiology is still poorly understood. Copy number variations (CNVs) are recognised as an alternative source of genetic instability that may influence cancer risk. In this sense we investigate CNVs in subtelomeric regions using MLPA technique (Multiplex ligation-dependent probe amplification) to better understand the genomic features associate to CRC. As a pilot study, 20 tissue samples were obtained by biopsy of 10 patients with CRC, being 10 neoplastic samples and 10 samples obtained from adjacent regions (10cm distant tumor tissue). The DNA was extracted using the QIAamp DNA Blood Midi Kit (QIAGEN, Valencia, California). Subsequently we performed MLPA (MRC-Holland®, Amsterdam, The Netherlands) with specific kits for the subtelomeric regions (P036 and P070). The results were analyzed using the software GeneMarker® (SoftGenetics, LLC, State College, PA - www. softgenetics. com). The results showed several different *indels*: deletions in *KDM5A*(12p), *TNFRS18* (1p), *CTDP1* (18q), *TRIML2(*4q*) IL-17RA*(22q), *MTA1*(14q), *GAS8* (16q) genomic regions; and duplications in *CHL1*(3p), *PSPC1*(13p*), SOX12*(20p), *SYBL1*(Xq), *CDC16*(13q), *PDCD6*(5p), *RABL2B* (22q) genomic regions. Structural variants could impact gene dosage and influence directly the number of functional copies of the gene that are available for transcription. We observed subtelomeric *indels* associated with cell migration and adhesion, transcriptional activation, cellular apoptosis and autoimmune diseases. Also the results showed that the cells are characterized by an unstable genome with heterogeneous and complex genomic abnormalities suggesting that “chromotripsis” could be a possible mechanism involved in these tumors. Moreover, modifications of subtelomeric genomic regions could play an important role in CRC predisposition and development.

2682T

Unusual revertant mosaicism in three siblings affected with Fanconi anemia group FA-G. *S. Chandrasekharappa[1], D. C. Kimble[1], F. Lach[2], J. Kennedy[2], A. Smogorzewska[2], A. D. Auerbach[2].* 1) Cancer Genetics and Comparative Genomics Branch, NHGRI/NIH, Bethesda, MD; 2) Laboratory of Genome Maintenance, The Rockefeller University, New York, NY.

Fanconi anemia (FA) is a genetically and phenotypically heterogeneous disorder characterized by congenital malformations, progressive bone marrow failure and predisposition to cancer. We describe here an unusual family followed by the International Fanconi Anemia Registry (IFAR) since 1989, with 4 of 11 siblings affected with FA. The index case (370_1) was diagnosed with FA around age 25, showing increasing pancytopenia after the birth of a child. She developed squamous cell carcinoma (SCC) of mouth and tongue in her mid-twenties, leading to death at age 39. She also had intra-ductal breast carcinoma at age 35 and SCC of the skin. A sister (370_2) had onset of pancytopenia at age 8, controlled with prednisone and later danizol. At age 37 she was diagnosed with invasive vulvar SCC with metastatic disease, and died soon thereafter. Sibling 370_3 had a matched sib BMT for MDS with complex cytogenetic abnormalities, and died at age 32 after a complicated post-transplant course. She had a history of multiple skin, cervical, perineal and perianal cancers and carcinomas-in-situ. An older brother (370_4) diagnosed at about age 10 with aplastic anemia was treated with prednisone, developed AML at age 18, and died in 1969 at age 19. DNA was not available from this sibling for genomic analysis. Sequencing of *FANCG* in lymphoblastoid cell lines from the 3 affected sisters indicated 3 different mechanisms accounting for reverse mosaicism (RM). The maternal variant c. 1158dupC in exon 10 was absent in the proband (370_1), likely from a gene conversion event correcting the mutant allele. The paternal alleles in each of 370_2 and 370_3 with inherited variant c. 1771dupC in exon 14 (p. L591PfsX14)(see http://www. rockefeller. edu/fanconi/) were found to be associated with *de novo* genomic variants c. 1729delC in exon 13 and c. 1761-1G>A in intron 13 respectively. cDNA analysis revealed, as predicted for 370_2, that the c. 1729delC results in correcting the defect caused by the inherited variant c. 1771dupC by restoring the reading frame except for altering amino acids encoding 577-589 from QTKGSHEDALWSL to KLRGHMKMLCGLS. Similarly, the *de novo* intronic variant c. 1761-1G>A in 370_3 causes c. 1761delG in the cDNA, resulting in only changing amino acids 587-589, from WSL to CLS. Three distinct alterations appear to be accounting for RM in the three affected siblings. The extent of RM, and the nature of the altered protein products might influence the phenotypic heterogeneity in FA.

2683W

The G2 micronucleus assay in lymphocytes shows increased radiosensitivity in in healthy BRCA1 mutation carriers. *K. BM. Claes[1], A. Baert[1-2], J. Depuydt[2], T. Van Maerken[1], B. Poppe[1], M. Van Heetvelde[1], K. De Leeneer[1], A. Vral[2].* 1) Center for Medical Genetics, Ghent University Hospital, Gent, Oost-Vlaandere, Belgium; 2) Department of Basic Medical Sciences, Ghent University Gent, Oost-Vlaanderen, Belgium.

The risk of breast cancer increases drastically in individuals carrying a germline *BRCA1* mutation. It is therefore crucial to subject individuals carrying a mutation to an intensified screening to detect possible tumours in an early stage. The general approach is to use mammography screening, complemented with MRI and/or echography. The screening typically starts at young age, resulting in a relatively high cumulative dose of ionizing radiation received during subsequent mammography screening. The exposure to ionizing radiation of *BRCA1* mutation carriers is contra intuitive since BRCA1 is active in the DNA damage response pathway. Moreover, BRCA1 is important in the initiation of homologous recombination and the G2/M checkpoint. In this study, radiosensitivity was determined in healthy *BRCA1* mutation carriers by means of the G2 micronucleus assay. We applied the assay on blood samples of carriers and non-carriers. We determined the radiosensitive phenotype using a robust grading system (grade 0-3) based on 4 parameters that investigate DNA repair as well as G2 arrest capacity. The robustness of this approach was demonstrated by a high reproducibility of the results obtained on samples taken at different time points from both carriers and non-carriers. Our analyses clearly show that healthy germline *BRCA1* mutation carriers are more radiosensitive compared to healthy non-carriers and we found that 39% of *BRCA1* mutation carriers had a radiosensitive grade of 2 or 3. These high grades were never observed in the non-carriers. Prospective studies are required to determine if these results may impact approaches for screening, e. g. avoiding mammography screening at young age in individuals demonstrating a high grade of radiosensitivity. In the current patient cohort, only patients with truncating mutations have been analysed. Further studies are ongoing to evaluate if a radiosensitive phenotype is also seen in carriers of pathogenic missense mutations. .

2684T

Identification of new target genes in microsatellite unstable colorectal cancer by exome sequencing. *J. Kondelin[1,2], E. Pitkänen[1,2], A. E. Gylfe[1,2], K. Palin[1,2], H. Ristolainen[1,2], R. Katainen[1,2], E. Kaasinen[1,2], M. Taipale[2,3], J. Taipale[2,3], L. Renkonen-Sinisalo[4], H. Järvinen[4], J. Böhm[5], JP. Mecklin[6], P. Vahteristo[1,2], S. Tuupanen[1,2], L. A. Aaltonen[1,2].* 1) Department of Medical and Clinical Genetics, University of Helsinki, Helsinki, Helsinki, Finland; 2) Research Programs Unit, Genome-Scale Biology Research Program, University of Helsinki, Helsinki, Finland; 3) Science for Life Center, Department of Biosciences and Nutrition, Huddinge, Sweden; 4) Department of Surgery, Helsinki University Central Hospital, Hospital District of Helsinki and Uusimaa, Helsinki, Finland; 5) Department of Pathology, Jyväskylä Central Hospital, University of Eastern Finland, Jyväskylä, Finland; 6) Department of Surgery, Jyväskylä Central Hospital, University of Eastern Finland, Jyväskylä, Finland.

Approximately 15% of colorectal cancers (CRCs) display microsatellite instability (MSI) caused by defective cellular mismatch repair. Most MSI CRCs are sporadic and result from biallelic inactivation of the *MLH1* gene, most often due to hypermethylation of its promoter. Cells displaying MSI accumulate a high number of mutations throughout the genome, especially in short repeat areas, microsatellites. These mutations typically lead to a premature stop codon resulting in a truncated protein product that may inactivate the gene, a mechanism known typical of tumor suppressor genes. To date, several genes have been proposed as MSI target genes based on high mutation frequency in targeted searches of microsatellites. New sequencing technologies, however, enable genome-wide investigation of mutations in human cancers. This has in turn led to identification of a vast amount of mutations in cancer, most of which are passenger mutations that do not confer selective growth advantage. The challenge is therefore to distinguish true driver genes from passengers. The high number of mutations in MSI tumors augments this challenge. The aim of this study is to identify new target genes in MSI CRC that function as true drivers in this tumor type. In this study, 24 sporadic MSI CRCs and their corresponding normal samples were exome sequenced to identify changes of somatic origin. An analysis and visualization program developed in our group, “RikuRator” (unpublished), was utilized along with a script to estimate the accurate somatic mutation frequency of each coding mononucleotide repeat. A statistical model based on exome sequencing data was developed that takes into account the background mutation frequency of microsatellites depending on the repeat length and nucleotide context. The most mutated repeats from the top 73 genes are validated by MiSeq-sequencing in a set of additional MSI CRCs. Functional studies will be carried out to further investigate the pathogenic effect of the new target genes identified. The systematic screening of mutations in microsatellites will improve our understanding on the mutation profile of these tumors. With our approach that takes into account the background mutation frequency of microsatellites, we aim to identify true driver genes in MSI CRC. A detailed understanding of the molecular background of this tumor type is important for the development of more efficient screening methods and personalized treatments for patients with MSI CRC.

2685W

Integrated landscape of molecular alterations in Uveal Melanoma. *H. Anbunathan[1], A. M. Bowcock[1], J. W. Harbour[3], R. Rabadan[2], R. D. Melamed[2].* 1) Imperial College London, London, LONDON, United Kingdom; 2) Columbia University, New York, USA; 3) Bascom Palmer Eye Institute, Miami, USA.

Uveal melanoma (UM) is the second most common form of melanoma and the most common primary cancer of the eye in adults. Tumors arise from neural-derived melanocytes of the uveal tract of the eye leading to tumors of the iris, ciliary body and choroid. Approximately 50% of tumors metastasize within 10 years, most commonly to the liver. Tumors can be subdivided into groups on the basis of gene expression differences that correspond to those low metastatic risk (class 1 tumors) and high metastatic risk (class 2 tumors). Class 2 tumors are frequently associated with monosomy 3. We previously used exome sequencing to identify driver mutations in *BAP1* gene in the metastasizing class 2 tumors with monosomy 3 and recurrent mutations in *SF3B1* in the class 1 tumors with less metastatic competency. *EIF1AX* mutations have also been identified in the low-grade class 1 tumors. We have now completed exome sequencing of 40 UM samples, 23 matched normal DNAs, 2 metastases, and 6 UM cell lines. In the case of one tumor we were able to exome sequence three independent segments and one metastasis which provided information about tumor evolution. We have also performed RNA-Sequencing on 22 UMs, and analysed 14 UM/normal pairs for CNVs by interrogating the Illumina OmniArray. UMs harbored an average of 79 somatic mutations, including 47 nonsynonymous mutations. Chromosomal copy number profile reveal 1q gain, monosomy 3, 6p gain, 6q loss, 8q gain and 9p loss which is consistent with previous findings. We also observed changes in chr11 and chr13 in 5 class 1 tumors. Gene fusions were identified with RNA-Seq data using more than one fusion detection tool (Defuse, Chimerascan and Tophat). These included *NAIP-OCLN*, *AZGP1-GJC3* (validated with RT-PCR), and *CHIA-TMEM101* fusions. The six cell lines harboured an average of 950 mutations per sample including 540 nonsynonymous mutations. Cell lines harboured known mutations in *SF3B1* and *EIF1AX* but none contained *BAP1* mutations. We found an average of 40 mutations in the two tumor metastases. However, there were no shared altered genetic drivers in the metastases other than those seen in the primary tumors. This suggests that for uveal melanoma, pathogenic mutations in *BAP1*, *EIF1AX* and *SF3B1* still are the best predictors for tumour progression. .

2686T

Somatic and genomic mutation spectrum of unilateral Vestibular Schwannoma (VS). *R. Birkenhager, S. Arndt, W. Maier, A. Aschendorff, R. Laszig.* Department of Otorhinolaryngology and Head and Neck Surgery, University Medical Center Freiburg.

A vestibular schwannoma (VS) (acoustic neuroma) is a benign tumors originated from schwann cells (SC) of the vestibular nerve and located in the cerebellopontine angle or the inner auditory canal. The tumor results from an over proliferation of SC; these cells wrap themselves around nerve fibres, often causing gradual hearing loss, tinnitus and dizziness. It can also affect with the facial nerve causing paralysis by compression. Early detection of the tumor is sometimes difficult, because the symptoms may be subtle and may not appear in the beginning. There are two types of VS: Bilateral and unilateral. Until now there is only one single gene known for bilateral VS, that leads to neurofibromatosis type 2 (NF2). Unilateral VS account for approximately 8 % of all cranial tumors. The exact cause of unilateral VS is unknown, most occur spontaneously. Our study group consists of 144 patients (from 139 independent families) of which 23 have a bilateral VS (NF2), 119 unilateral VS and 2 with an initial suspicion of VS. All patients were clinically clearly defined. Genetically we performed a mutation analysis on the NF2 gene at the genomic and tumor level. So far we identified different missense/nonsense and splice mutations in NF2 gene, on the genomic level and in some cases exclusively in tumor cells, which all lead to a loss of function of the gene product Merlin and the appearance of VS. In addition to somatic changes and mutations there are possible indications of a genomic predisposition to unilateral vestibular schwannoma.

2687W

Comprehensive genomic characterization of three spatially and temporally distinct tumors from different organs in a single patient exhibiting both BRCA2 and VHL germline mutations. *S. M. Boyle[1], M. J. Clark[1], T. H. Ho[2], M. L. Stanton[2], E. P. Castle[2], A. S. Parker[3], R. Alla[1], E. Helman[1], S. Luo[1], M. Snyder[4], J. West[1], R. Chen[1].* 1) Personalis, Menlo Park, CA; 2) Mayo Clinic, Inc. , Scottsdale, AZ; 3) Mayo Clinic, Inc. , Jacksonville, FL; 4) Department of Genetics, Stanford university, Stanford, Ca.

Inherited genetic mutations have been strongly associated with heightened risk for a broad range of cancer subtypes. Over the last two decades, large efforts have focused on recognizing cancer driver genes and better understanding their role in cancer progression, resulting in identification and characterization of important drivers, such as BRCA1, BRCA2, , PTEN, and VHL. In very rare cases, a single individual has germline mutations in more than one of these cancer driver genes. To profile multiple tumor progression when two susceptibility gene mutations are detected, we applied next generation sequencing to analyze and characterize the development of three distinct tumors in a 54-year-old female with germline mutations in both BRCA2 and VHL. BRCA2 is well known for its role in breast, ovarian,, and skin cancers. Damaging germline mutations in VHL are rare (1/36,000 individuals) and cause Von Hippel-Lindau, a rare autosomal dominant disorder that predisposes to a range of distinct cancers, including renal cell carcinoma (RCC), hemangioblastomas. To our knowledge, extensive genetic profiling has not been performed on a patient with both BRCA2 and VHL mutations. Over the course of 5 years, this patient was diagnosed with pheochromocytoma, RCC, and basal cell skin cancer. Each of the tumors was identified more than one year apart, was treated, and is currently under active treatment for RCC. Using the ACE Extended Cancer Panel, a targeted enrichment sequencing platform including over 1,300 cancer genes and 200 miRNAs, we sequenced each neoplastic site as well as adjacent normal tissues. Both the DNA and RNA were sequenced to high depth, and small variants, gene expression, copy number alterations, and gene fusions were assessed. Additionally, by sequencing both DNA and RNA, we evaluated allele specific expression and verified copy number changes affecting gene expression. This is the first study to investigate the progression of multiple tumors arising from dual BRCA2/VHL germline mutations in a single patient. In addition to identifying the impact of inherited BRCA2 and VHL mutations in each tumor, we extensively characterized driver mutations in all tumors, identifying private mutations in each. We discuss how each contributes to overall cancer progression and demonstrate the ways in which an integrated RNA/DNA approach with a cancer focused augmented enrichment panel enables a greater understanding of the genetics of cancer in a particular case.

2688T

Down regulation of TGFB3 plays a critical role in chordoma formation. *W. Chen[1], L. Wang[2], Q. Hu[1], Z. Wu[2], J. Zhang[2], C. Zeng[1].* 1) Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China; 2) Department Neurosurgery, Beijing Tian Tan Hospital, Capital Medical University, Beijing, China.

Chordoma is rare malignant tumor, arising from notochordal remnants and occurs along the spine. Study on chordoma is challenging owning to its low prevalence and lack of normal controls. Although the expression of T gene was seen as a diagnostic marker, its direct effect to tumor genesis remains unclear and thus far no other gene is consistently reported to be involved in chordoma development. To identify the important pathway in the formation of chordoma, we conducted miRNA and mRNA sequencing in two chordoma tissues and two fetal notochords, which were seen as normal control tissue of chordoma. We found the differentially expressed mRNA-miRNA pairs were enriched in TGFB pathway. And the chordoma up-regulated miR-29b-3p and its down-regulated target gene TGFB3 were further validated in additional 8 chordoma tissues and 8 notochords. There is a significant negative correlation between the expression of TGFB3 and miR-29 in chordoma samples (r2=0. 87). Lacking of such a correlation in notochords suggests the down-regulation of TGFB3 in chordoma were mainly caused by up-regulation of miR-29. We further compared the copy number of miR-29 gene loci between chordoma tissues and their paired blood samples. Among the 8 chordoma patients, 5 had somatic copy number gain at miR-29b1 and/or miR-29b2 loci. In addition, in seven chordoma samples with enough blood controls, somatic copy number loss of TGFB3 were observed in 4/7 samples. To verify the interruption of TGFB3 in chordoma formation, we knocked down TGFB3 on zebrafish eggs by morpholino approach. As the zebrafish lavas developed to 48 hours, about a half of them showed curved shape phenotype. Chordoma like phenotype in zebrafish notochords was then identified by H&E staining and immunohistochemistry. Our findings demonstrate that down-regulation of TGFB3 directly causes chordoma formation and TGFB3 or miR-29 may act as a plausible drug target for chordoma treatment.

2689W

Genomic alterations in *BCL2L1* and *DLC1* contribute to drug sensitivity in gastric cancer. *S. Y. Cho[1], H. Park[1,2], H. Kim[2], D. C. Na[1], J. Y. Han[1], J. Chae[3], C. Park[3], O. K. Park[3], K. Min[1], J. Kang[1], B. Choi[4], J. Min[4], Y. S. Suh[5], S. H. Kong[5], H. J. Lee[4,5], E. Liu[2], J. I. Kim[3], S. Kim[6], H. K. Yang[4,5], C. Lee[2].* 1) Department of Life Science, Ewha Womans University, Seoul, Korea; 2) The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA; 3) Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul, Korea; 4) Cancer Research Institute, Seoul National University College of Medicine, Seoul, Korea; 5) Department of Surgery, Seoul National University College of Medicine, Seoul, Korea; 6) Department of Molecular Medicine and Biopharmaceutical Sciences, Seoul National University, Seoul, Korea.

Gastric cancer (GC) is the third leading cause of cancer-related deaths worldwide. Recent high-throughput analyses of genomic alterations reveal several driver genes and altered pathways in GC. However, therapeutic applications from genomic data are limited, largely due to the lack of druggable molecular targets and preclinical models for drug selection. To investigate therapeutic targets for GC, we performed whole exome sequencing and array Comparative Genomic Hybridization (aCGH) of DNA from 103 GC patients. Pathway analysis showed recurrent alterations in the Wnt signaling (*APC, CTNNB1* and *DLC1*), the ErbB signaling (*ERBB2, PIK3CA* and *KRAS*) and the p53 signaling/apoptosis (*TP53* and *BCL2L1*) pathways. Notably, in 18% of GC cases (19/103), amplification of *BCL2L1* was observed and subsequently a BCL2L1 inhibitor was shown to markedly decrease cell viability in cell lines and in similarly altered patient-derived GC xenografts, especially when combined with other chemotherapeutic agents. In 11% of cases (6/55), mutations in *DLC1* were found and also shown to confer a growth advantage for these cells via activation of Rho-ROCK signaling, rendering these cells more susceptible to a ROCK inhibitor. Taken together, our study implicates *BCL2L1* and *DLC1* as potential druggable targets for a subset of GC cases.

2690T

Flying Blind: Building a Cancer Genomic Standard. *M. Clark, E. Helman, S. Boyle, R. Alla, S. Luo, D. Church, N. Leng, S. Kirk, P. Sripakdeevong, A. Huang, J. West, R. Chen.* Personalis, Inc. , Menlo Park, CA.

Cancer genomes are a hodgepodge of mutational types, including small variants as well as more complex structural changes, directed by strong selection pressures to enhance oncogenesis, or simple genomic instability. Tumor heterogeneity is common, such thatcancer genomes are not a single entity, but rather a population of genomes with varied mutations. Tumor samples range from solid tumors to liquid biopsies, and are often contaminated with adjacent normal tissue. Moreover, cancer samples are treated with formalin and embedded in paraffin for archival purposes, causing additional genomic variations. Much attention has been paid to recent efforts to define gold standard human genomes, but thus far they lack many features that are characteristic of cancer genomes. No single cancer sample contains all of these various features, and any given cancer sample is very limited in quantity. There are some single-sample cancer reference standards, including the Horizon Quantitative Multiplex, an engineered FFPE cell line mix with variants at allele frequencies (AFs) from 1. 0 to 41. 5%. Sequencing this cell line, we detected 18/18 SNVs and 4/4 indels. While samples of this type have limited utility, there is a clear need for samples containing the full spectrum of mutation types and frequences. Therefore, we made a collection of 28 cancer cell lines to generate a cancer genome standard set including 817 known SNVs, 62 small indels, 21 deletions, 23 amplifications, and 14 gene fusions. We simulated tumor heterogeneity by mixing the cell lines at various ratios, generating variant allele frequencies down to 1% and emulated reduced purity by mixing cell lines with paired normals at ratios down to 10%. These experiments required sequencing over 140 samples and mixes. Using an augmented oncome panel (targeting >1,500 cancer genes), these multiplexes enabled us to interrogate substantial quantities of variants. We detected 16136/16146 SNVs at 5% AF, 639/646 indels at 10% AF, 29/30 CNAs at 20% purity, and 14/14 gene fusions using this panel. These high variant counts enabled us to establish tighter confidence intervals and validate this oncome platform for detection of all major cancer variant types. .

2691W

Genetic profile of tumorigenesis in Brazilian NSCLC patients. *P. GP. Couto[1], F. M. MELO[1], R. GUIMARAES[2], S. D. PENA[2], A. VILHENA[1], E. FRIEDMAN[3], L. BASTOS-RODRIGUES[4], L. DE MARCO[1].* 1) Departments of Surgery, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; 2) Departments of Biochemistry, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil; 3) The Susanne Levy Gertner Oncogenetics Unit, Chaim Sheba Medical Center, Tel-Hashomer, Israel; 4) Department of Basic Sciences, Universidade Federal de Juiz de Fora, Brazil.

Introduction: Lung cancer is the most common type of neoplasm and the leading cause of cancer-related deaths worldwide. To improve the survival rate of lung cancer patients, a better understanding of tumor biology is required to help the development of new therapeutic strategies. Over the past decade, it has become evident that subsets of NSCLC can be further defined at the molecular level by recurrent 'driver' mutations that occur in multiple oncogenes including *AKT1, ALK, BRAF, EGFR, HER2, KRAS, MEK1, MET, NRAS and PIK3CA*. Next-generation sequencing technologies have revolutionized cancer genomics research by providing an unbiased and comprehensive method to detect genome abnormalities. **Materials and methods**: Our study analyzed seven samples each from smokers and non-smokers Brazilian patients with lung adenocarcinoma through whole exome sequencing. Samples were subjected to whole exome capturing and sequencing using the Roche NimbleGen V2 chip and the Illumina HiSeq2000 sequencing platform. The generated files were analyzed separately using *Ingenuity® Variant AnalysisTM* and *Mendel, MD* softwares. Variant calling of both samples resulted in 102,933 single nucleotide variations (SNVs) being 13,837 registered as novel variants. **Conclusion**: Our data shows that genes other than those that have already been published are present in Brazilian patients. Therefore, further studies are needed to investigate molecular alterations in tumor-related genes and help define a genetic profile of tumorigenesis in Brazilian NSCLC patients.

2692T

Breast cancer mutations are enriched in ERVK elements. *D. Ebrahimi[1,2,3], G. J. Starrett[1,2,3], N. A. Temiz[1,2,3], R. S. Harris[1,2,3].* 1) Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, MN 55455 USA; 2) Masonic Cancer Center, University of Minnesota, Minneapolis, MN 55455 USA; 3) Institute for Molecular Virology, University of Minnesota, Minneapolis, MN 55455 USA.

Nearly half of the human genome is composed of repetitive elements including DNA transposons and two main classes of retrotransposons. Human cells use several mechanisms including silencing by CpG methylation and inhibition by APOBEC3 editing enzymes to repress the activity of these endogenous elements (1). Both of these mechanisms can lead to C-to-T mutations through deamination of methylcytosine and cytosine bases respectively, however in different sequence contexts. Methylation drives mutation within CpG motifs and editing by all APOBEC3 enzymes except one occurs mainly within TpC (2). Recent studies have shown that C-to-T mutations within the contexts of CpG and TpC are also dominant in many cancers including breast carcinoma (3, 4). These observations suggest that the mechanisms responsible for inhibiting endogenous elements may also play a role in inducing cancer mutations. Importantly a large number of cancer mutations occur in non-genic regions composed of mainly endogenous elements. However studies of cancer mutations are mostly limited to exonic regions. Here we quantify the distribution of mutations in whole genome breast cancers to investigate whether certain mutations, in particular CpG methylation and APOBEC3 signatures are enriched within the endogenous element sequences of cancer genomes. An enrichment would suggest a link between the repression of endogenous elements and cancer mutations. Our analyses of 21 whole breast cancer genome sequences indicate that all elements are susceptible to 'spontaneous' mutagenesis but somatic mutations are enriched in ERVK elements. The most enriched mutations were C-to-T within CpG motifs suggesting a link between the methylation of these elements and their mutation in breast cancer. C-to-T mutation within the APOBEC3 preferred target TpC was not enriched in the endogenous retroviral sequences. These results suggest that a retrotransposon methyl-maintenance mechanism may be deregulated in some breast cancers. 1. N. Zamudio, D. Bourc'his, Transposable elements in the mammalian germline: a comfortable niche or a deadly trap? *Heredity* 2010, 105, 922. L. Lackey *et al.* , Subcellular localization of the APOBEC3 proteins during mitosis and implications for genomic DNA deamination, *Cell Cycle* 2013, 12, 7623. M. B. Burns *et al.* , APOBEC3B is an enzymatic source of mutation in breast cancer, *Nature* 2013, 494, 3664. L. B. Alexandrov *et al.* , Signatures of mutational processes in human cancer. *Nature* 2013, 500, 415.

2693W

How deep does intra-tumor genetic heterogeneity run in breast cancer? Identifying multiple different mutations in a single gene (intra-gene heterogeneity) within individual breast cancer tumors. *B. Gottlieb[1,2,3], F. Babrzadeh[5], K. Klein[1], C. Alvarado[1], C. Wang[5], B. Gharizadeh[5], M. Basik[2,4], L. K. Beitel[1,3,4], M. Trifiro[1,2,3,4].* 1) Lady Davis Inst Med Res, Montreal, QC, Canada; 2) Segal Cancer Center, Jewish General Hospital, Montreal, QC, Canada; 3) Dept Human Genetics, McGill University, Montreal, QC, Canada; 4) Dept Medicine, McGill University, Montreal, QC, Canada; 5) Stanford Genome Technology Center, Stanford University, Palo Alto, CA.

Recent studies that have revealed the degree and complexity of intra-tumor genetic heterogeneity within breast cancer tumors and between single tumor cells have been limited to identifying genetic alterations in different genes, but have yet to investigate the degree of genetic heterogeneity within the same gene (intra-gene heterogeneity, IGH). While next generation sequencing (NGS) has technically made this feasible, there are a number of issues that must be resolved before we can actually do so. In particular, present sequence analysis algorithms are designed so that they effectively ignore IGH, as there is the assumption that IGH is very unlikely to occur. Thus, most NGS analysis tools use filters and other techniques such as specific sequence alignment tools whose design has resulted in the reduced possibility of finding IGH in tumor tissues. To resolve this issue, we have developed a new approach to sequencing and sequence analysis that has allowed us to examine IGH within the androgen receptor gene (*AR*), even while removing sequence variance due to sequence or sequence analysis errors. The examination of a number of breast cancer tumors using this approach has revealed the presence of IGH within the *AR.* Initial data indicate that the vast majority of *AR* mutations in the breast cancer tumors exist only in minority forms, we believe these would have been totally missed using traditional sequencing and sequencing analysis tools. In light of the recent attention being paid to *AR* mutations, particularly in triple negative breast tumors, identification of minority *AR* mutations in breast cancer tissues could be important in considering treatment options. Furthermore, the inability to identify estrogen receptor mutations in most breast cancer patients, may be because they exist as minority variants within breast cancer tissues, Indeed, the presence of IGH could explain how both carcinogenesis and resistance to chemotherapy can be associated with rapidly occurring mutations in both tumor suppressor and oncogenes. It could also explain their significance in cancer ontology and treatment.

2694T

Complementing NGS panel sequencing by high rersolution custom array CGH increases the mutation detection rate in hereditary breast and ovarian cancer. *K. Hackmann[1], K. Kast[2], P. Wimberger[2], E. Schrock[1], A. Rump[1].* 1) Institut fuer Klinische Genetik, Medizinische Fakultaet Carl Gustav Carus, TU Dresden, Dresden, Sachsen, Germany; 2) Klinik und Poliklinik fuer Frauenheilkunde, Medizinische Fakultaet Carl Gustav Carus, TU Dresden.

As a diagnostic routine we perform next generation sequencing (NGS) and apply Illumina's TruSight Cancer panel to identify mutations that are causative for familial cancer predisposition in hereditary breast and ovarian cancers (HBOC). In order to also conduct a thorough detection of deletions and duplications we designed a customized array that covers all 94 genes that are present on the TruSight Cancer panel. In 11 families and 15 persons out of 232 individuals we detected copy number variants that were causative for tumor predisposition and affected the 10 genes that are recommended for diagnostic analysis by the German Consortium for Hereditary Breast and Ovarian Cancer . In addition to previously reported variants we found unreported deletions in *ATM*, *RAD51C* and an intragenic duplication in *BRCA2*. Besides adding about 5% to the total detection rate in HBOC cases, precise knowledge of the breakpoint positions makes it possible to carry out PCR for subsequent predictive testing of family members or segregation analysis. Generally, we can recapitulate copy number changes from NGS data to a certain degree. But in genomic areas with low read coverage we are not confident that NGS coverage analysis is sufficiently reliable. This may change with an increasing number of cases and appropriate software.

2695W

Somatic mutation detection in urological cancers from liquid biopsies. *C. Ionescu-Zanetti[1], R. Brobey[2], K. Rosenblatt[3], M. Dehghani[3], M. Schwartz[1], R. Amato[2].* 1) Research and Development, Fluxion Biosciences, South San Francisco, CA; 2) University of Texas Health Science Center at Houston; 3) Research and Development, CompanionDx.

Purpose: Next generation sequencing (NGS) of blood-derived nucleic acids is an emerging paradigm for determining the mutational status of cancer patients over time. Both circulating tumor cells (CTC) and cell-free circulating DNA have been proposed as possible sample types for extracting tumor DNA. Here, we present data from a CTC enrichment modality that results in tumor cell purities in excess of 10% followed by a high sensitivity NGS data analysis workflow. This study is aimed at urological cancers (kidney, prostate) that historically have not performed well for CTC analysis. **Methods:** Urological cancer blood samples were enriched for CTCs using immunomagnetic separation in a microfluidic chamber. CTCs were lysed and amplified, followed by DNA purification. Targeted libraries were sequenced and data was analyzed using a customized variant pipeline based on standard alignment tools and functional interpretation. Matched blood samples were enumerated to determine the CTC load (CK+, CD45-, nucleated cells). **Results:**Multisite analytical validation data, based on spiking of cells into whole blood, demonstrated a detection limit down to 10 cells from a blood draw with a false positive rate of below 0. 1 calls per sample. Clinical data from two different urological cancer pilot studies (prostate and kidney) demonstrated the detection of somatic variants for a majority of samples with significant overlap between detected mutations and known somatic mutation sites. We detected common mutations (i. e. TP53, PTEN and APC genes) that are similar to the population distribution of mutation rates in tissue biopsies. Over 90% of samples tested had CTCs above the detection limit (10 tumor cells). **Conclusions:** This assay makes possible the detection of somatic variants from urological cancer patients without the need for a tissue biopsy.

2696T

Case analysis of advanced cancers with Watson Genomic Analytics. *T. Koyama[1], S. Jones[2], F. Utro[1], Y. Ma[2], K. Rhrissorakrai[1], Y. Shen[2], B. Carmeli[3], M. Jones[2], Z. Waks[3], E. Plesance[2], R. Norel[1], R. Moore[2], E. Bilal[1], A. Mungal[2], K. Beaty[1], J. Schein[2], V. Michelini[4], M. Marra[2], A. Royyuru[1], J. Laskin[2].* 1) IBM Research, Yorktown Heights, NY; 2) British Columbia Cancer Agency, Vancouver, British Columbia, Canada; 3) IBM Research, Haifa, Israel; 4) IBM Watson, New York, NY.

The personalized onco-genomics project (POG) at British Columbia Cancer Agency (BCCA) aims to use paired tumour/normal whole genome and transcriptome sequence information from a cancer patient to inform treatment options. The analysis and interpretation of this large and complex data set is difficult to complete within a clinically relevant timeframe. As whole genome sequencing becomes more readily available there is a need to scale-up these interpretive processes but maintain a high degree of evidence and data integrity. This project compares the output of the human curated literature review and data analysis methods employed at the BCCA with that of Watson Genomic Analytics (WGA), a novel IBM Watson solution. The POG project has biopsied, sequenced and analyzed over 100 metastaticcancers since 2012. Each case has whole genome and transcriptome sequencing andthe data is mined to identify somatic mutations or disrupted signaling cascadesthat might be functionally critical for understanding driver events within anindividual's cancer or useful for rationalizing treatment planning. Anonymized data from 29 cases was input into WGA which readsmutation, copy number and gene expression files. It performs a molecularprofile analysis to identify possible driver events, followed by a pathway analysis to find possible pharmaceutical interventions including phase 3 candidates. Pathways, off-target pharmacological effects, and pharmacogenomics/biomarker information are augmented with relationshipsdiscovered by natural language processing from millions of PubMed abstracts and full text biomedical journals. Theoutput of WGA consists of a list of drug candidates with rationales and scoresalong with literature evidence that supports their consideration. The results were compared to the human curated reports produced by the POG team. In all 29 instances of retrospective analysis, WGA finds actionable insightsand identifies potential drugs for consideration along with visualization for molecular profile analysis and pathways. The automated generation ofthese insights is achieved by WGA in minutes. With increasing adoption of molecular targeted therapies, there is a strongneed to continuously update the relationships between genomic alterations,cancer biology and drugs. This is a task that requires extensive and time consuming review of latestscientific literature. WGA may allow for more efficient androbust way of performing the analysis in precision medicine.

2697W

Highly Sensitive and Cost-Effective Detection of *BRCA1* and *BRCA2* Cancer Variants in FFPE Samples Using Multiplicom's MASTR Technology & Single Molecule, Real-Time (SMRT®) Sequencing. *S. Kujawa[1], A. Sethuraman[1], P. Baybayan[1], J. Del Favero[2].* 1) Pacific Biosciences, Menlo Park, CA; 2) Multiplicom N. V. , Niel, Belgium.

Specific mutations in *BRCA1* and *BRCA2* have been shown to be associated with several types of cancers. Molecular profiling of cancer samples requires assays capable of detecting the entire spectrum of variants. Next-Generation Sequencing (NGS) has been a powerful tool for researchers to better understand cancer genetics. Single Molecule, Real-Time (SMRT) sequencing offers several advantages, including sensitive detection of variants and long read lengths capable of spanning entire protein coding regions. We examined the detection sensitivity of SMRT sequencing in combination with the BRCA MASTR Dx kit from Multiplicom, amplifying all coding exons of *BRCA1* and *BRCA2* covered by 181 overlapping amplicons ranging in size between 120 bp and 230 bp. FFPE derived DNA samples were multiplex PCR amplified followed by a second round of PCR amplification enabling sample multiplexing by adding a unique set of 16-bp PacBio barcodes to either end of each amplicon. The resulting amplicons were further purified, concentrations normalized, and mixtures made prior to sequencing. SMRTbell™ libraries were constructed and sequenced using standard Pacific Biosciences® P6-C4 chemistry and protocols. With a 90-minute movie collection protocol, the average polymerase read lengths were approximately 8,000 bp across samples, with 5% of reads longer than 20,000 bp. For coverage and variant detection analysis, we sequenced a number of sample combinations, from 1 sample per SMRT Cell up to several samples per SMRT Cell. With 4 samples multiplexed in a single SMRT Cell, we obtained enough circular consensus sequencing (CCS) coverage of all 181 amplicons for efficient downstream variant analysis. Using either a reference-based approach or a *de novo* approach for variant analysis, we accurately and reliably detected all the expected variants for the samples in the mix. The random error profile, long read lengths, and high accuracy CCS reads make it possible to call variants from as few as 10 molecules. While SMRT sequencing can identify cancer mutations with high accuracy, the associated low cost of consumables and fast turnaround time make this an ideal platform for clinical research applications.

2698T

Clinical validation of a droplet PCR-based targeted gene panel for the detection of somatic variants in solid tumors. *L. Liang[1], F. B. Patel[1], Q. Tan[1], H. Gaspard[1], W. M. Hussain[1], A. I. Brooks[1], J. A. Tischfield[1], A. Sahota[1,2].* 1) Department of Genetics, Rutgers University, Piscataway, NJ; 2) Department of Pathology and Lab Medicine, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ.

Introduction: Next generation sequencing enables the sensitive detection of sequence variants in multiple genes from small amounts of tumor tissue. The resulting mutational profile may serve multiple purposes, including personalized patient care, enrollment in clinical trials, or basic research into the molecular heterogeneity of cancer. Here, we describe the clinical validation of the ThunderBolts (TB) Cancer Panel from RainDance Technologies for the detection of sequence variants in DNA from FFPE tissue from a variety of solid tumors. This panel targets 230 amplicons spanning 50 genes. Many of the targeted mutations are in hotspot regions and therapies are available for a subset of the molecular pathways involving these mutations. **Methods:** DNA was extracted from FFPE tissue and a minimum of 10 ng DNA in a total volume of 40 µl was analyzed using the TB panel. The reaction mixture was subjected to picoliter-volume droplet formation and the droplets then amplified by PCR. The amplified products were purified and a second round of PCR was carried out to add barcoding sequences. Following further purification, PCR products from up to 16 samples were pooled and sequenced on the Illumina MiSeq. Nucleotide sequence data were analyzed using MiSeq Reporter and associated software. **Results:** We analyzed DNA from 59 FFPE samples (51 tumor and 8 normal adjacent tissue) and three well-characterized human cell lines. For the FFPE samples, the average read per sample was 2. 1x106, with an average read of 10,083 per amplicon and over 99% of the amplicons had a read depth greater than 1,000. The assay was highly reproducible, both for cell line and FFPE DNA. Based on analysis of cell line DNA, the analytical sensitivity and specificity was 98. 6 and 100%, respectively. The Life Technologies Ion AmpliSeq Cancer Hotspot Panel v2 was used to confirm, in 27/59 of the FFPE samples, the sequence variants identified by the TB panel. Sequence variants down to 5% could be readily detected using cell line mixtures and this was confirmed in clinical samples. **Conclusions:** The TB panel eliminates the need for traditional library preparation and it has additional advantages over other targeted panels, including the ability to amplify degraded DNA and precise detection of low-prevalence somatic mutations based on single molecule PCR. The assay is highly reproducible and we have validated it for use in the clinical laboratory for the detection of sequence variants in solid tumors.

2699W

***Cis*-regulatory drivers in colorectal cancer.** *H. Ongen[1], C. L. Andersen[2], J. B. Bramsen[2], P. G. Ferreira[1], T. F. Orntoft[2], E. T. Dermitzakis[1].* 1) Department of Genetic Medicine and Development, University of Geneva Medical School, 1211 Geneva, Switzerland; 2) Department of Molecular Medicine, Aarhus University Hospital, Aarhus, Denmark.

The cis-regulatory effects have been shown to be important in colorectal cancer (CRC) development (Ongen H et al. , *Nature*, 2004). To better characterise the impact of the regulatory effects on CRC tumorigenesis, we conducted an RNA-sequencing experiment on 300 matched tumour and adjacent normal colon mucosa samples from CRC patients, which are also germline genotyped. On average there are 529 significant allele-specific expression (ASE) signals (FDR = 1%) per sample. Utilizing the changes in ASE in matched pairs of samples we discover 261 Genes with Allelic Dysregulation (GADs – or genes with cis regulatory driver mutations), enriched for TCGA PANCANCER drivers. We also correlated genotypes and gene expression to identify expression quantitative trait loci (eQTLs) and find 7114 and 2462 cis-eQTLs in normals and tumours, respectively. We estimate that 10% of the tumour cis-eQTLs are exclusive to CRC. Importantly tumour specific cis-eQTL genes also accumulate more cancer somatic mutations when compared to the shared cis-eQTL genes, raising the possibility that some of them may be germline-derived cancer regulatory drivers. We have investigated the trans effects on gene expression, and preliminary results show the there are 9 and 7 trans-eQTLs, in normal and tumour respectively, which are shared between the cell types. Furthermore we are assessing the trans effects of cis-eQTLs in order to increase our power in detecting regulatory variants that may act in trans. We are also using ASE differences in matched samples and the correlation of these differences with clinical characteristics to discover a cascade of regulatory changes that happen in tumorigenesis. In conjunction with this approach we are also assessing whether eQTL effects correlate with tumour progression. Furthermore we are investigating if ASE differences classify CRC samples into subtypes and whether these correlate with CRC phenotypes. We are also extending our analysis for finding GADs to un-matched normal and tumour samples, which will facilitate discovery of putative somatic regulatory drivers in un-matched cohorts. Moreover, the GADs are being assessed for their driver potential by comparing the expression levels in tumour vs. normal samples. Collectively the integration of genome and the transcriptome reveals a substantial number of putative somatic and germline cis-regulatory drivers.

2700T

Gene polymorphisms vis-a vis susceptibility to breast cancer and benign breast disease: A case-control study from Jammu region of J&K State, India. *R. K. Panjaliya[1], R Sharma[2], J Kour[3], S Bhardwaj[4], T Ajad[5], P Kumar[1, 6].* 1) Human Genetics Research cum Counselling Centre, University of Jammu, Jammu, J & K, India; 2) Human Genetics Research cum Counselling Centre, University of Jammu, Jammu, J & K, India; 3) Human Genetics Research cum Counselling Centre, University of Jammu, Jammu, J & K, India; 4) Department of Pathology, Govt. Medical College, Jammu, J&K; 5) Department of Surgery, Govt. Medical College Jammu, J&K; 6) Department of Zoology, University of Jammu, J&K.

Breast cancer is the most common cause of death in women all over the world. It is the most frequent cancer of both developed and developing regions. In India, breast cancer is the second most common cancer in females after cervical cancer and accounts for 22. 2% of all new cases diagnosed and 17. 2% of all cancer deaths in women. Molecular epidemiological studies have revealed the association of different gene polymorphisms with breast cancer risk. In the study under report, gene polymorphisms namely MTHFR C677T, ESR- PvuII, ESR -XbaI, CYPI-AI-Msp I, ACE I/D, GSTMI and GSTT1 have been genotyped in a total of 316 females including 60 confirmed patients of sporadic breast cancer, 105 benign breast disease and 150 age matched healthy females from Jammu region of J&K state. Polymerase chain reaction (PCR) was used to determine the genotypes in case of ACE I/D, multiplex PCR for GSTMI and GSTTI and PCR-RFLP method for remaining polymorphisms. Besides, modifiable and non-modifiable risk factors were also considered for the gene-environment interaction study and gene-gene interaction was also performed by using MDR analysis. Out of these polymorphisms GSTMI null genotype showed a significant association with the risk of breast cancer. T allele of MTHFR C677T polymorphism also conferred increased risk for breast cancer with significant p value. All other polymorphisms showed a lack of association with risk of breast cancer and benign breast disease as well. The small sample size is limitation of the present research work which seeks a study on large sample size that we intend to do.

2701W

Whole exome sequencing and copy-number variation analysis of 20 neurofibromatosis type 2-associated spinal and cranial meningiomas. *A. Pemov[1], R. Dewan[2], N. F. Hansen[3], S. Chandrasekharappa[3], J. C. Mullikin[3], A. R. Asthagiri[4], D. R. Stewart[1], NIH Intramural Sequencing Center, Comparative Sequencing Program.* 1) NIH/NCI, Bethesda, MD; 2) University of Maryland School of Medicine, Class of 2017, Baltimore, MD; 3) NIH/NHGRI, Bethesda, MD; 4) Department of Neurosurgery, University of Virginia School of Medicine, Charlottesville, VA.

Recent studies in high-throughput sequencing of genomes of sporadic meningiomas have revealed that inactivation of *NF2* is the most frequent genetic event in the tumors and is mutually exclusive with mutations occurring in genes such as *TRAF7, KLF4, AKT1* and *SMO*. It is accepted that somatic inactivation of the second copy of the *NF2* gene is associated with tumor initiation in patients with NF2, however little is known about what other genes or pathways are involved in the tumor formation and progression in patients with this disorder. To investigate this question we performed whole exome sequencing (WES) (Illumina Hi-Seq 2500 platform, 96 Mb SeqCap EZ Exome + UTR Library, NimbleGen) and SNP-array analysis (HumanOmniExpressExome-8, v1. 2 arrays, Illumina) of 20 spinal and cranial meningioma samples from seven NF2 patients. We identified 4 nonsense germline *NF2* mutations, a splice site mutation and an intronic mutation near the acceptor splice site (-9 bp) in six patients. SNP-array analysis revealed a large deletion in chromosome 22 containing the *NF2* gene in all but one meningioma sample. The second most frequent chromosomal aberration was a deletion within chromosome 1p (2 tumors) followed by the entire chromosome X deletion and complex rearrangements in chromosome 17q. The remaining samples exhibited the genomic architecture of a normal diploid cell. Next, we investigated the somatic mutation burden in the genomes of the tumors. We identified ~20 point substitutions and small indels in 20 tumor samples. Most of the mutations were observed only once. We identified a few mutations that were common among multiple samples, however none of these mutations were shared among the samples originating from different patients. Validation of these mutations by an orthogonal sequencing technology is pending. In conclusion, our genomic study revealed that somatic inactivation of *NF2* is the most frequent and the only recurrent genetic event in NF2-associated meningiomas. Large LOH events such as large deletions in chromosome 22 are by far the most prevalent mechanism of the second hit. Large LOH events elsewhere in the genome, especially on chromosome 1, could represent a common path of tumor progression toward a more advanced stage. Somatic single nucleotide substitutions and small indels are relatively rare in these tumors and understanding the role of these types of mutations in NF2-associated meningioma tumorigenesis will require further studies.

2702T

Driver gene mutations and fusion genes in patients with sezary syndrome. *A. Prasad[1], R. Rabionet[1], L. Zapata[1], A. Puig[1], Y. S. Trujillo[1], A. M. Puiggros Metje[2], B. Espinet[2], F. Gallardo[2], R. M. Pujol[2], X. Estivill[1].* 1) Bioinformatics and Genomics, Centre for Genomic Regulation (CRG), Barcelona, Barcelona, Spain; 2) Laboratorio de Citogenetica Molecular, Servicio de Patologia, Hospital del Mar, Barcelona, Spain.

Sezary syndrome is one of the most common types of cutaneous T-cell lymphoma and is characterized by extensive red, itchy rashes all over the body. The prevalence of Sezary syndrome is around 0. 3 cases per 100,000 people. The genetic etiology of the disease is poorly understood with chromosomal abnormalities and mutations in some genes shown to be involved in the disease. The goal of our study is to understand the genetic basis of the disease by looking for driver gene mutations and fusion genes in a matched tumor and normal pair of 17 Sezary cases. We have undertaken a whole-exome, RNA and small RNA sequencing approach. We have identified somatic point mutations in genes including *TP53*, *ITPR1*, *DSC1*, *PLCG1*, *CARD11*, *GLI3* and some of these genes are mutated in more than one individual. We have observed several somatic copy number variations including a recurrent deletion on 17p (~15 Mb) in six of our cases. Furthermore, we have discovered several fusion RNAs. Of particular interest are the ones involving *RASA2* and *BCR* genes due to their relevance in a variety of human cancer types. In addition, we see specific profiles of deregulation of microRNAs in our cases. Currently, we are correlating the clinical phenotype of patients with somatic mutation spectrum, fusion RNAs, RNA and microRNA expression profiling to better understand the disease pathogenesis.

2703W

Microfluidic Single Cell Exome-seq and RNA-seq analysis of Tumor Composition. *I. Ragoussis[1], P. Savage[3], C. Batchu[2], Y-C. Wang[1], T. Ravil[1], D. Badescu[1], E. Iacucci[1], L. Letourneau[1], A. Monast[3], N. Bertos[3], A. Omeroglou[4], M. Park[3].* 1) McGill University & Genome Quebec Innovation Centr, McGill University, Montreal, Quebecc, Canada; 2) Fluidigm Corporation,7000 Shoreline Boulevard, Suite #100 South South San Francisco, CA 94080 United States; 3) The Rosalind and Morris Goodman Cancer Research Centre, McGill University, Montreal, Quebec, Canada; 4) Department of Pathology, McGill University, Montreal, Quebec, Canada;.

Human breast tumors have been shown to exhibit extensive inter- and intra-tumor heterogeneity. While recent advances in genomic technologies have allowed us to deconvolute this heterogeneity, few studies have addressed the functional consequences of diversity within tumor populations. Here, we identified an index case for which we have derived a patient-derived xenograft (PDX) as a renewable tissue source to identify subpopulations and perform functional assays. On pathology, the tumor was an invasive ductal carcinoma which was hormone receptor-negative, HER2-positive (IHC 2+, FISH average HER2/CEP17 2. 4), though the FISH signal was noted to be heterogeneous. On gene expression profiling of bulk samples, the primary tumor and PDX were classified as basal-like. We performed single cell RNA and exome sequencing of the PDX to identify population structure. Using a single sample predictor of breast cancer subtype, we have identified single basal-like, HER2-enriched and normal-like cells co-existing within the PDX tumor. Genes differentially expressed between these subpopulations are involved in proliferation and differentiation. Functional studies distinguishing these subpopulations are ongoing. Microfluidic whole genome amplification followed by whole exome capture of 81 singe cells showed high and homogeneous target enrichment with >75% of reads mapping uniquely on target. Variant calling using GATK and Samtools revealed founder mutations in key genes as BRCA1 and TP53, as well as subclonal mutations that are being investigated further. Loss of heterozygocity was observed in 16 TCGA cancer driver genes and novel mutations 7 cancer driver genes. These findings may be important in understanding the functional consequences of intra-tumor heterogeneity with respect to clinically important phenotypes such as invasion, metastasis and drug-resistance.

2704T

***Helicobacter* DNA integrations in the gastric cancer genome.** *K. M. Robinson[1], N. Kumar[1], J. Torres[2], J. C. Dunning Hotopp[1,3,4].* 1) Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA; 2) Unidad de Investigación en Enfermedades Infecciosas, UMAE Pediatría, IMSS, México City, DF, México; 3) Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD, USA; 4) Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, MD, USA.

There are 10 times more bacterial cells in the human body than human cells, and various bacteria are known to influence carcinogenesis. *Helicobacter pylori* is considered a carcinogen by the World Health Organization due to its ability to promote carcinogenesis in gastric tissue. While viral DNA integrations in the human genome have been shown to promote carcinogenesis, bacterial DNA integrations (BDI) into the human genome are rarely investigated. Our previous analysis of publicly available sequencing data from the Cancer Genome Atlas showed BDIs from *Pseudomonas* spp. rRNA into proto-oncogenes of gastric cancer samples in the 5'-UTR of four proto-oncogenes as well as in *Ig* (the immunoglobulin gene). However, our inability to obtain the materials required to validate our findings in these samples left us with many unanswered questions. Therefore, we sought to sequence a different cohort of patients in order to identify BDIs in gastric tumors using transcriptome sequencing data with further sequencing and analysis of whole genome and whole exome sequencing for BDI positive samples. Using our previously published BWA-based pipeline, we have identified putative BDIs from *Helicobacter pylori* rRNA into numerous genes in one tumor sample as well as in its adjacent matched sample, including BDIs in the *Ig* locus. Validation of these results is ongoing and will be presented. Following validation, tests examining the carcinogenic potential of these insertions can be conducted. Given these results, more consideration should be given to bacterial DNA integrations into the human genome when bacterial associations with diseases are suspected.

2705W

Somatic mutation load and risk of colorectal cancer: NGS-study of a large panel of healthy and diseased individuals. *N. T. Rodchenko[1], M. P. Salomon[1,2], P. Marjoram[2], S. V. Nuzhdin[1].* 1) Molecular and Computational Biology, University of Southern California, Los Angeles, CA; 2) Division of Biostatistics, Department of Preventive Medicine, Los Angeles, CA.

DNA mutations are the source of evolution, but also a hindrance to organism survival. There is selection to reduce the germline mutation rate, that is known to occur at a rate of 10^{-10} to 10^{-8} changes per nucleotide per generation among higher eukaryotes. However, the somatic mutation (SM) rate is under a lesser selection constraint as the soma is disposable. From selection assay experiments it appears that SMs arise one or two orders of magnitude more rapidly than germline variants. Mutations, ranging from single base substitutions to large genome rearrangements, gradually accumulate in cells as the organism ages with an intensity dependent on the cells' exposure to damage, level of mis-functioning of repair genes and the speed of tissue proliferation. There are many outcomes of accumulation of SMs, with age being a major determinant of risk of human cancer and other diseases. A study of a large number of case and control samples from accessible healthy tissues, such as blood, is required to understand whether overall rate of SMs and presence of particular hot spot mutations can be indicative of a higher risk of the disease.

In this work we examined the abundance of single-nucleotide somatic mutations (SNSMs) in blood samples of two large panels: (1) individuals with known colorectal cancer (CRC) diagnosis and (2) healthy subjects (~3700 samples in total). We used targeted genomic sequencing data for regions spanning CRC-associated GWAS peaks (~5. 5 Mbp per sample) and developed a pipeline to call SNSMs in a conservative way using a subset of bi-allelic genomic positions sequenced at a depth of at least 100x. We used available phenotypic data (age, sex, ethnicity and medical records of the individuals) to examine associations between elevated SM rates and phenotypic parameters.

Results show association between the number of SNSM's and age, sex and ethnicity of the individuals, as well as significant differences between colon cancer patients and healthy individuals with respect to proportions of transitions and transversions in coding regions . Follow-up mutation annotation shows the presence of SNSMs that have been found to be functional in this group of cancers in previous studies and that are potential hot spots in the colorectal cancer cohort.

2706T

Enhanced error correction and increased sensitivity in variant calling in FFPE samples with cancer hot spot amplicon panels. *T. Singer, I. Khrebtukova, H. Zhang, A. Barr, L. Way, A. Wong, G. DeSantis, G. P. Schroth.* Genomic Applications, Illumina Inc. , San Diego, CA.

Low frequency variant calling in FFPE DNA is notoriously difficult due to high false positive error rate in those samples. We designed a palette of small amplicon cancer "hot spot" panels to perform targeted amplification of 50 oncogenes and tumor suppressor genes to investigate various error correction methods. The probe panels were designed by modifying the design pipeline of the TruSeq custom amplicon workflow to allow tiling designs of very small overlapping amplicons targeting both strands of DNA. Each panel represents a single tube assay and is comprised of ~300 amplicons with an average size of 110bp, targeting 1,917 Cosmic SNPs. A 100% coverage of target regions was achieved. The panels were designed in "mirror" fashion, to enable interrogation of opposite strands of DNA. In addition, panels were modified to include bar codes or UMIs ("unique molecular identifiers") to enable molecular counting. These amplicon panels are uniquely suited to enable investigation of low frequency somatic variant calling in degraded FFPE samples. We will present data testing various strategies to reduce false positive calls in Illumina dual strand sequencing data. A bioinformatics analysis pipeline was developed to reduce error rate, increase quality scores and enhance low frequency variant calling in next generation sequencing reads. A comparison of dual strand amplicon strategy, the use of replicate FFPE samples and benefits of UMIs will be presented.

2707W

Evaluation of single-molecule real-time long-read sequencing as a rapid-turnaround tool for validation of somatic mutations in cancer genomics. *A. Uzilov, L. Newman, I. Bourzgui, C. Lau, E. E. Schadt, R. P. Sebra, R. Chen.* Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY.

Background: Next-generation sequencing of tumor DNA is on the rise in research and in clinical diagnostics. Because tumor specimens are heterogeneous, limited, and may have low tumor content, clinical-grade accuracy is a challenge. Efficient validation by an orthogonal platform is needed. **Methods:** We assess targeted single-molecule real-time (SMRT) sequencing on the PacBio RSII platform as a high-coverage, long-read (>1000 nt) method to validate mutations found in lower-coverage, short-read (2x100 nt) whole-exome sequencing (WES) on Illumina HiSeq 2500. Actionable somatic mutation candidates (SNVs and indels) were identified from running WES on tumor/normal genomic DNA from cancer patients as part of a study on clinical actionability of cancer sequencing. Amplicons of size ~500 nt were made that included ~250 nt flanks around each WES-identified mutation. Each circularized amplicon molecule was read several times by a continuous long read; consensus reads were assembled, aligned, and analyzed using PacBio and in-house tools. **Results:** Normal/tumor DNA from five patients were sequenced using SMRT for validation (13 to 27 WES candidates selected per patient). 96-100% of the candidates were validated by SMRT sequencing. For one patient (27 candidates), a primary tumor (frozen, ~50% purity based on WES data) and a metastasis (FFPE, ~10% purity based on WES data) were available. Two mutations called as unique-to-primary from WES were found in the metastasis on SMRT (allelic fractions <2%, mean coverage 1993X of metastatis sample sequencing run), showing the tumors were more similar than would be inferred from WES alone (mean coverage 223X of metastatis sample sequencing run). Allelic fractions from SMRT and WES agreed that metastatic tumor was < 10% pure. Turnaround time per patient from submission of candidates and DNA to sequencing data can be as little as four days, needing 10-50ng of gDNA per candidate. **Conclusions:** SMRT sequencing is an effective method for validation of variants on a clinical timescale, resolving low-purity tumors that would be harder to interrogate with Sanger-based or qPCR assays. This can also be used for personalized, low-cost tracking of tumor genome evolution, e. g. regions identified from WES could be cheaply, rapidly re-sequenced on SMRT in recurrent and metastatic tumors to see how the cancer genome evolves during disease course.

2708T

Comprehensive characterization of the cancer genome by integrating targeted DNA and RNA sequencing. *P. Van Hummelen[1,3], R. Abo[1,3], S. Hunter[1,3], L. Lin[1,3], M. Ducar[1,2], A. Thorner[1,3], W. C. Hahn[1,3,4], M. L. Meyerson[1,3,4], E. P. Garcia[2], L. M. Sholl[2], L. E. MacConaill[1,2,3].* 1) Center for Cancer Genome Discovery, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA; 2) Department of Pathology, Brigham & Women's Hospital, Boston, MA; 3) Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA; 4) Broad Institute of MIT and Harvard, Cambridge, MA.

Genome variant analysis (DNA-seq) and transcriptome profiling by sequencing (RNA-seq) provide multiple levels of insight into the cancer genome. These approaches generate endpoints such as somatic mutations, rearrangements, copy number aberrations and expression analysis. Combining DNA and RNA-seq data will add functional confirmation by quantifying the expression level of the genomic variant or its effect on gene expression down stream of its pathway. Furthermore, chimeric transcripts that result from structural rearrangements in the genome can also be corroborated by detection and quantification at the RNA level. This poster will give an overview of our DNA&RNA targeted sequencing platform, including specific tools for rearrangement detection (Breakmer), copy number analysis (RobustCNV) and allele specific expression analysis. Our targeted DNA-seq pipeline has been established and translated to the clinical setting in 2013 as part of PROFILE, a collaboration between Dana-Farber Cancer Institute, Brigham and Women's Hospital and Boston Children's Hospital, to preemptively screen every admitted cancer patient for a targeted genome analysis of the tumor. For RNA-seq we have developed and characterized protocols and analysis procedures and compared three different RNA-seq strategies (whole-transcriptome, exome and targeted RNA-seq) investigating the implications of the choice of strategy on the downstream analysis and results. In addition, we show a proof of principle for verifying DNA variants by RNA-seq, including activating mutations in *EGFR* and loss of function variants in tumor suppressor genes such as *STK11* and *RB1*. Our protocols are made compatible with archival clinical tissues and amenable for clinical sequencing on a large scale.

2709W

Genomic alterations associated with clonal hematopoietic expansion and malignancies inferred in normal human blood samples. *M. Xie[1,2,3], M. Wendl[1,2,4], L. Ding[1,2,3,5].* 1) McDonnell Genome Institute, Washington University, Saint Louis, MO; 2) Department of Medicine, Washington University, Saint Louis, MO; 3) Department of Genetics, Washington University, Saint Louis, MO; 4) Department of Mathematics, Washington University, Saint Louis, MO; 5) Siteman Cancer Center, Washington University, Saint Louis, MO.

Blood cancer formation is a long-term clonal expansion process. It often takes a couple of decades for a patient to develop a clinically detectable tumor that is severe enough to be diagnosed or produce symptoms, so some early mutations are more likely to exist many years in patients who have no apparent disease symptoms. These mutations have the potential to confer advantages in self-renewal or proliferation, resulting in clonal expansion, and causing hematopoietic malignancies by cooperating with subsequent mutations. Therefore, identifying these early mutations and monitoring clonal expansions in normal individuals could facilitate development of prevention strategies to reduce the risk of progression to cancer. Here, we collected whole-exome sequencing data from 6,099 TCGA cancer patients who only have first-time primary cancers and have had no treatment with radiation and/or chemotherapy. To systematically detect the variants in blood samples, we applied three widely used softwares, VarScan (for SNV and indel detection), GATK (for SNV and indel detection), and Pindel (for indel detection). After false-positive filtering and variant allele fraction comparison, we identified 81 recurrently mutated genes with blood-specific somatic truncations that are present in blood sample (variant allele fraction > 10%) but not present or present only at very low levels in either the tumor samples or tumor-adjacent normal samples. As expected, *DNMT3A*, *ASXL1* and *TET2* are the top 3 most recurrently mutated genes. Besides, we also found *KDM6A*, *MLL3*, *PPM1D*, *PTPRN2* are also frequently mutated in normal individuals. To validate the functional impacts of these new identified genes or mutations on clonal hematological expansion, further analysis is still needed. This study will provide a comprehensive profiling of genes and variants that initiate clonal expansion. It will enable to shed light on understanding of mechanism underlying hematological malignances, and facilitate to develop new strategies for early detection and prevention of hematologic cancer.

2710T

Frequent alterations in cytoskeleton remodeling genes in primary and metastatic lung adenocarcinomas. *X. Zhao[1], K. Wu[1], X. Zhang[2], F. Li[1], D. Xiao[2], Y. Hou[1], S. Zhu[1], D. Liu[1], X. Ye[1], X. Xu[1], H. Yang[1], K. Kristiansen[4], J. Wang[1], N. Zhong[2], J. Wang[1], Q. Pan-Hammarström[3], J. He[2].* 1) Beijing Genomics Institute (BGI-shenzhen), Shenzhen, Guangdong, China; 2) The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China; 3) Karolinska Institutet, 14186, Stockholm, Sweden; 4) University of Copenhagen, Copenhagen 1599, Denmark.

Lung adenocarcinoma results in a high mortality globally. The landscape of genetic alterations in tumours derived from Asian patients is however largely uncharacterized. Here we present an integrated genomic and transcriptomic analysis of 244 primary lung adenocarcinomas and 35 corresponding lymph node metastases from Chinese patients, of which 60% were at an advanced disease stage. A total of twelve significantly mutated genes were identified, including two novel mutation targets, *RHPN2* and *GLI3*. *TP53* was the most commonly mutated gene and mutations in this gene were significantly enriched in tumours from patients harbouring metastases. Genes regulating cytoskeleton remodeling processes were frequently altered, especially in metastatic samples, of which the high expression level of IQGAP3 was identified as a marker for poor prognosis. Functional studies further demonstrated that IQGAP3 facilitates tumour cell invasion. Our study represents the first large-scale sequencing effort on lung adenocarcinoma in Asian patients and provides a comprehensive mutational landscape for both primary and metastatic tumors. Our findings thus form a basis for future personalized medical care and may also shed light on the molecular pathogenesis of metastatic lung adenocarcinoma.

2711W

Mutation spectra of *BRCA* genes in Iranian women with early onset breast cancer, 15 years experiences. *V. R. Yassaee[1,2], Z. Soltani[1], Z. Ravesh[1], F. Hashemi-Gorgi[1], S. M Poorhosseini[1,2], R. Anbiaee[3], A. Joulaee[4].* 1) Genomic Resreach Center, Shahid Beheshti University of Medical Sciences, Tehran, IRAN; 2) Dept. of Medical Genetics. Faculty of Medicine Shahid Beheshti University of Medical Sciences Tehran, IRAN; 3) Dept. of Radiotherapy & Oncology, Imam Hossein hospital, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran; 4) Dept. of Surgery, Mahdieh, Women's hospital, School of Medicine, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Breast cancer is the most common cancer in Iran. In the recent years an upward trend (10%) has been observed in the Iranian population. Early detection by molecular approaches may reduce breast cancer morbidity and mortality. We consulted 3781 women diagnosed with early onset breast cancer during the past 15 years (1999-2014). To establish a data set for *BRCA* gene alterations in at risk Iranian families, two hundred and fifty three women who met the criteria were selected. A total number of 46 alterations including 17 variants with unknown clinical significance (37%), 19 Missense mutations (41%), 8 Indels (17%) and 3 large rearrangement sequences (6%) were identified. Further scanning of affected families revealed that 48% of healthy relatives harbor identical causative mutations. This is the first report of comprehensive *BRCA* analysis in Iranian women with early onset breast cancer. Our findings provide valuable molecular data supporting physicians as well as patients for the best decision making on disease management.

2712T

iCAGES: integrated CAncer GEnome Score for comprehensively prioritizing cancer driver genes in personal genomes. *C. Dong[1], H. Yang[1], Z. He[3], X. Liu[2], K. Wang[1].* 1) Keck School of Medicine, University of Southern California, Los Angeles, CA, 90089, USA; 2) Division of Epidemiology, Human Genetics and Environmental Sciences, The University of Texas Health Science Center at Houston, Houston, TX, 77030, USA; 3) Department of Computer Science, New York University, New York, NY 10012, USA.

All cancers arise as a result of the acquisition of somatic mutations that drive the disease progression. Even though a number of computational tools have been developed to identify driver genes for a specific cancer from a group of cancer samples, it remains a challenge to identify driver mutations/genes for an individual patient and design drug therapies. To tackle this challenge, we developed iCAGES (integrated CAncer GEnome Score), a novel statistical framework to rapidly analyze patient-specific cancer genomic data, prioritize personalized cancer driver events and predict personalized therapies. iCAGES includes three consecutive layers: the first layer utilizes machine learning models and integrates contributions from coding, non-coding and structural variations to infer driver variants. The second layer identifies driver genes for different cancer subtypes, using information from the first layer and integrating prior biological knowledge on gene-gene and gene-phenotype networks. The third layer prioritizes personalized drug treatment, by classifying potential driver genes into different categories and querying drug-gene databases. Compared to currently available tools, iCAGES achieves better performance by correctly classifying point coding driver mutations (AUC=0. 97, 95% CI: 0. 97-0. 97, significantly better than the second best tool with P=0. 01) and genes (AUC=0. 93, 95% CI: 0. 93-0. 94, significantly better than MutSigCV with P<1X10-15). We also illustrated two examples where iCAGES correctly nominated two targeted drugs for two advanced cancer patients with exceptional response, based on their somatic mutation profiles. iCAGES leverages personal genomic information and prior biological knowledge, effectively identifies cancer driver genes and predicts treatment strategies. iCAGES is available at http://icages. usc. edu.

2713W

Sensitive Mutation Detection By Sequencing Circulating Cell-Free DNA. *N. Fang, R. Tolun, K. Keitz, C. Baron.* QIAGEN, Hilden, Germany.

Circulating DNA, the cell-free DNA (cfDNA) found in serum or plasma, has become a powerful tool in non-invasive prenatal testing (NIPT), as well as in cancer liquid biopsy. In cancer testing, it has been shown that the quantity and integrity, as well as the mutation content of the cfDNA in cancer patients differ from that in healthy controls. Therefore, cfDNA may serve as a biomarker for cancer diagnosis, prognosis, and stratification. High-throughput sequence analysis of the cfDNA using next generation sequencing (NGS) technologies provides a highly sensitive and specific method in detecting and characterizing somatic mutations in cancer samples. However, the cfDNA concentration in serum or plasma is normally very low, which makes sequencing library construction challenging. Here, we describe an optimized library construction protocol that combines high-efficiency adaptor ligation and unbiased library amplification to deliver a sequencing library with high yield and complexity from as little as 1 ng cfDNA. The sequencing library can be used in combination with target enrichment for sensitive and reliable mutation detection in cfDNA samples.

2714T

Low cost, broad panel (100kb) liquid biopsy with reduced DNA sequencing by depletion of wild-type sequence. *A. Marziali[1,2], J. Pel[1], M. Despotovic[1], P. Davies[1], L. Gelinas[1], D. Broemeling[1].* 1) Boreal Genomics, Vancouver, BC, Canada; 2) University of British Columbia, Vancouver, BC, Canada.

The most significant barrier to widespread clinical deployment of a highly sensitive circulating tumor DNA (ctDNA) assay (liquid biopsy) is the high cost of the assay compared to potential reimbursement. Assay cost is currently dominated by the large amount of DNA sequencing required to achieve coverage of a broad gene panel, and by the high read depth required for high clinical sensitivity. The common practice of attaching unique molecular barcodes to ctDNA fragments for the purpose of error reduction further increases the sequencing requirements, making liquid biopsy commercialization in many clinical applications impractical. Since the majority of sequencing reads are wasted re-sequencing wild-type cell free DNA (cfDNA) fragments, a powerful solution to this challenge is to somehow separate mutant cfDNA from wild-type cfDNA so that only relevant tumor molecules are sequenced. Given that in many clinical cases ctDNA represents only 0. 1% – 0. 01% of the total cfDNA, this approach has the potential to reduce required sequencing by 1,000 – 10,000 fold. While we have previously demonstrated such separation for specific mutations, the larger challenge is to achieve this separation and enrichment of ctDNA without any prior knowledge of the mutation site or sequence, and over a broad set of relevant genes. We present a new version of our electrophoretic enrichment technology that can deplete a sample of wild-type genomic DNA prior to sequencing without *a priori* knowledge of the mutation sequence or location, and over large regions of the genome, in the range of 100kb or more. By narrowing the sample to contain primarily mutant ctDNA, both the sequencing cost and false positive sequencing errors are greatly reduced. We present a methodology used to deplete wild-type sequence over large genomic regions, as well as data from clinical samples. The addition of wild-type DNA depletion to NGS assays enables lower cost, high sensitivity, and high specificity liquid biopsy tests to be developed, enabling commercialization in applications with limited reimbursement, potentially including early cancer detection.

2715W

Detection of somatic mosaicism in children with suspected DICER 1 syndrome using high sensitivity sequening with molecular tag-containing Haloplex [HS]. *L. de Kock[1,2], Y. C. Wang[3], B. Rivera Polo[1,2], D. Badescu[3], T. Revil[3], E. Weber[4], N. Sabbaghian[2], M. Wu[1,2], C. Sandoval[5], S. M. J. Hopman[6], J. H. M. Merks[6], A. van Hagen[7], D. A. Plager[8], A. Ramasubramanian[8], C. Cher[9], N. Hamel[4], D. Bouron-Dal Soglio[10], J. R. Priest[11], I. Ragoussis[3], W. D. Foulkes[1,2,4,12].* 1) Department of Human Genetics, McGill University, Montréal, Québec H3A 0G4, Canada; 2) Lady Davis Institute, Segal Cancer Centre, Jewish General Hospital, Montréal, Québec H3T 1E2, Canada; 3) McGill University and Genome Quebec Innovation Centre & Department of Human Genetics, McGill University, Montréal, Québec H3A 0G4, Canada; 4) Department of Medical Genetics, Research Institute of the McGill University Health Centre, Montréal, Québec H3H 2R9, Canada; 5) Department of Pediatrics, New York Medical College and Maria Fareri Children's Hospital, Valhalla, New York 10595, USA; 6) Department of Pediatric Oncology, Emma Children's Hospital, Academic Medical Center, 1105 AZ Amsterdam Zuidoost, The Netherlands; 7) Department of Clinical Genetics, VU University Medical Center, 1081 HZ Amsterdam, The Netherlands; 8) Glick Eye Institute, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA; 9) Agilent Technologies; 10) Department of Pathology, CHU-Sainte Justine and University of Montreal, Montréal, Québec H3T 1C4, Canada; 11) Minneapolis, Minnesota 55414, USA; 12) Program in Cancer Genetics, Department of Oncology and Human Genetics, McGill University, Montréal, Québec H2W 1S6, Canada.

DICER1 is a protein critical for the processing of mature microRNAs and germ-line mutations in the *DICER1* gene cause a rare cancer syndrome, the DICER1 Syndrome, or the pleuropulmonary blastoma (PPB) familial tumor and dysplasia syndrome (PPB FTDS). The syndrome is characterized by PPB along with other rare childhood sarcomas and dysplasias. Four children who presented with multiple primary tumors associated with the DICER1 syndrome were each found to carry a specific *DICER1* "hotspot" RNase IIIb mutation in multiple tumor biopsies from different sites. However, the mutations were not detected in the patients' germ-line using conventional technology, leading to the suspicion of somatic mosaicism. We investigated and characterized the suspected mosaic origin of the *DICER1* RNase IIIb mutations by deep sequencing of multiple tissues: Both tumor and normal tissues from the four patients were deep sequenced. An in-house, custom made *DICER1* Fluidigm Access array (which selectively captures all protein coding regions and intron-exon boundaries of *DICER1*) and a novel, commercially unavailable Haloplex array (which captured a 250kb region including *DICER1*) were used prior to next generation sequencing on the Illumina HiSeq at high coverage (>10,000x). Using the molecular tag technology, the relative abundance of the previously-identified mutations was assessed between tumor and non-tumor samples from the respective patients with the aim of confirming or refuting the hypothesis of a mosaic origin. By applying the Haloplex HS to multiple tissues from each patient, we have identified low-frequency mutations (at 0. 24%) and we have been successful in testing somatic tissues. In conclusion, we have confirmed the hypothesis that *DICER1* RNase IIIb mosaicism is the cause of the rare DICER1-associated tumors in these children. The Haloplex HS provided the sensitivity required to generate and analyze deep coverage data and accurately detect rare low-frequency mosaic mutations.

2716T

Reconstructing clonal evolutionary process among copy number variants in tumor. *A. Tai[1], W. Hsieh[1], C. Peng[2].* 1) Institute of Statistics, National Tsing Hua University, Hsinchu, Taiwan, R. O. C; 2) Departments of Resource Center for Clinical Research, Chang Gung Memorial Hospital, Taoyuan, Taiwan, R. O. C.

Clonal evolution of cancer is a complex issue and needs to be reconstructed with the following three important parts. The first part is to clearly measure the variants such as single nucleotide variants and structure variants. Subclones can then be identified according to the proportion of presence of those variants. The final step is to restore the order of their emergence with a tree structure. Recently, many studies investigated the cancer-clone evolution by analyzing the sequencing data. They are mostly built on single nucleotide variants with some correction from the copy number effects. The potential evolutionary sequence of copy number variants has not been touched yet. In this study, we propose a method based on Poisson distribution to integrate the above three issues specifically for copy number variants with sequencing data. This method first determines the copy number states by a Poisson mixture model. With the copy number variants detected, the second step is to predict the evolutionary relationship between each pair of copy number variant locus with a likelihood-based approach. The likelihood function is constructed to distinguish the following three potential relations: being present at the same subclone, being present at subclones descended from each other, or being present at subclones with farther/independent relationship. This strategy can cluster the locus into their emerged subclones and at the same time build the evolutionary relationship among subclones. We will demonstrate the accuracy of the relationship assignment through simulations under different scenarios.

2717W

Successful detection of 40 COSMIC hotspot mutations at allelic frequency below 0. 5%. *A. Mongan, R. Chien, D. Brinza, K. Bramlett, F. Hyland.* Thermo Fisher Scientific, South San Francisco, CA.

Research has shown that circulating cell free DNA (ctDNA) is informative of tumor load and tumor evolution in both solid and hematological cancer. Detecting mutations in these ctDNA holds the promise for an accurate and non-invasive approach to assess minimum residual disease as well as treatment response in the future. However, as ctDNA makes up only a small fraction of cell free DNA recovered from the plasma, traditional methods of targeted sequencing often face a poor signal-to-noise ratio that can not be overcome with deep coverage. Here we present a novel research method that is capable of detecting ultra-rare mutations at allelic frequency below 0. 5%. This approach leverages target multiplexing capabilities of the Ion AmpliSeq™ technology with some important modifications to the sample preparation procedures. The new protocol requires as little as 20 ng of input DNA and offers a sample-to-answer turn-around time under 24 hours. To support the analysis of this new approach, we have further developed a novel Bayesian statistics that models the propagation of potential artifacts introduced during amplification and sampling effects during sequencing to differentiate false positives (variants observed in sequencing data that were not present in input DNA) from true mutations that were present at very low levels in the original sample. We successfully applied this new method to detect spike-in mutant DNA in both cell line (Coriel GM23485) and cfDNA samples. Specifically, we demonstrated the detection of 40 COSMIC genomic aberrations found in frequently mutated genes including EGFR, ALK, PIK3CA, KRAS, NRAS, MET, PTEN, KIT, ATM, RB1, TP53 and others. At hotspot locations of 40,000x coverage, the method achieved 90% sensitivity and 100% specificity with no false positive results. In summary, we are reporting, for the first time, a novel research method that is capable of detecting ultra-rare mutations at allelic frequency below 0. 5% using the Ion Torrent™ platform.

2718T

The transcriptional landscape and mutational profile of follicular thyroid neoplasm. *H. Cho[1,4], SK. Yoo[1,6], HG. Jee[2], SJ. Kim[3], S. Lee[1], JY. Shin[1], YJ. Park[1,5], KE. Lee[1,3], JS. Seo[1,4,6].* 1) Genomic Medicine Institute (GMI), Medical Research Center, Seoul National University, Seoul, Korea; 2) Research Institute, National Medical Center, Seoul, Korea; 3) Department of Surgery, Seoul National University Hospital & College of Medicine, Seoul, Korea; 4) Department of Biomedical Sciences, Seoul National University Graduate School, Seoul, Korea; 5) Department of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea; 6) Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul, Republic of Korea.

Follicular thyroid carcinoma accounts for 10% of the thyroid neoplasm, and similar cytological feature of follicular thyroid carcinoma (FTC) and follicular thyroid adenoma (FA) is huge diagnostic challenge in clinics. We performed RNA sequencing (RNA-seq) for 32 FTC and 19 paired-normal tissues, 25 FA and 18 paired-normal tissues. We then additionally analyzed 82 papillary thyroid carcinoma (PTC), 53 follicular variant papillary thyroid carcinoma (fvPTC). Driver mutations in 83% of follicular thyroid neoplasms (FTN) were discovered, and 41% follicular thyroid neoplasms harbored RAS (KRAS, NRAS, and HRAS) mutations. Furthermore, DICER1 and EIF1AX are second most common drivers, which accounts for 6% of FTN, respectively. Unlike previous reports PAX8-PPARG fusion gene was detected in only one case in FTC (0. 01%) which represents ethnic difference and no other fusion driver was detected. Moreover, chromosome 22q deletion was more recurrent in RAS driven neoplasms (chi-square test p = 2. 701e-41). Principal component analysis of gene expression and three gene expression based on the scoring systems represented all of thyroid neoplasms; they are grouped by BRAF-like, RAS-like regardless of cytological classification. In addition, we discovered some FTNs show up-regulatited TCA cycle and other metabolic pathways. Also they represented oncocytic variant feature (Fisher's exact test p = 6. 214e-08). In the present study, we successfully extended transcriptomic landscape of thyroid neoplasm and figured out dominant role of driving genetic alterations in follicular and papillary thyroid neoplasms.

2719W

Haplotyping and structural variant detection from whole exome sequencing of 1ng of cancer cell line DNA. *K. M. Giorda, M. Schnall-Levin, S. Kyriazopoulou-Panagiotopoulou, G. X. Y. Zheng, P. Marks, P. Mudivarti, H. Ordonez.* 10X Genomics, Pleasanton, CA.

Structural changes, and particularly gene fusions, are known driving mutations in many cancers. In many cases they have also proven to be effective drug targets. However, sample input mass is often limited (<50 ng) and detecting fusions is a challenge with existing short-read sequencing technologies, particularly when using exon target enrichment approaches to achieve the ultra-deep coverage required to sensitively detect important mutations in heterogeneous cancer samples. The new GemCode system was used to obtain high-depth whole exome sequencing data from 1ng of cancer cell line DNA. The GemCode Platform massively partitions and barcodes DNA, producing sequencing-ready libraries with >100,000 specific barcodes. Custom algorithms use this barcode information to map reads back to original long molecules of DNA, creating Linked-Reads that span many 10's to 100s of kilobases. Since GemCode barcodes are incorporated during library construction, targeted whole exome libraries retain long-range information used for haplotype phasing and structural variant detection including gene fusions. Haplotype phasing of GemCode whole exome samples produced N50 phase blocks >125 kb, with more than 95% of genes shorter than 100kb fully phased. Variant calling from duplicate removed high depth sequencing (≥160X) starting from 1ng input achieved sensitivity and precision comparable to ligation-based libraries using 200ng of DNA. Linked-Read information was used to detect structural variants in cancer cell lines. We detected EML4/ALK and ALK/PTPN3 fusions in the lung cancer cell line NCI-H2228, an NPM1/ALK fusion in the lymphoma cell line SU-DHL-1 and SLC26A/PRKAR2A fusion in the triple negative breast cancer cell line HCC38. Furthermore a complex genome rearrangement known as chromoplexy was uncovered in the VCaP prostate cancer cell line. The closed quartet chromosome-shuffling event includes TMPRSS2/ERG fusion that is found in half of prostate cancers. In summary, comprehensive tumor profiling includes phasing and structural variant detection from precious samples. We employed the 10X Genomics platform to generate highly complex libraries for gene phasing and fusion detection while simultaneously following intricate chromosome rearrangements including chromoplexy from just 1ng of DNA.

2720T

The Genomic Landscape of Allelic Imbalance in the Normal-Appearing Airway Field of Cancerization. *H. Kadara[1], Y. A. Jakubek[1], W. Lang[1], S. Vattathil[1], M. Garcia[1], L. Huang[1], W. Lu[1], C. Chow[1], Z. Weber[2], G. E. Davies[1], C. Behrens[1], N. Kalhor[1], C. Moran[1], J. Fujimoto[1], R. J. Mehran[1], J. Fowler[1], E. A. Ehli[2], I. I. Wistuba[1], P. Scheet[1].* 1) MD Anderson Cancer Center, Houston, TX; 2) Avera Inst. For Human Genetics, Sioux Falls, SD.

The phenomenon of field cancerization has been observed in various cancers, including those of the lung. We have recently demonstrated that "normal" airway cells carry expression profiles that are often characteristic of the adjacent tumor. A better understanding of mechanisms driving these field changes may provide important biological insights into lung tumorigenesis. Loss-of-heterozygosity (LOH) and other forms of acquired chromosomal alterations (allelic imbalance; AI) have an established role in oncogenesis. However, the relationship between AI and field cancerization has not been studied comprehensively across the genome. Here we address this void by interrogating a rich collection of normal airways from non-small cell lung cancer (NSCLC) patients. We applied Illumina 1M SNP arrays to characterize whole genome copy number alterations in 435 samples from 45 early-stage NSCLC patients [31 adenocarcinomas (ADCs), 14 squamous cell carcinomas (SCCs)]. Each patient set comprised samples from the primary tumor and adjacent airways paired with blood cells and/or uninvolved normal lung tissue. A subset of these included brushings from large mainstem bronchi and from the nasal cavities as well as multi-region tumor biopsies for intra-tumoral analysis. To characterize the field in airways at a genome-wide scale, we applied a haplotype-based computational program, hapLOH, to profile AI events (loss, gain, copy neutral LOH) in a paired mode contrasting signals in the blood or normal lung. We detected 247 AI events in airways of 21 of 45 patients. Of the 21 patients, 19 had events in the adjacent airway, 3 had events in the large airway, and no events were observed in nasal brushings, indicating a pronounced AI field gradient. We detected AI in the airways of ~30% of ADCs regardless of smoking status (2 of 7 non-smokers, 8 of 24 smokers), and 79% (11 of 14) of SCCs, clearly indicating squamous histology as a greater predictor of genomic field effects ($P < 0.01$). The most frequently observed airway alterations were in 9p and 9q, affecting 13 smoker patients. Finally, we note that AI events were present in the airways of 4/5 (80%) of patients with recurrence and only in 17/40 (43%) patients without recurrence, signaling a prognostic value in studying the field in NSCLC. Although preliminary, our findings suggest that chromosomal aberrations are common in the airway field of cancerization and can provide insights into the biology of lung cancer pathogenesis.

2721W

Towards understanding the genomic architecture of cancer genomes. *E. T. Lam[1], A. Hastie[1], M. Imielinski[2], C. Zhong Zhang[2], J. Wala[2], Z. Džakula[1], H. Cao[1].* 1) BioNano Genomics, Inc. , San Diego, CA; 2) Broad Institute of Harvard and MIT, Center for Biomedical Informatics, Cambridge, MA.

Understanding the genetic architecture of cancer requires whole-genome and integrative approaches. Cancers often feature genomic alterations that range from single-base changes to large-scale structural rearrangements. Having a complete catalogue of mutations in cancer is crucial for identifying key drivers and providing accurate diagnosis, prognosis, and targeted therapy. Next-generation sequencing (NGS) platforms have limited power to decipher large, complex rearrangements frequently observed in cancer. Genome mapping represents a complementary technology that provides critical long-range structural information. It involves high-throughput analysis of single molecules spanning hundreds of kilobases in NanoChannels. Long-range information is preserved and direct interrogation of complex structural variants made possible. Therefore, leveraging the strengths of these complementary platforms would give a comprehensive view of a cancer genome. Here, we present our analysis of well-studied cell lines as well as fresh patient tumor samples (multiple myeloma and prostate cancer). We constructed complete *de novo* genome map assemblies with N50 lengths of more than 1 Mb. We derived multi-sample normalized copy number profiles of matched tumor-control pairs based on genome mapping data. We observed that tumor samples had highly variable copy number profiles, corresponding to focal and chromosome-scale changes. Copy number breakpoints were shown indicative of translocation events. We also present a pipeline to integrate NGS and genome mapping data to validate and refine translocation calls. Genome mapping data helped bridge and phase neighboring translocation events. Finally, we present a computational approach to identify translocations by clustering single molecules with abnormal alignment to the reference and by performing local assemblies of these molecules. Overall, integrating NGS and genome mapping data provides a comprehensive view of a cancer genome.

2722T

Direct Structural Variation Analysis of FFPE Samples Using Long Mate Pair NGS Libraries. *D. Mead[1], D. Smith[2], S. Monsma[1], M. Lodes[1].* 1) Lucigen, Middleton, WI; 2) Mayo Clinic, Rochester, MN.

The vast majority of cancer biopsy samples are routinely formalin fixed and embedded in paraffin (FFPE) in order to preserve the morphological features of suspected tumor samples. While FFPE samples are valuable for archiving histological data, the formalin treatment is detrimental to nucleic acids, resulting in fragmentation, oxidation, deamination, and protein-DNA crosslinks. Although methods for PCR amplification and exome capture from FFPE samples have been successfully developed and validated, little long-span information is available to aid the study of structural variation (SV) in cancer samples. SV analysis would benefit greatly from construction of mate pair libraries from FFPE samples, but such studies have not been described, and this capability has been declared unachievable. In spite of these challenges, we have developed methods to construct mate pair libraries from genomic DNA extracted from FFPE samples. The purpose of this study is to measure the efficiency of coverage and ability to uncover SV from FFPE samples using a new NGS tool for direct mate pair library construction and analysis. Results from normal and cancer samples indicate feasibility of 2-3 kb mate pair libraries, for example a library constructed from FFPE prostate tissue exhibited a mean mate pair distance of 1,690 +/- 720 bp, and a library constructed from FFPE kidney exhibited a mate pair distance of 2,068 +/- 840 bp. The ability to query these samples for structural variations could unleash a new era in cancer diagnostics and enable detection of.

2723W

Genomic alterations profile in triple negative breast tumors with loss of BRCA1 nuclear expression. *T. Tapia[1], A. Aravena[2], C. Alvarez[1], V. Cornejo[4], W. Fernández[4], M. Camus[5], A. Maass[2,3], P. Carvallo[1].* 1) Departamento de Biologia Celular y Molecular, Pontificia Universidad Catolica de Chile, Santiago, Santiago, Chile; 2) Centro de Modelamiento Matemático, Universidad de Chile, Santiago, Chile; 3) Departamento de Ingeniería, Universidad de Chile, Santiago, Chile; 4) Unidad de Anatomía Patológica, Hospital San Borja Arriarán, Santiago, Chile; 5) Centro de Cáncer, Pontificia Universidad Católica de Chile, Santiago, Chile.

Triple-negative breast cancer (TNBC), defined by the absence of expression of estrogen and progesterone receptors and epidermal growth factor receptor 2 (HER2), represent an aggressive group of tumors lacking directed therapy. TNBC tumors are frequently observed in women carrying a germline mutation in the *BRCA1* and in approximately 30% of non-carrier women with loss of expression of BRCA1 in tumors, suggesting that loss of function of BRCA1 could be relevant in TNBC tumor development and/or progression. A knowledge of the genes involved in this type of cancer may have relevance for treatment. Also, identifying regions affected by copy number alterations may help to detect genes in molecular or signaling pathways that could be potential treatment targets. We studied 48 TNBC by immunohistochemistry for cytokeratins 5 and 14, EGFR and BRCA1. Using array-CGH we characterized common gains and losses between tumors. Analysis of BRCA1 by immunohistochemistry revealed different patterns of expression and localization of this protein in TNBC tumors. We identified 31% of tumors with absent or reduced nuclear expression of BRCA1. This group was mainly comprised of TNBC/Basal-like tumors. Considering all TNBC tumors, array-CGH analysis revealed common gains (5q15. 33, 20q13. 33, 21q22. 3) and losses (6p11. 2, 9q33. 1, 11q22. 3). Clustering of TNBC tumors separated two main groups, one of which correlates with the mislocalization of BRCA1. Moreover, we found copy number alterations significantly different among subgroups of tumors with different expression/localization of BRCA1. These results suggest that loss of BRCA1 expression or its mislocalization is relevant for tumor progression. Analysis of the genes contained in these regions revealed specific biological processes and signaling pathways. Specifically, in TNBC/Basal-like tumors with absent/low nuclear expression of BRCA1 we identified genes related to epithelial cell differentiation and RAS protein transduction, suggesting that this signaling pathways may be relevant in the progression of BRCA1 deficient tumors. Our results suggest that TNBC/Basal-like tumors with loss of BRCA1 nuclear expression may be selected for personalized treatment with PARP inhibitors and DNA-damaging agents such as platinum. FONDECYT 1080595, CONICYT 24091058.

2724T

Identifying somatic copy number alterations for cervical cancer in the Latino population. *C. Xu[1], A Joseph[1], J Ordonez[2], L Palmer[3], C Camarillo[2], A Torabi[5].* 1) Department of Pediatrics, Texas Tech University Health Sciences Center, El Paso, TX; 2) Departments of Biomedical Sciences, Paul L. Foster School of Medicine, Texas Tech University Health Science Center, El Paso, TX, USA; 3) University of Texas at El Paso, TX, USA; 4) Departments of Medical Education, Paul L. Foster School of Medicine, Texas Tech University Health Science Center, El Paso, TX, USA; 5) Departments of Pathology, Paul L. Foster School of Medicine, Texas Tech University Health Science Center, El Paso, TX, USA.

Cervical cancer affects millions of Americans. The rate for the Latino population is approximately twice that for non-Latinas. However, the etiologies of cervical cancer are still not fully understood. Recent advances in genome studies have led to the discovery of one important type of variation, copy number alterations (CNAs), with high throughput technology. A number of somatic mutations, including TP53, PIK3CA, PTEN, STK11 and KRAS and several CNAs in the pathogenesis of cervical carcinomas were suggested in the non-Latino populations. Therefore, to identify somatic mutation events in the Latino population, we conducted a pilot study of genome wide copy number alteration analysis using 2. 5 million markers in four diagnostic groups: healthy normal tissue (N=2), low grade dysplasia (N=4), high grade dysplasia (N=5), and invasive carcinoma (SCC, N=5). The source of the DNA was derived from these tissues and CNA was detected with used Illumina HumanOmni2. 5-8 BeadChip Kit follow by statistical and bioinformatics analyses using the PennCNV and cnvPetition since the use of multiple algorithms has been shown to increase the reliability of observations with different degrees of confidence. A total of 83 CANs (8. 4% insertions and 91. 5% deletions) over 100 kb was identified in only low grade dysplasia, high grade dysplasia and SCC. Recurrent CNAs included 1q21. 2 (2 SCC, 592 – 330 kb deletions), 3p21. 31 (3 high grade dysplasia, 3 SCC where RBM6, tumor suppressor gene located, 219-3379 kb deletions), 17q21. 2 (2 low grade dysplasia, >1215 kb deletions, where BRCA1, tumor suppressor gene located), 19p13. 2 (high grade dysplasia, SCC, >650 kb deletions), and 9q34. 12 (high grade dysplasia and SCC, ~ 242 kb deletion and insertion, where ABL1, oncogene located). Moreover, difference sizes of deletions were also discovered in 16p13. 3 (5 SCC and 4 high grade dysplasia) close to the location of CREBBP, tumor suppressor gene. This is first report of CNAs identified as known (tumor suppressor or oncogene) and as novel loci for cervical cancer in the Latino population. We are aware of the major limitation of the study, a small sample size. Thus, currently we are confirming the findings with more sample and analyzing our data using public available data of normal tissue (N=40) as references. However, additional studies with a large sample are needed to confirm the current findings.

2725W

Racial differences in molecular cytogenetic abnormalities in consecutive black and white patients with multiple myeloma. *Y. S. Zou[1, 4], Y. Huang[2], Z. Feng[2], S. Chan[3], I. Lee[4], Z. Singh[4], M. R. Baer[5], A. Z. Badros[5].* 1) Clinical Cytogenetics Lab, University of Maryland School of Medicine, Baltimore, MD; 2) Department of Mathematics and Statistics, University of Maryland, Baltimore County, MD; 3) Department of Human Genetics and Genomic Medicine; 4) Department of Pathology; 5) Department of Medicine and Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, MD.

Background:Multiple myeloma (MM) is a clonal plasma cell malignancy with disparities in prevalence and incidence betweenblacks andwhites. Although a few studies have reported differences in molecular cytogenetic abnormalities betweenblacks and whiteswith MM, no study has been performed inconsecutiveblack and white patients with MM, which can avoid the bias from differences due to chromosome culture conditions, fluorescence*in situ* hybridization (FISH) methods/probes used in different centers, etc. We sought to define the prevalence of cytogenomic alterations in consecutive patients with MM using integration of chromosome and interphase FISH in CD138-positive cells. **Methods**:-Consecutive patients seen at the University of Maryland Greenebaum Cancer Centerwere diagnosed with MM by clinical and hematopathology criteria. Routine chromosome analyses andFISH were performed. CD138 cells were isolated from bone marrow samples and analyzed by FISH for eight cytogenetic subtypes (monosomy 13/del(13q), monosomy 17/del(17p), gain of 1q, t(11;14), t(4;14), t(14;16), IGH rearrangements, and other abnormalities by FISH). IGH rearrangements included t(11;14), t(4;14), t(14;16) and other IGH rearrangements. **Results**:From July 2013 to January 2015, 180 consecutive patients with MM were studied, of whom 85 were the black (mean age = 58. 3) and 95 were the white (mean age = 62. 2). White patients were older than black patients (p=0. 007) and included a higher proportion of males (p=0. 01). With regard to racialdifferencesin molecular cytogenetic abnormalities, black patients had significantly lower frequencies of -17/17p- (7. 1% versus 20. 4%; p=0. 01) and -13/13q- (28. 6% versus 43. 0%; p=0. 047) than whites. Black patients also had a trend toward lower frequencies of IGH rearrangements, t(14;16) and abnormal karyotypes than white patients (0. 05<p<0. 10). Within age groups of < 60 and ≥ 60 years, black patients had significantly lower frequencies of the -17/17p- abnormality than whites (p=0. 0037), and had lower frequencies in abnormalities of -13/13q- and IGH than white patients (0. 05<p<0. 06). Within race groups, younger patients showed significantly higher frequencies of the -17/17p- abnormality than older patients (p=0. 02). **Conclusions**:To the best of ourknowledge, this is the first molecular cytogenetic study of consecutive black and white patients with MM. This study showed significant differences in cytogenomic profiles between black and white patients with MM.

2726T

Genetic Hotspots of Glioma Progression by Array CGH: A Systematic Review. *C. J. DeGraffenreid, T. Dickyi, B. Vo, S. Romee, J. Gu, V. Hopwood, M. Zhao.* Cytogenetics, University of Texas MD Anderson Cancer Center School of Health Professions, Houston, TX.

Glioma is a common form of primary brain tumor that arises from the three principle types of glial cells and is categorized into low grade (grade I and II) and high grade (grade III and IV). The use of whole genome array and sequencing technology has revealed that chromosomal copy number variations (CNVs) are commonly seen in glioma, and increase in frequency with tumor progression. Survival is very poor once glioma has progressed to high-grade form, and therefore it is important to establish which copy number aberrations harbor genes that lead to glioma tumor progression. The purpose of this study was to determine which chromosome segments show significantly increased chromosomal copy number variations in the progression from low to high-grade glioma. We hypothesized that chromosomal regions that contain cell cycle regulator genes would show significantly increased aberrations between low and high grade glioma tumors, including chromosome 7p (*EGFR* gene), 9p (*p16* gene), 10q (*PTEN* gene), and 17p (*p53* gene). A systematic review was performed by searching PubMed, Scopus, and Cinahl using keywords glioma, array CGH, sequencing, and whole genome. Data involving CNV in glioma found using array and sequencing technology of the whole genome was extracted. Our result showed significant increases in copy number variation between low and high-grade glioma. Chromosomes showing an increase in aberration gain frequency include 1q (39% increase), 7p(17%), 9p(12%), and 10q(15%) and those showing an increase in aberration loss frequency include 9p(30%), 10q(18%), 17p(26%), 17q(13%), and 22q(10%). Increased gains and losses of chromosome segments in high-grade glioma was found to be consistent with our hypothesis, as those regions where cell cycle regulator genes are located were increasingly gained or lost in high grade glioma as compared to low grade. Other regions found to show increased aberration possibly contain other mutated genes involved in tumor grade advancement. More research is needed to identify the specific genes on the regions showing aberration that could be affecting tumor growth and progression.

2727W

Hepatitis C Virus and Schistosomiasis as A Causative Factor for hTERT Amplification in Hepatocellular Carcinoma. *O. M. Eid, N. A. Helmy, W. Ezzat, I. Fadel, M. M. Eid.* National Research Centre - Egypt, Giza, Egypt.

Background Telomere abnormalities appear to play a role in carcinogenesis. hTERT is the catalytic component of the telomerase complex. It is regulates telomerase activity in cancers such as hepatocellular carcinoma (HCC). Amplification of the hTERT gene have been detected in about 21% of HCC. HCC occurs frequently in patients with chronic viral hepatitis or liver cirrhosis. Hepatitis B virus (HBV) or hepatitis C virus (HCV) induce HCC. The schistosomal parasite plays a role in the development of HCC through modify the course of hepatitis C accelerating progression to hepatitis C-associated fibrosis and thus quicker progression to HCC. **Objectives and methods**Evaluation of hTERT amplification in HCC in association with HCV and schistosomiasis. Fifteen surgically resected HCC (Grade two) and fourteen cirrhotic resected liver were examined using FISH technique. **Results** Amplification of hTERT gene were found in 80% (21/15) of HCC which is much higher than the previously reported data. Eight samples were HCV and schistosomiasis positive, 4 were HCV positive, 2 were schistosomiasis positive and one was HCV and schistosomiasis negative. Amplification of the hTERT gene was not observed in cirrhotic tissues except two cases which is probably due to extension of adjacent HCC. **Conclusions**From our results, the coexistence of HCV and/or schistosomiasis with hTERT amplification, we can conclude that the HCV and schistosomiasis may play an important role in upregulation of this gene. Consequently, this may lead to the development of HCC rather than just a marker for cancer development. However this conclusion might need further investigation.

2728T

Translocation t(5;16)(q32;p13) and trisomy 8 in a patient with Acute Eosinophilic Myelomonocytic Leukemia. *X. Montenegro[1], R. Céspedes[2], Y. Llimpe[1,3], A. Arias[1].* 1) Equipo Funcional de Genética y Biología Molecular, Instituto Nacional de Enfermedades Neoplásicas, Lima, Perú; 2) Universidad Nacional Federico Villarreal (UNFV), Facultad de Ciencias Naturales y Matemáticas. Lima-Perú; 3) Universidad Nacional Mayor de San Marcos (UNMSM), Facultad de Medicina. Lima-Perú.

Acute eosinophilic myelomonocytic leukemia (AML-M4Eo) is considered an entity with unique cytogenetic, morphologic and prognostic qualities. AML-M4Eo presents at a young median age, with organomegaly and high peripheral white blood cell count, and usually has a good response to chemoterapy although relapses in central nervous system are common. Bone marrow morphology is characterized by blast cells, monocytic cells and eosinophills showing positivity for chloroacetate esterase (CAE) and peroxidase. AML-M4Eo is frequently associated with inv(16)(p13q22) or the variant t(16;16) which involves the genes: *CBFB* at 16q22, that codes for beta subunit of a heterodimeric core-binding transcription factor belonging to the PEBP2/CBF transcription factor family (CBFb) and the gene *MYH11* at 16p13 that codes for smooth muscle myosin belonging to the myosin heavy chain family (SMMHC), generating a chimeric protein CBFβ-SMMHC, that blocks myeloid cell differentiation. We report the case of a one year old female patient, that presented with pallor skin, fever, hepatosplenomegaly and lymphadenopathy. A Peripheral blood study showed hemoglobin 10,1 g/dL, leukocytes 15,7x109/L, lymphocytes 20%, monocytes 17%, neutrophils 13%, blasts 50% and platelets 48x109/L. Bone marrow aspirate was hypercellular and heterogeneous with 23% of blasts, Neutrophils and their precursors, monocytes and their precursors each comprise more than 20% of bone marrow cells. The eosinophils count was elevated (10%) and showed prominent basophilic-staining granules. Bone marrow cytochemical studies revealed CAE and peroxidase positive. Flow cytometry demonstrated mature monocytes positive for CD14, CD64, CD4; polyclonal lymphocytes T CD/CD8 ratio of 1,18 and inmature non granular myeloid population (9,39%) positive for CD45, CD34, CD13, CD33, CD15, HLA, DR, CD117, MPO, CD71, CD19. Bone marrow cytogenetics revealed 47,XX,t(5;16)(q32;p13),+8. She was diagnosed with AML-M4Eo and received chemotherapy treatment but died ten months after diagnosis. The translocation t(5;16)(q32;p13) is not very common. It affects the genes *PDGFRB* on chromosome 5 and *NDE1* at chromosome 16. It is important to note that the genes *MYH11* and *NDE1* overlap and are transcribed in opposite ways, therefore *NDE1* is altered in more than 90% of cases with inv(16)(p13q22). In our case, t(5;16) is present with trisomy 8 in the same clone, even though t(5;16)(q32;p13) may be a variant from t(16;16), the prognosis is uncertain.

2729W

Sensitive cancer fusion detection and discovery in FFPE samples by RNA-Seq. *L. C. Watson, S. M. Gross, A. M. Mai, F. Schlesinger, I. Khrebtukova, G. DeSantis, G. P. Schroth.* 5200 Illumina Way, San Diego, CA 92122.

Gene fusion detection in cancer samples can provide tumor-specific information for cancer research, clinical diagnosis and targeted treatment. Common fusion detection methods such as qPCR and FISH are restricted to known fusion junctions and limited in the number of genes that can be detected in parallel. In contrast, RNA sequencing is a powerful approach for simultaneous discovery of all possible fusion junctions in a single reaction. But, the sequencing depth required for sensitive detection of fusions from whole-transcriptome libraries can be cost-prohibitive. Here we describe a cancer-specific capture-based approach for fusion detection by RNA sequencing that requires only a fraction of the sequencing depth of whole-transcriptome methods. We designed oligo probes that densely target coding regions of over 200 clinically relevant gene fusions and cancer-associated genes. This oligo panel was used to capture cancer-specific fusions from total RNA-Seq libraries. We used commonly studied cancer cell lines including MCF-7, K562, PC-3, LnCAP, A431 and Universal Human Reference RNA (UHRR) to compare the sensitivity of fusion detection across three RNA-Seq library prep methods: (1) cancer panel library capture (2) whole-transcriptome library capture and (3) PolyA selection. We show that probes targeting individual exons can robustly capture well-characterized cancer gene fusions such as *BCR-ABL* and *BCAS4-BCAS3,* as well as translocations where fusion junctions are unknown. Furthermore, these comparisons demonstrate the enhanced sequencing efficiency of the targeted cancer panel, while maintaining highly accurate quantitation of gene expression. We show that selective enrichment of RNA-Seq libraries with cancer-specific capture probes enables high-resolution mapping of genomic rearrangements in patient cancer samples, even those derived from FFPE, facilitating sequencing studies that were not previously possible.

2730T

A complex variant of t(7;12) with loss of RB1 locus in a case of Acute Erythroblastic Leukemia. *A. Yenamandra, A. Hollis, T. Leftwich, D. Head.* Pathology, Microbiology and Immunology, Vanderbilt University Medical Center, Nashville, TN.

Recurrent and cryptic rearrangements of t(7;12) involving the ETV6 locus at chromosome 12p13 have been identified in children of age two years or younger with various types of leukemia. Heterogeneityin the breakpoints of chromosome 7 ranging from 7q32 to 7q36 have been reported in myeloid disorders. We report here a case of Acute Erythroblastic Leukemia with a variant translocation involving three chromosomes, t(5;12;7), with additional loss of the RB1 locus on 13q. Sequential cytogenetics and metaphase FISH were performed to identify the partner chromosomes with the ETV6 locus rearrangement. A two- year old male with fatigue, bruising, petechiae and fever presented for evaluation of normocytic anemia and thrombocytopenia in February, 2014. His initial bone marrow (BM) biopsy was packed, with a predominance of erythroids and a large percentage of immature forms. Flow cytometry was positive for CD33, CD7, CD36 and CD71, consistent with erythroid differentiation of immature cells. The diagnosis was FLT3-, NPM1-, and KIT17 D816V- AML. Cytogenetics revealed seventeen of the twenty cells analyzed had 46,XY,t(5;12;7)(p15. 1;p13;q32),del(13)(q12q14)[17]/46,XY[3], with deletion at 5p15. 1 and 12p13 of der(5) and der(12) respectively, along with a deletion of 13q12q14 involving the RB1 locus. The patient entered remission with induction chemotherapy, received consolidation chemotherapy, and at most recent follow up 16 months after initial diagnosis bone marrow studies were negative for leukemia by morphology and cytogenetics. The variant translocation described here has not been previously reported. The putative disrupted gene in this translocation at 7q32 is unknown, but this breakpoint appears to have a better prognosis than translocations involving the 7q36 locus, as most patients with the 7q36 breakpoint have had a poor outcome. In summary, we present a case with a novel rearrangement involving the important hematopoietic regulatory gene ETV6 in a case of pediatric primary AML. References:

1. t(7;12)(q36;p13) and t(7;12)(q32;p13)translocations involving ETV6 in children 18 months of age or younger with myeloid disorders. Slater RM1, von Drunen E, Kroes WG, Weghuis DO, van den Berg E et al. Leukemia. 2001 Jun;15(6):915-20. 2. Three-way complex translocations in infant acute myeloid leukemia with t(7;12)(q36;p13): the incidence and correlation of a HLXB9 overexpression. Park J, Kim M, Lim J, Kim Y, Han K et al. , Cancer Genet Cytogenet. 2009 Jun;191(2):102-105.

2731W

Transformation of 5q- Syndrome to Chronic Myelogenous Leukemia with a Novel Complex BCR/ABL1 Translocation with Rapid Transformation to Acute Myelogenous Leukemia: An Update. *A. Zaslav[1], B. Burke[1], M. Schuster[2], S. Itwaru[2], B. Jalilizeinali[2], T. Mercado[1], E. Knorr[1], D. Tully[1], T. Ahmed[1].* 1) Pathology - Cytogenetics Lab, Stony Brook Medicine - SUNY Stony Brook, Stony Brook, NY; 2) Medicine, Stony Brook Medicine-SUNY Stony Brook, Stony Brook, N. Y.

We report a case of a 72 year old female who initially presented in August, 2010 with Myelodysplastic Syndrome (MDS). G-banded chromosome analysis of bone marrow (BM) revealed a deletion of part of the long arm of chromosome 5 (i. e. del(5)(q12q33), in 11/20 metaphase cells confirmed by FISH. A diagnosis of MDS with isolated 5q deletion was made. In June 2014 she was found to have increased blasts on BM. Cytogenetic and FISH analysis of BM revealed the del(5q) in 1. 5% of nuclei, and a complex BCR/ABL1 translocation [i. e. , 45,XX,t(9;15;22)(q34;p10;q11. 2),-22] in 20/20 metaphase cells and in 10/10 metaphase FISH. In July 2014 G-banding and FISH evaluation of unstimulated blood (UB) revealed the presence of blasts, the complex translocation and clonal evolution. An additional der(15)t(9;15;22)(q34;p10;q11. 2) in 5/20 metaphases was observed. The del(5) was not present in this analysis. In November of 2014 BM the revealed the same clones as July 2014 and the del(5) in metaphase (5%) and FISH (0. 5%) analyses. In April 2015 UB nuclear FISH revealed the del(5) (0. 5%) and a variant BCR/ABL1 translocation in 3% and three copies of ABL1 in 16% of the nuclei. All FISH studies were performed using the LSI5q EGR1/D5S23 (5q31,5p15. 2) and the LSI BCR/ABL DC DF (9q34,22q11. 2) (Abbott, Des Plaines, IL) probes on nuclei and/or metaphase spreads. Molecular testing demonstrated a p210 BCR/ABL1 transcript. Flow cytometry revealed 18% blasts. These results indicated that the patients' MDS transformed to chronic myelogenous leukemia (CML) with rapid transformation into acute myelogenous leukemia (AML). The patient was placed on the tyrosine kinase inhibitor Tasigna and was in stable condition. However, as clonal evolution occurred, she is being readied for transplant. Transformation from MDS to CML has rarely been reported. To our knowledge transformation with this complex translocation has never been described. This case illustrated the importance of careful monitoring to definitively determine the nature of this disease and course of patient treatment.

2732T

Multi-omics analysis defines core genomic alterations in pheochromocytomas and paragangliomas. *N. Burnichon[1,2,3], LJ. Castro-Vega[1,2], E. Letouze[4], A. Buffet[1,2,3], PF. Plouin[1,2,5,8], J. Bertherat[2,6,7,8], L. Amar[1,2,5], A. De Reynies[4], J. Favier[1,2], AP. Gimenez-Roqueplo[1,2,3,8].* 1) INSERM, UMR970, Paris-Cardiovascular Research Center, PARIS, France; 2) Paris Descartes University, Sorbonne Paris Cité, Faculté de Médecine, PARIS, France; 3) Department of Genetics, Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges Pompidou, PARIS, France; 4) Programme Cartes d'Identité des Tumeurs, Ligue Nationale Contre Le Cancer, PARIS, France; 5) Hypertension Unit, Assistance Publique-Hôpitaux de Paris, Hôpital Européen Georges Pompidou, PARIS, France; 6) INSERM, U1016, Institut Cochin, PARIS, France; 7) Department of Endocrinology, Assistance Publique-Hôpitaux de Paris, Hôpital Cochin, PARIS, France; 8) Rare Adrenal Cancer Network COMETE, PARIS, France.

Pheochromocytomas and paragangliomas (PCCs/PGLs) are rare neural crest tumors derived from chromaffin cells in adrenal medulla or extra-adrenal sympathetic paraganglia, respectively. PCCs/PGLs are characterized by a remarkable genetic heterogeneity with at least 12 identified susceptibility genes comprising two oncogenes (*RET* and *HIF2A*) and ten tumor suppressor genes (*NF1, VHL, SDHA, SDHB, SDHC, SDHD, SDHAF2, FH, TMEM127* and *MAX*). PCCs/PGLs present a high degree of heritability with up to 40% of affected patients carrying a germline mutation in one of these genes. Moreover, loss-of-function mutations in *NF1* and *VHL*, as well as activating mutations in *RET, HIF2A* and *HRAS,* have been reported at the somatic level in about 30% of these tumors. Here we report the first integrated genomic examination of a collection of 202 PCC/PGL. Whole-exome sequencing, single-nucleotide polymorphism (SNP) array, as well as mRNA, miRNA and DNA methylation profiling were performed aiming to characterize the major genomic alterations underlying PCCs/PGLs. SNP array analysis reveals distinct copy-number patterns associated with genetic background. Whole-exome sequencing shows a low mutation rate of 0. 3 mutations per megabase, with few recurrent somatic mutations in common cancer genes not previously specifically associated with PCC/PGL including *TP53*, *CDKN2A* and *MET*. DNA methylation arrays and miRNA sequencing identify DNA methylation changes and miRNA expression clusters strongly associated with messenger RNA expression profiling. Overexpression of the miRNA cluster 182/96/183 is specific in *SDHB*-mutated tumors and induces malignant traits, whereas silencing of the imprinted DLK1-MEG3 miRNA cluster appears as a potential driver in a subgroup of sporadic tumors. Altogether, the complete genomic landscape of PCC/PGL is mainly driven by distinct germline and/or somatic mutations in susceptibility genes and reveals different molecular entities, characterized by a set of unique genomic alterations. In the near future, omics-based tests should be developed to offer access to a precise molecular classification of PCCs/PGLs. The knowledge of specific genomic alterations should guide the choice of targeted therapy for malignant cases. Therefore, our findings pave the way towards omics-based clinical management and personalized medicine for patients with PCCs/PGLs.

2733W

A specific "leukemia diagnostic panels" based on selection of novel relevant biomarkers for leukemiadiagnosis, prognosis and therapeutic decision making. *M. Delledonne[1], G. Martinelli[2], R. Calogero[10], M. Fortuna[3], L. Bullinger[4], P. Vandenberghe[5], L. Farinelli[6], S. Pospisilova[7], J. Hernandez Rivas[8], W. Kern[9], E. Zago[11], L. Xumerle[11], I. Iacobucci[2], A. Ferrarini[1], M. Garonzi[1], L. Bettinetti[1].* 1) Di Biotenologie, Universita Degli Studi Di Verona, Verona, Italy; 2) Alma Mater Studiorum-Universita Di Bologna, Italy; 3) Sinaptica IT SRL, Italy; 4) Universitaet Ulm, Germany; 5) Katholieke Universiteit Leuven, Belgium; 6) Fasteris SA, Switzerland; 7) Masarykova univerzita, Czech Republic; 8) Fundacion De Investigacion Del Cancer De La Universidad De Salamanca, Spain; 9) MLLMunchner Leukamielabor Gmbh, Germany; 10) Universita Degli Studi Di Torino, Italy; 11) Personal Genomics Srl.

Leukemia is an hematological malignancy characterized by an high etiologic, risk and response heterogeneity among subsets of patients. To improve the diagnosis and prognosis it's thus crucial to develop tools able to identify and discriminate different types of leukemia. Within the project "Next Generation Sequencing platform for targeted Personalized Therapy of Leukemia" (NGS-PTL; FP7 european grant) we are involved in the setup of a bioinformatic pipeline for the identification of biomarkers for different types of leukemia from the analysis of Next Generation Sequencing (NGS) data. The pipeline setup is based on the identification of somatic mutations in control-tumor pairs coupled with statistical and functional analysis aimed to identify genes carrying potential driver mutations. Somatic mutations identification step was setup with a combined approach that integrates GATK software for calling of Indels and MuTect for calling of SNPs which allowed to achieve a 89% validation rate. Different strategies can be employed to detect putative genes carrying driver mutations based for example on the not-random distribution of driver mutations in the genome or on the predicted effect of mutations. To maximize the sensitivity of the detection of drivers to be used as candidate leukemia biomarkers we choose to employ a combined approach of three complementary statistical methods. The first method identifies genes significantly enriched in somatic mutations (MutSigCV), the second is based on mutation clustering on protein domains (OncodriveClust) while the third identifies genes accumulating putative driver mutations with high functional impact (OncodriveFM). Finally candidate driver genes detected by one or more methods are mapped on a functional network to identify driver genes perturbing the same metabolic pathway or functional module. Using the method setup we analyzed the sequencing data generated from cohorts of patients affected by 2 different leukemia types (AML,ALL) detecting 41 putative biomarkers which will be used as a base to build diagnostic kits. Using the method setup we analyzed the sequencing data generated from cohorts of patients affected by 4 different leukemia types (AML,ALL,CLL,ET) detecting 47 putative biomarkers which will be used as a base to build diagnostic kits.

2734T

Identification of Two Families with Li Fraumeni Syndrome on Multi-Gene Panel Testing. *C. Csuy[1], JM. McDonald[2].* 1) Novant Health Presbyterian Medical Center, Charlotte, NC; 2) Novant Health Gynecologic Oncology.

Li Fraumeni Syndrome (LFS) was thought to be a rare disorder affecting approximately 1 in 20,000 individuals. LFS has typically been defined to include very young onset breast cancer, adrenocortical tumors, sarcomas, leukemias and lymphomas. It is thought that 50% of patients have cancer by the age of 30 and 90% of patients will develop cancer by the age of 60. In the past year, two patients presented to our clinic that were not classic presentations for LFS however were found to have *TP53* mutations. Both patients were identified utilizing pan-cancer panel testing. The first patient presented because of a personal history of bilateral breast cancer at the age of 46. She documented that her sister was diagnosed with colon cancer at the age of 21. Her father was diagnosed with colon cancer at the age of 45 and her paternal grandmother was diagnosed with breast cancer at the age of 50 and colon cancer in her eighties. Her mother was diagnosed with bladder cancer at the age of 60. A maternal aunt (a maternal half-sister of her mother) was diagnosed with breast cancer in her forties and a maternal great-aunt was diagnosed with breast cancer under the age of 50. The *TP53* mutation was identified ("c. 97-2_99dup). The other patient was unaffected at the age of 42. Her father passed away from an unknown primary at the age of 49. A paternal aunt was diagnosed with breast cancer at the age of 40 and a paternal uncle was diagnosed with prostate cancer at the age of 50. A paternal first cousin was diagnosed with breast cancer in her forties. Her paternal grandfather was diagnosed with colon cancer and passed away at the age of 49. Her mother was adopted but the patient documented that her maternal grandmother was diagnosed with breast cancer in her fifties. The *TP53* mutation "p. R267W" was identified. Neither of these patients were classic in their presentation for LFS however, their results have drastically changed their medical management and follow up. Panels have changed how genetic testing has been offered in the clinical setting and will continue to help clarify syndromes and the prevalence of these conditions.

2735W

≤Evaluation of *IGK* and *IGL* Molecular Gene Rearrangements According to the BIOMED-2 Protocol for Clinical Diagnosis of Hodgkin Lymphoma≥. *S. Ghorbian[1], I. Jahanzad[2], G. R Javadi[3], E Sakhinia[4].* 1) Department of Molecular Biology, Ahar Branch, Islamic Azad University, Ahar, Ira, Ahar, Iran; 2) Department of Pathology, Imam Khomeini Hospital Complex, Medical Sciences / University of Tehran, Tehran, Iran; 3) Department of Biology, Science and Research Branch, Islamic Azad University, Tehran, Iran; 4) Department of Medical Genetics, School of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran.

Molecular clonality of light chains (*IGK* and *IGL*) gene rearrangements assessed by applying BIOMED-2 protocols in Hodgkin's Lymphoma (HL) cases. We evaluated an protocol in HL cases, which has been suggested previously as an gold standard method for molecular clonality analysis on formalin fixed, paraffin-embedded (FFPE) tissue in Non-Hodgkin Lymphoma (NHL) patients, especially. We recruited 50 consecutive FFPE tissue of HL samples to evaluate *IGK* and *IGL* clonality gene rearrangements using BIOMED-2 method. Our finding revealed a total of 94% (47/50) positive clonality, which consisted of 70% (35/50) for *IGK* and 44% (22/50) for *IGL*. In three of cases, clonality was no detected in any of the immunoglobulin gene segments. Analysis of clonality gene rearrangements in *IGK* and *IGL* genes using BIOMED-2 protocols, could be implemented as a valuable method for increase sensitivity (94%) and accuracy of HL diagnosis similarly to NHL by clonality detection.

2736T

BRCA1/2 Mutation Status Is an Independent Factor of Improved Survival for Advanced Stage Ovarian Cancer. *R. Janavicius[1, 2], V. Rudaitis[1], T. Žvirblis[1], L. Griškevičius[1].* 1) VUH Santariskiu Clinics, Santariskiu st. 2, Vilnius, Lithuania; 2) State Research Institute Innovative Medicine Centre, Vilnius, Lithuania.

The aim of this study was to evaluate BRCA1 and BRCA2 mutation impact on prognosis of advanced-stage (III-IV) ovarian cancer patients after standard treatment. Methods: A total of 521 patients with advanced-stage (primary) epithelial ovarian cancer (EOC) were identified from a clinical database during year 1998-2013 and enrolled in a prospective, single-center study. All cases with available germline DNA (n=297) were screened for BRCA1 and BRCA2 gene mutations using combination of methods (HRM, Sanger/Next Generation Sequencing, MLPA); a total of 155 mutation carriers were identified. Progression-free survival (PFS) and overall survival (OS) was assessed between BRCA1/2 mutation carriers and BRCA1/2 wild-type patients. To reduce survivorship bias, analysis was limited for cases with BRCA testing performed <36 month from the diagnosis. Various clinical risk factors for PFS and OS were assessed by univariate and multivariate Cox regression analysis with stepwise model selection process. Results: Older age (hazard ratio [HR], 1. 032; 95% confidence interval [CI], 1. 010-1. 055; P=0. 0047), nonoptimal cytoreduction (HR, 3. 170; 95% CI, 1. 986-5. 060; P=0. 0001), and BRCA1/2 wild type (HR, 1. 625 [1. 003-2. 632]; P=0. 0486) were significantly associated with shorter PFS in multivariate Cox regression analysis. Nonoptimal cytoreduction (HR, 2. 684; 95% CI, 1. 264-5. 701; P=0. 0102) and BRCA1/2 wild type (HR = 1,612 (95% CI 1,16 - 2,23; P=0. 0002) were statistically significant risk factors for shorter OS. The overall 5-year survival for the hereditary case patients was better than that of the nonhereditary patients, however after that time no survival advantage was apparent. The carriers of Baltic founder BRCA1 c. 4035delA mutation had worse survival than other BRCA1 carriers (p=0,014) till 4 years. Conclusions: Advanced ovarian cancer patients harboring BRCA1/2 mutation treated with debulking surgery and platinum-based adjuvant chemotherapy have a longer PFS and OS not longer than 5-years. Available data suggest that*BRCA1* c. 4035delA mutation may be associated with poor survival.

2737W

Gene expresion analysis of *TNF-α, 5-LOX* and *iNOS* in tumoral and normal adjacent tissue of sporadic colorectal cancer patients. *U. Santana-Bejarano[1,2], L. Bobadilla-Morales[1,2], B. González-Quezada[1,2], V. Maciel Gutierrez[3], M. Centeno-Flores[3], J. Corona-Rivera[1], H. Pimentel-Gutiérrez[1,2], C. Ortega-de la Torre[1,2], A. Corona-Rivera[1,2].* 1) Laboratorio de CitoGenética, Genotoxicidad y Biomonitoreo, Instituto de Genética Humana “Enrique Corona”, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Sierra Mojada 950, Col. Independencia, c. p. 44340. Guadalajara, Jal; 2) Unidad de CitoGenética, Servicio de Hematología y Oncología Pediátrica . Hospital Civil de Guadalajara. “Dr. Juan I. Menchaca”. Salvador de Quevedo y Zubieta No. 750. Guadalajara, Jalisco, MX; 3) Servicio de Colon y Recto del Hospital Civil de Guadalajara “Dr. Juan I. Menchaca”. Salvador de Quevedo y Zubieta No. 750. Guadalajara, Jalisco, MX.

Colorectal cancer (CRC) is one of the most frequent malignant diseases caused by accumulation of genetic changes. *TNF-α, iNOS* and *5-LOX* are complementary genes to critic cancer alterations, their function is associated with the control between tumor cells and their environment. It’s thought that their cooperation is essential to complete tumor development. To analyze the expression of *TNF-α, iNOS* and *5-LOX* genes in tumor tissue we compared to normal adjacent tissue from the same patient with sporadic CRC in different tumor stages. The patients were classified into 2 groups: advanced stages (III and IV, n = 20) and early stages (I and II, n = 7). Tumor and normal adjacent tissue were collected and underwent to RNA extraction. Starting concentration of 1 µg of RNA was used to synthesize cDNA. QRT-PCR was performed for *TNF-α, iNOS* and *5-LOX genes*, the assay were validated using the constitutive genes *GUS, ABL* and *ACTB*. The results were analyzed by comparative CT method (χCT). The use of *GUS* and *ACTB* to validate expression assays in CRC was confirmed and also the use of *ABL* aims to validate similar studies. Genetic overexpression in tumor tissue compared to normal adjacent tissue from the same patient befall in the 3 genes, at the same time gradually increasing gene expression was also observed comparing advanced stages with early stages in all of them. In both groups, overexpression of *TNF-α* was observed in tumor tissue compared to normal adjacent tissue in 100% of cases. According to our results, we suggest that *TNF-α, iNOS* and *5-LOX* genes exhibit overexpression in tumor tissue compared to adjacent normal tissue, it could be due to the importance of their activity in biological processes during the establishment of CRC and that the association of overexpression with increasing tumor stage is due to their increasing actions during CRC development.

2738T

Clinical multiplex genetic testing using a standardized panel approach in a population-based hereditary cancer clinic. *K. A. Schrader[1,2,3], T. R. Docking[1,3], I. Bosdet[1], S. Young[1], T. Tucker[1], S. van den Berg[2], M. He[1], Z. Lohn[1], J. Nuk[1], M. Taylor[1], J. Scott[1], A. Mindlin[1], C. Portigal-Todd[1], C. Cremin[1], M. J. Asrat[1], K. Turner[1], J. Sawkins[1], J. Thompson[1], HCP. clinical team[1], L. Armstrong[2], B. McGillivray[2], R. Moore[1], A. Mungall[1], L. A. Swanson[1], L. P. Yang[1], J. Slind[1], M. Balasundaram[1], J. Schein[1], G. Mitchell[1], A. Karsan[1,2].* 1) Hereditary Cancer Program, Michael Smith Genome Sciences Centre, Department of Pathology & Laboratory Medicine, Department of Molecular Oncology, BC Cancer Agency, Vancouver, BC, Canada; 2) Department of Medical Genetics and Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, Canada; 3) equal contributions.

The BC Cancer Agency's Hereditary Cancer Program (HCP), is the primary provider of cancer geneticsservices to families and individuals across British Columbia (BC) and the Yukon. Services include genetic counselling (GC) and testing and cancer risk management advice in high-risk families. A pilot study at the HCP showed economic feasibility for conducting multiplex genetic testing of a 14-gene panel. The pilot study of 90 patients including 2 positive controls, revealed expected genetic diagnoses in 4 patients (4%), an unexpected diagnosis in 1 patient (1%), and variants of uncertain significance (VUS) in 36 patients (40%). The remaining 49 patients (54%) had no presumed pathogenic variants or VUS. These results highlight the possibility of unexpected diagnoses even in small panel testing and the significant number of VUS. Since October 2014, clinical implementation of panel testing for eligible HCP patients has replaced single gene testing for associated genes. This has expedited diagnoses and has avoided costly iterative single gene testing, although it has presented challenges necessitating alterations in GC. GC for panel testing must now highlight to patients issues related to simultaneously testing multiple genes, most of which are not predicted to contribute to the patient's clinical and/or family history. Genes in this panel were chosen to represent those that confer increased cancer risk for breast, ovarian and colorectal cancers, but can be applied to phenotypically diverse conditions. The most common indication for testing is for hereditary breast and ovarian cancer (85%) followed by Lynch syndrome (14%). The other less common hereditary cancer susceptibilities account for 1% of patients tested. While there are multiple benefits to panel testing, it is clear there are scenarios that result in potential complexities in GC, such as a higher rate of VUS, the identification of variants in which risk is only associated with specific ethnic groups, and the identification of presumed pathogenic variants associated with susceptibility to cancer types not present in the person or family history. The challenge for GC is to present sufficient information within the limited appointment time to ensure informed consent, without overwhelming patients with information unrelated to their presenting history. We will present our experience to date regarding panel testing implementation of affected individuals with a clinical suspicion for hereditary cancer.

2739W

Systematic *BRCA1/2* genetic testing in unselected epithelial ovarian cancer- results from the GTEOC study. *M. Tischkowitz[1,2], I. Plaskocinska[1,2], J. Drummond[2], E. Thompson[2], G. Sagoo[3], B. Newcombe[4], E. Barter[5], P. Ridley[6], R. Ng[6], S. Miller[7], F. Thompson[8], A. Dann[8], V. Licence[8], H. Webb[9], C. Hodgkin[4], L. T. Tan[7], M. Daly[9], S. Ayers[5], B. Rufford[6], C. Parkinson[10], H. Earl[10], T. Duncan[11], P. Pharoah[12], S. Abbs[2], N. Hulbert-Williams[13], R. Crawford[10], J. Brenton[10,14], H. Shipman[1,2].* 1) University of Cambridge, Cambridge, Cambridgeshire, UK; 2) East Anglian Medical Genetics Service, Cambridge University Hospitals NHS Foundation Trust, UK; 3) PHG Foundation, Cambridge, UK; 4) Cambridge Cancer Trials Centre, Cambridge University Hospitals NHS Foundation Trust, UK; 5) Clinical Oncology, Peterborough and Stamford Hospitals NHS Foundation Trust, UK; 6) Cancer Services, The Ipswich Hospital NHS Trust, Ipswich, UK; 7) Clinical Cancer Services, Hinchingbrooke Health Care NHS Trust, Huntingdon, UK; 8) Cancer Research Team, Norfolk & Norwich University Hospitals NHS Foundation Trust, Norwich, UK; 9) Department of Oncology, The Queen Elizabeth Hospital King's Lynn NHS Foundation Trust, Kings Lynn, UK; 10) Cancer Services, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK; 11) Department of Obstetrics and Gynaecology, Norfolk and Norwich University Hospitals NHS Foundation Trust, Norwich, UK; 12) Department of Primary Care and Public Health, University of Cambridge, Cambridge, UK; 13) Department of Psychology, University of Chester, Chester, UK; 14) Cancer Research UK Cambridge Institute, Cambridge, UK.

Background: Genetic testing for germline mutations in BRCA1/2 has been available for many years, but until recently technical complexity and cost has limited testing access to high-risk cases. The advent of targeted therapies for BRCA1/2 mutation carriers necessitates new service delivery models for genetic testing. The Genetic Testing in Epithelial Ovarian Cancer (GTEOC) Study explores the feasibility and acceptability of offering genetic testing to all women recently diagnosed with EOC. **Methods:** From 1st July 2013 to 30th June 2015 women newly diagnosed with EOC were recruited though 6 sites in East Anglia, UK, which has an outbred population of 2. 5m with no BRCA1/2 founder mutations. Eligibility was irrespective of patient age and family history of cancer. Consent for genetic testing was obtained by the research team after participants reviewed the study information sheet. BRCA1/2 testing was done by NGS and MLPA. The psychosocial arm of the study utilised quantitative questionnaires (Depression Anxiety and Stress Scale, DASS, Impact of Event Scale, IES) and qualitative interviews subjected to Interpretive Phenomenological Analysis. **Results:** 208 women have been recruited and testing completed in 186. The mean age at diagnosis was 65 years (range 30-90). 13 mutations were detected (9 in BRCA1, 4 in BRCA2) giving a yield of 7%. The mutation yield is 10. 5% in unselected women <70 (11/105) and 2. 5% in unselected women 70+ (2/81). Testing only those with a positive family history (1o or 2o) increases the mutation yield to 13% in women <70 (9/70), but 15% of mutations in this age group will be missed. Preliminary analysis of the first 81 completed questionnaire responses showed that IES and DASS scores in response to genetic testing were significantly lower than equivalent scores of IES and DASS in response to cancer diagnosis ($p<.$ 001). Correlation tests revealed that whilst older age is a protective factor against any traumatic impacts of genetic testing (as measured by the IES, $p<.$ 05), no significant correlation exists between age and distress outcomes. **Conclusion:** The mutation yield in an unselected cohort of women diagnosed with EOC from a heterogeneous population with no founder mutations is 7% in all ages and 10. 5% in women under 70. Population-based genetic testing appears to be acceptable to patients and is less resource-intensive than standard practice where all patients have a full assessment by the genetics team prior to testing.

2740T

Desmoid tumor as a presenting symptom of FAP in a 54 year old man with a disease causing mutation in APC and family history of FAP. *C. Vinkler[1], K. I. Dayan[2], D. Lev[1], A. Singer[4], R. Brener[3].* 1) Inst Med Gen, Wolfson Med Ctr, Holon, Israel; 2) Inst Oncology, Wolfson Med Ctr. Holon, Israel; 3) Dpt of Surgery, Wolfson Med Ctr. Holon, Israel; 4) Medical Genetic Unit, Barzilai Medical Center, Ashkelon Israel.

Sporadic desmoid tumors is estimated at 1 to 2 per million whereas more than 15% of patients suffering from familial adenomatous polyposis (FAP) develop such extraintestinal tumors. FAP is a colon cancer susceptibility syndrome in which hundreds to thousands of colonic polyps develop. Typically beginning at the 2nd decade of life. By age 35 years, 95% of individuals with FAP have polyps. Different phenotypes ofFAP have been described and some genotype-phenotype correlations have been raised, associated with different sites of germline mutations in the adenomatous polyposis coli (APC) gene. We report on a 53-year-old male presenting first with intestinal obstruction. Colonoscopy revealed two small adenomatous polyps in the colon. Gastroscopy demonstrated more than 200 polyps located in the fundal part of the stomach. CT showed a large abdominal mass . A large tumor of the peritoneum was removed by surgery and was later diagnosed as desmoid tumor. The pedigree revealed other relatives with adenomatous polyposis. They werepreviously diagnosed as FAP mostlyduring their 3rd-4rddecade of life. In those relatives the disease manifested with polyposis of the colon requiring preventive surgery or colorectal cancer with death at an early age. The probandhadnormal colonoscopies and was asymptomatic until a short time before he was admitted to the hospital. DNA sequencingin the proband showed a frameshift mutationin exon 8 c. 896_897delCT(het); p. Ser299Cysfs*27; rs397515735. Desmoid tumor at an older age, with no GI manifestation is rarely reported as a presenting symptom of FAP. This unusual presentation demonstrates the variability in expression of this mutation. It may be postulated thatdifference in genetic background along with environmental factors may be involved in modifying the phenotypic expression in this case.

2741W

A comparative analysis of the mutaome profile of 10 frequently mutated genes in newly diagnosed and refractory and relapsed acute myeloid leukemia. *Y. Zhang, F. Wang, X. Chen, J. C. Fang, W. Teng, X. P. Han, M. Y. Wang, H. Wang, C. R. Tong, H. X. Liu.* Molecular Medical Laboratory, Ludaopei Hematology and Oncology Center, Beijing, China.

Objective To investigate the mutaome profiling of 10 frequently mutated genes in newly diagnosed and refractory and relapsed acute myeloid leukemia (AML) patients. **Methods** 125 newly diagnosed AML patients and 148 refractory and relapsed AML were enrolled, in which the median age is 24 and 26 respectively. Mutations of 10 frequently mutated genes were detected by Sanger sequencing in bone marrow samples, including kinase gens FLT3 and KIT, transcription factor genes CEBPA, NPM1 and PHF6, as well as epigenetic regulation genes ASXL1, DNMT3A, TET2, IDH1 and IDH2. **Results** (1) Mutations of the 10 genes were detected in 70. 40% of newly diagnosed AML and 62. 16% of refractory and relapsed AML. CEBPA (25. 58%), NPM1 (13. 95%) and FLT3-ITD (13. 18%) were the most frequently mutated three genes in newly diagnosed AML, and that were FLT3-ITD (19. 59%), KIT (12. 84%) and CEBPA (11. 49%) in refractory and relapsed AML. Concurrent mutations of more than one gene is statistically significant higher in newly diagnosed AML (28. 00% vs 10. 14%, P<0. 01). Mutations of CEBPA in newly diagnosed AML is also higher (25. 60% vs11. 49%, P<0. 05). (2) We observed consistent rules in the occurrence of gene mutation in two groups. KIT mutations usually occurred alone, while IDH1 and IDH2 often mutated accompanied with other genes. The kinase genes FLT3 and KIT were mutually exclusive. 80% among patients who carried more than one gene were accompanied with transcription factor gene mutation. Transcription factor genes mutated accompanied with kinase genes and/or epigenetic regulation genes is the most common mutated combination. (3) When analyzed the two groups according to age of ≤35 and >35, refractory and relapsed AML patients who were in the group of ≤35years carry more mutations of single gene (60. 19% vs 31. 11%, P<0. 05), in which KIT and FLT3-ITD account for 49. 21%. While patients in the group of >35years carry more mutations of more than one gene(20. 00% vs 6. 80%,P<0. 05). Moreover, the mutation of NPM1 were higher in>35 yearsin both groups (28. 13% vs 9. 68%, 20. 00% vs 2. 91%, P<0. 05). **Conclusions** The results showed that there were certain rules in the mutaome profiling of AML patients, which was related to the function classification of genes and age. So analysis of the mutaome profiling comprehensively is contribute to investigate its role in the development of AML as some genes affected clinical prognosis alone while some required coordinated with other genes.

2742T

Alternative genetic treatment technique for Leukemia. *S. Rashmi.* BIOTECHNOLOGY, AMITY UNIVERSITY, NOIDA, India.

The widely used chemotherapy for the hematological malignancy acquired genetic disease has transformed the disease from fatal to chronic. The development of Trypsin- inhibitor drug imatinib targeting the Philadelphia chromosome was considered as a boon in the treatment of chronic myeloid leukemia. But the question lies, is inhibition of *BCR-ABL* activity really helping in the complete eradication of the stem cell? There is a unanimous agreement that the stem cell survival don't depend on the presence or absence of the chimeric gene activity. Imatinib, a lifelong taken drug, inhibits the kinase activity exhibited by the *BCR-ABL* fusion gene without affecting the Leukemia stem cell survival. These stem cells with the help of cytokines survive and proliferate. Such cells have the potential to revert back. Out of the 95% patients responding to the medicine most of them are detected by RT-PCR with the presence of *BCR-ABL* transcripts. While the others with no such detection experiences recurrence on discontinuing the drug which proves the potential of the stem cells surviving on cytokines to develop cancer. Scientists are engaged in studying the mechanism behind the imatinib inhibition and persistence of *BCR-ABL* gene so as to come up with more cost efficient, with lesser side effects and secondary complications, and restrains the survival of the root cause that are the stem cells. It has been observed in literature that the RNA-binding protein Musashi2 (*MSI2*) plays a significant role in the self renewal programming of the leukemic stem cell by maintaining efficient translation of *Hoxa9*, *Myc*, and *Ikzf2* mRNAs. However, reduction of the colony formation, decrement in proliferation, and increment of apoptosis can be done by the depletion of MLL targeting the *Ikzf2* in leukemia stem cell. MSI can act as a potential therapeutic target for myeloid leukemia.

2743W

Hereditary Leiomyomatosis and Renal Cell Cancer (HLRCC): A 10-year retrospective chart review of Eastern Ontario referrals. *P. T. Bhola, C. Gilpin, G. E. Graham.* Regional Genetics Program, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada.

Hereditary leiomyomatosis and renal cell cancer (HLRCC) is caused by autosomal dominant germline mutations in the fumarate hydratase (*FH*) gene. It is characterized by cutaneous leiomyomas, uterine fibroids and early onset renal malignancies that are typically of a type 2 papillary histology and aggressive. Estimates of renal cancer development in HLRCC are approximated at 15%. The rate of positive genetic testing with more than one cutaneous leiomyoma is 90-100%. Using a retrospective chart review, we sought to characterize the population of patients referred to the Regional Genetics Program at the Children's Hospital of Eastern Ontario for HLRCC from 2004-2014 and compare our experience with that described in the literature. Fifty-six patients, including 10 families, were seen for this indication from 2004-2014. Most of the referring providers were from primary care (family history of HLRCC) and dermatology (personal history of cutaneous leiomyoma(s)). Less common reasons for referral included papillary renal cell carcinoma (RCC) and a family history of fumarase deficiency. Fifty-four of the 56 patients (96%) pursued *FH* testing for HLRCC. Forty (74%) were positive for a clearly pathogenic mutation or likely pathogenic variant. The average age at presentation of the positive test population was 37. 8 years; 17 were male and 23 were female. Three patients presented with papillary RCC at an average age of 25. 7 years; all were the probands in their families and all were *FH* mutation positive. Of the positive test population, 20 patients were referred for a positive family history; when evaluated in our clinic, 3 had suspected cutaneous leiomyomas on exam. Of the new probands referred for cutaneous leiomyoma(s), 16 of 17 pursued genetic testing and 13 (81%) were positive for a clearly pathogenic mutation or likely pathogenic *FH* variant. Of the 3 mutation-negative probands referred for cutaneous leiomyoma(s), 2 had only a single leiomyoma but 1 had several leiomyomas, uterine fibroids and a family history of RCC. Our HLRCC population is similar to those described. Our patients with RCC were younger at diagnosis than the average age in the literature but all were the presenting probands in their families. Our experience confirms reports that multiple cutaneous leiomyomas and papillary RCCs are sensitive markers of HLRCC, suggesting a need to target dermatologists, oncologists and urologists for awareness-raising initiatives.

2744T

***APC* mosaicism in a young woman with juvenile fibromatosis.** *E. Grindedal[1], O-J. Norum[2], T. Berg[3], T. Hølmebakk[4], K. Aaberg[3], S. Steigen[3,5], AT. Stormorken[1].* 1) Section for Inherited Cancer, Department of Medical Genetics, Oslo University Hospital, Oslo, Norway; 2) Section for Orthopaedic Oncology, Department of Orthopaedic Surgery, Oslo University Hospital, Oslo, Norway; 3) Department of Clinical Pathology, University Hospital of North Norway, Tromsø, Norway; 4) Department of Abdominal and Pediatric Surgery, The Norwegian Radium Hospital, Oslo University Hospital, Oslo, Norway; 5) Department of Medical Biology- Tumor Biology Resarch Group, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway.

Background: Familial adenomatosis polyposis (FAP) is caused by germline *APC* mutations. The classic form is characterized by hundreds to thousands of adenomas in the colorectum and early onset colorectal cancer (CRC) if left untreated. FAP is also associated with numerous extra-colonic manifestations. It is often referred to as Gardner syndrome when observed in combination with osteomas and soft tissue tumours (epidermoid cysts, fibromas and desmoid tumours). Somatic mosaicism has been reported in a few sporadic cases of polyposis, but may be an underestimated cause of the disease. This case report presents the detection of a mosaic *APC* mutation in a young woman with juvenile fibromatosis. **Clinical presentation:** Our patient had been diagnosed with juvenile fibromatosis as a child following removal of several subcutanous soft tissue tumours and epidermoid cysts throughout childhood. At 26 years she was referred to the Sarcoma group at the Norwegian Radium Hospital with a large soft tissue lesion in her left thigh, described histologically as inactive fibromatosis. A subcutanous soft tissue tumour was detected in her right buttock, histologically described as fibrous tissue but not fibromatosis. Based on her medical history, Gardner syndrome was suggested and upper and lower endoscopic examinations performed. Fifty adenomas were detected in her colon and 50 to 70 fundic gland polyps and adenomas in her stomach. Sanger sequencing of DNA from peripheral blood indicated presence of a pathogenic mutation c. 4348C>T (p. Arg1450*) in the *APC* gene. The signal representing the mutation was very weak, indicating somatic mosaicism. DNA was extracted from cells of her normal colonic mucosa and from adenoma tissue. The mutation was present at apparent heterozygous levels in both. There was no history of polyps or CRC in her family, and it was confirmed that she is a *de novo* carrier. The mutation is situated in the area of the *APC* gene associated with desmoid development. Based on her high risk of CRC, her present desmoid and risk of future desmoids, prophylactic proctocolectomy was recommended over colectomy with ileorectal anastomosis. Oesophagogastrosduodenocopies will be performed annually. **Conclusion:** Our case illustrates that a history of soft tissue tumours, and not just polyposis, may be an indication for *APC* genetic testing, and that *APC* mosaicism may be the cause of single cases of polyposis but also of the extra-colonic features of FAP.

2745W

Phenotypic delineation and bioinformatic detection of the pathogenic variant 5083DEL19 in *BRCA1* in two brazilian patients. *R. M. Minillo, P. M. Oliveira, C. L. Clozato, R. S. Francisco, M. C. Cervato, N. H. Muto, F. T. Lima, R. Sitnik, J. R. R Pinho, C. L. P. Mangueira.* Molecular Genetics, Albert Einstein Israelite Hospital, São Paulo, Brazil.

Introduction: Prevalence of pathogenic variants associated with the Hereditary Breast and Ovarian Cancer Syndrome varies according to different ethnicities and geographical regions. Next generation sequencing (NGS) is highly used for detecting these variants, although some of them are tricky to detect, either because of their chromosomal location and/or differences in the algorithms and parameters applied in the commercially available variant calling softwares (VCS) that integrate the bioinformatics pipelines (BP). One of these variants, c. 4964_4982del19 (5083del19), in *BRCA1*, is a founder mutation of South Italy (Calabria), associated with breast tumor that lack hormonal receptors expression, high levels of proliferation markers and ovarian serous adenocarcinoma. It was identified in two Brazilian patients with personal history of breast and/or ovarian cancer. **Objective:** To clinically describe the variant 5083del19 in *BRCA1* detected by NGS in two unrelated Brazilian patients. **Methods:** Review of medical records and BP used for the detection of the variant. **Results:** Patient 1 had triple negative breast carcinoma (with ductal infiltration) at age 50, skin cancer at age 55, ovarian serous adenocarcinoma at age 57 and rectal cancer at age 61. Her family is from South Italy; her mother had breast cancer at age 38, just like one niece of the maternal grandmother and one niece of the paternal grandfather. Patient 2 had ovarian serous adenocarcinoma ate age 40. Her family is from Minas Gerais (a southeastern Brazilian state); three aunts and two paternal cousins had cancer among ages 40 and 50. There is no information on anatomic pathology of the relatives' tumors. Both patients presented the same pathogenic variant, which was identified in each patient separately. We found that this variant is neglected when the default parameters of the VCS are applied, which makes it necessary to customize specific filters in the softwares for its correct identification. **Conclusion:** The patients herein described presented typical tumors associated with the 5083del19 in *BRCA1*. At least one of them quotes Italian ancestry, and also had rectal cancer. The high prevalence of this variant in individuals of Italian ancestry makes it epidemiologically relevant in Brazil, since there are about 17 million Italian-descendent individuals in Southeastern Brazil. We recommend that this variant should be manually screened by BP in *BRCA1* e *BRCA2* NGS panels.

2746T

Neuroendocrine tumors of the lower gastrointestinal tract in hereditary colon cancer syndromes. *C. Pedley, A. Blanco, M. Myers.* University of California, San Francisco , San Francisco, CA.

Neuroendocrine tumors are relatively rare; in 2004 the estimated incidence was 5. 25/100,000. In addition to being rare in the general population, neuroendocrine tumors of the lower gastrointestinal (GI) tract have not traditionally been associated with hereditary gastrointestinal cancer syndromes. However, with the arrival of next generation sequencing panels, the spectrum of reported cancers within each syndrome is growing, and there are documented case reports of neuroendocrine tumors of the lower GI tract arising in Lynch syndrome [MIM 120435] and Familial Adenomatous Polyposis (FAP [MIM 175100]). Six patients seen in the University of California, San Francisco (UCSF) Cancer Genetics and Prevention Program will be presented, each diagnosed with pathology confirmed neuroendocrine tumors of the lower GI tract and molecularly confirmed mutations in either *APC* [MIM 611731], *MLH1* [MIM 120436], *MSH2* [MIM 609309], *MSH6* [MIM 600678], *CHEK2* [MIM 604373] or *MUTYH* [MIM 604933], the latter two being homozygous mutations. Neuroendocrine tumors of the lower GI tract, while rare, are seen in inherited gastrointestinal cancer syndromes. Patients presenting with neuroendocrine tumors and other features to suggest an inherited syndrome, such as age of diagnosis and personal and/or family history of cancer should be referred for genetic counseling and genetic testing. Genetic testing with a next generation sequencing panel should be considered due to the occurrence of neuroendocrine tumors of the lower GI tract in several different inherited cancer syndromes.

2747W

Double heterozygosity for *BRCA1* and *BRCA2* pathogenic variants in a French metastatic breast cancer patient. *A. Vilalta[2], E. Curtit[1], G. Meynard[1], C. Villanueva[1], L. Mansi[1], M. Chaix[1], J. Kuo[2], M. Villa[2], J. Neidich[2], A. Tomar[2], A. Arianpour[2], P. Lebahar[2], X. Pivot[1].* 1) Service Oncologie Médicale, CHRU Besançon - IRFC , 3 Boulevard A. Fleming 25030 BESANCON Cédex; 2) Pathway Genomics Corporation, 4755 Nexus Center Drive, San Diego, CA 92121.

Double heterozygosity is an extremely rare occurrence in hereditary breast and ovarian cancer syndrome (HBOC) where two pathogenic variants, one in *BRCA1* and one in *BRCA2*, are found in an individual. To date, only a few case reports and case series have been reported in the literature (1-3). Furthermore, little is known about the clinical characteristics, family history, and tumor histology in these patients. In this study, we utilized targeted gene testing with next-generation sequencing (NGS) technology in an metastatic breast cancer patient from France. We evaluated germline variants using Pathway Genomics' BRCATrue NGS test, which analyzes variants in all exons and exon flanking regions in both *BRCA1* and *BRCA2* genes. All variant calls were determined after alignment and mapping to the GRCh37/hg19 reference genome. Variant calls were confirmed by Sanger sequencing. In this patient, a c. 1016dupA (p. V340GfsX6) frameshift variant was found in *BRCA1* along with a c. 6814delA (p. R2272EfsX8) frameshift variant in *BRCA2*. Both frameshift variants are predicted to truncate the BRCA proteins. The *BRCA1* c. 1016dupA variant is considered a Norwegian founder mutation but has also been observed in individuals of other ancestries (4-7). The *BRCA2* c. 6814delA (p. R2272Efs*8) pathogenic variant, also known as 7042delA, is predicted to truncate the BRCA2 protein and has been identified in individuals with a personal or family history of breast and/or ovarian cancer (8,9). To the best of our knowledge, the combination of these two pathogenic variants in an individual has not been previously reported. In a clinical diagnostic setting, the possibility of double heterozygosity of pathogenic variants in more than one susceptibility gene should be considered, especially in patients with early-onset metastatic cancers. Furthermore, genetic testing and genetic counseling should also be indicated for high-risk family members. meta1. (2012) *Breast cancer research and treatment* **134**, 1229-12392. (2011) *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* **22**, 964-9663. (2012). *Breast cancer*4. (1996) *American journal of human genetics* **59**, 486-4875. (2012) *Nucleic acids research* **40**, D992-10026. (1999) *American journal of human genetics* **65**, 671-6797. (1994) *Nature genetics* **8**, 392-3988. (2012) *International journal of oncology* **41**, 1619-16279. (2014) *Maturitas* **77**, 68-72.

2748T

The changing landscape of hereditary cancer testing: Lessons learned from 1703 patients seen at an NCI designated cancer center. *G. L. Wiesner[1,2], S. Lewis[1,2], J. Holt[1,3], R. H. Morgan[1], D. A. Riddle[1], S. W. Trump[1], K. McReynolds[1,2].* 1) Vanderbilt-Ingram Cancer Center, Vanderbilt Univ, Nashville, TN. , USA; 2) Department of Medicine, Vanderbilt Univ, Nashville, TN, USA; 3) Department of Biomedical Informatics, Vanderbilt Univ, Nashville, TN, USA.

Intro: Next generation sequencing (NGS) multi-gene panels were introduced in 2012 for genetic testing of patients at risk for hereditary cancer. These tests are controversial due to the lack of clinical guidelines related to rare genes on the panel and the possibility of a VUS result. It is also possible that more at risk patients will be identified with pathogenic mutations compared to usual testing protocols. Methods: We tested this hypothesis by examining the outcomes of genetic test reports from 1703 patients seen in the Hereditary Cancer Program at Vanderbilt-Ingram Cancer Center from July 2012 through Dec 2014. The tests were tabulated by type of test (single site, single genes such as BRCA1/2, or NGS panel), type of provider, and date of test. Results: Most of the 1703 patients were female (88%) and were seen for an indication of a personal history of breast cancer (38%) or a family history of breast cancer (72%) regardless of cancer history. The group was predominately white non-Hispanic (77%). Genetic tests were submitted on 1145 patients (67%), with 186 having a single site (16%), 346 (30%) single gene, and 613 (54%) multi-gene panel tests. 180 (16%) patients were identified with a pathogenic mutation, a VUS (239; 21%) or both (23; 2%). 70 (34%) pathogenic mutations were identified by multi-gene panels, 47 (23%) by single gene tests, and 86 (46%) by single site tests. There was no difference in the overall rate of pathogenic mutations identified by panel (11. 4%) compared to single gene tests (13. 6%). However, when the comparison was restricted to 497 tested women with breast cancer, 13 of the 189 (7%) tested with Sanger BRCA1/2 had pathogenic mutations compared to 40 mutations found in 308 (13%) with multi-gene testing (P=0. 035). There was a significant difference between MD or NP providers (67%) compared to genetic counselors (57%) in preference for multi-gene panels compared to single gene tests (P=0. 002). Over the 30-month study period, the proportion of panel tests sent increased markedly (P=0. 0001). Conclusion: The testing paradigm has shifted from testing a small number of genes to a larger multi-gene sets, with the potential to identify a greater number of pathogenic mutations, at least in patients with breast cancer. We also found clear differences in ordering practices by type of provider. Future studies should examine whether the clinical indication, cancer diagnosis, and patient desires are important factors guiding the choice of test.

2749W

Genetic counseling and analysis plan in Von Hippel-Lindau syndrome and pheochromocytoma. *W. Smaoui[1], A. Bahloul[1], H. Fourati[1], K. Chaker[1], O. Kaabi[2], R. Louati[2], MN. Mhiri[1], NB. Abdelmoula[2].* 1) Dep Urology, HB Univ Hospital, Sfax, Tunisia; 2) Lab Histology, Univ Medicine, Sfax, Tunisia.

Pheochromocytomas (PCCs) are rare catecholamine-secreting tumors that usually arise within the medulla of the adrenal gland or the sympathetic ganglia. The hormone imbalance affects the cardiovascular system leading to hypertension and other metabolic processes. PCCs are genetically determined in approximatelyone thirdofpatients by germline/somatic mutations in 11 known susceptibility geneslike RET,VHL, NF1, HIF2A and SDH. PCC treatment involves treatment of hypertension and surgical removal of the tumor. When PCC is the presenting tumor of syndromic forms, diagnosis may be difficult or missed, as often the first symptoms are labile hypertension that can be confused for other diagnoses. We report two Tunisian female patients who were surgically treated for pheochromocytoma associated to renal lesions. Case Report 1 was a 50-year-old woman who presented multiple cerebellar hemangioblastomas, cervical spinal cord and recurrent retinal angiomas. She was surgically treated for left clear cell renal cell carcinoma (ccRCC) and one year after for right PCC. She had also multiple pancreatic cysts and an ovarian cyst. Case Report 2 was a 39-year-old woman who underwent surgery for right PCC. At the same time clinical management showed bilateral polycystic kidneys as well as an ovarian cyst. Skin lesions and a positive family history with a spectrum of clinical phenotypes, evoking familial von Hippel–Lindau (VHL) disease were recorded for our two patients. As VHL (MIM 193300) is an autosomal dominant disorder caused by heterozygous mutations in the VHLtumor suppressor gene at 3p25. 3 linked tomTOR pathway, a genetic screening forVHLgene germline mutation detection was proposedat genetic counseling. Regarding phenotype presentation with PCC classed as VHL type 2B, VHL R167Q will be our first genetic analysis purpose, even though no therapeutic approach is available to target it. In fact, this missense mutation is the most common VHL point mutation in hereditary VHL disease. It is also a representative type 2B mutation that predispose to a high risk of ccRCC and finally it has been found in sporadic PCCs.

2750T

Examining severity of cancer history in patients with pathogenic variants in *ATM*. *J. Abernethy, E. Rosenthal, J. Kidd, P. Vail, S. Manley.* Myriad Genetic Laboratories, Inc. , Salt Lake City, UT.

Background: We tested the hypothesis that different types of pathogenic variants (PVs), particularly missense versus truncating variants, confer different levels of cancer risk in *ATM* carriers and their families. **Methods:** Patients ascertained for suspected hereditary cancer risk were tested with a clinical 25-gene hereditary cancer panel between September 20, 2013 and March 13, 2015. Clinical history data for those with PVs in *ATM* was obtained via healthcare provider report on the test request forms. 23 patients with PVs in more than one gene were excluded from the cohort. Variants classified as Deleterious or Suspected Deleterious by our laboratory were regarded as pathogenic. Many *ATM* missense variants are classified as pathogenic based on their contribution to the recessive condition Ataxia-Telangiectasia (A-T). **Results:** Of 59 female patients with a missense pathogenic variant in *ATM,* 17 (28. 8%) had a personal diagnosis of breast cancer at or before age 50. Of 310 female patients with a truncating variant in *ATM*, 86 (27. 7%) had a personal history of early onset breast cancer. Of 60 patients (males and females) with a missense variant in *ATM*, 32 (53. 3%) had a high-risk family history (3 or more diagnoses in the family, including the patient where applicable, of breast and/or pancreatic cancer). Of 316 patients with a truncating variant in *ATM*, 138 (43. 7%) had a high-risk family history. **Discussion:** This study identified no evidence for a significant difference in cancer risks associated globally with pathogenic missense compared to truncating variants, although wecannot rule out higher or lower penetrance associated with individual variants. As additional families are identified, these findings can be confirmed and the relative cancer risks associated with individual *ATM* variants can be further characterized.

2751W

Renal manifestations of TSC and early onset of hypertension: About two Tunisian familial cases. *N. Abdelmoula[1], W. Smaoui[2], K. Chaker[2], O. Kaabi[1], H. Fourati[2], MN. Mhiri[2].* 1) Lab Histology, Univ Medicine, Sfax, Tunisia; 2) Dep Urology, Univ HB Hospital, Sfax, Tunisia.

Renal manifestations of tuberous sclerosis complex (TSC) include angiomyolipomas, simple cysts, polycystic kidney disease, and renal-cell carcinoma. These lesions likely arise in infancy or early childhood, increasing in size and number with age. After neurologic complications, renal involvement is the second most common cause of morbidity and mortality in TSC. The most common renal lesions are angiomyolipomas found in 80% of TSC patients and renal cysts found in 17% of children and 47% of adults. Renal cysts are frequently multiple and bilateral. Polycystic kidney disease may also occur and is a more severe lesion with innumerable cysts that enlarge, replacing renal parenchyma, and causing renal insufficiency and hypertension typically at an early age. Here, we describe two Tunisian patients in whom TSC diagnosis was confirmed at the clinical level for both and molecularly for one of them. The first case was a 36-year-old man and harbored multiples bilateral renal cysts with two right angiomyolipomas rapidly transformed into renal cell carcinoma (RCC). In fact, surgical resection of five tumors with a very large one in the central regionleaded to phenotypic characterization of two central lesions as a Grawitz tumor and three peripheral ones as angiomyolipomas. Familial germinal small TSC1 deletion at codons 213-214 of exon 5 of the gene was confirmed in this patient. The second patient was a 25-year-old woman harboring at the renal level, bilateral multiples angiomyolipomas leading to total left kidney nephrectomy. Immunohistochemical analysis of renal tumors showed typical perivascular epithelioid cell phenotype. These two patients presented an early onset hypertension. High blood pressure in STC is generally related to renal cysts. We suggest through our observations that hypertension may be associated directly to angiomyolipomas genetic evolution and mTOR signaling dysregulation. In fact, relationships between hypertension, cardiac hypertrophy, Angiotensin II as well as dysregulation of signaling pathways and structural remodeling of vessel walls in angiomyolipomas were studied in animal models and needed confirmation and delineation in TSCs positive patients. .

2752T

Bilateral, multifocal renal tumors diagnosed as Birt-Hogg-Dubé syndrome confirmed by genetic analysis. *S. Park, S. Park, E. Kang, H. Cheong, D. Kim, K. Oh, Y. Kim, K. Joo, H. Lee.* Seoul National University Hospital, Seoul, South Korea.

Introduction: Birt-Hogg-Dubé syndrome (BHD) is an autosomal dominant disorder characterized by skin fibrofolliculomas, pulmonary cysts and spontaneous pneumothorax, and renal cancers. The syndrome is caused by germline mutations of the *FLCN* gene located in *17p11. 2* encoding folliculin. The risk of renal cancer is seven times higher in BHD patients. Bilateral, multifocal and chromophobe renal cell carcinoma is characteristic of renal cancers in BHD. Among about 110 pathogenic mutations reported in BHD, only twenty six of them have been associated with renal cancer. In this study, we described a case of 50-year-old woman with chromophobe renal cell carcinoma who had c. 1557delT mutation in the *FLCN* gene which is novel in BHD-associated renal cancer. **Case report:** A 50-year-old-woman presented with flank pain. A CT scan of abdomen showed multifocal tumors in both kidneys. To differentiate the origin of the tumors, a chest CT and a whole body PET-CT were carried out, although there was no other lesion except a few lentiform cysts at basal lungs. Histologic evaluation through a needle biopsy revealed that the mass was chromophobe type renal cell carcinoma. The patient underwent partial nephrectomy. Given multifocal distribution of the renal cell carcinoma, we suspected an inherited form of kidney cancer. Although she denied any specific familial history, the cystic change of both lungs and chromophobe renal cell carcinoma pointed towards the possibility of BHD. DNA sequencing of the entire *FLCN* gene identified a heterozygous c. 1557delT mutation in exon 14 [p. Phe(TTT)519Leu(TTA) fs*18]. **Discussion:**Once multifocal renal masses are diagnosed as renal cell carcinoma, histologic subtype, family history and other clinical features should be considered to recognize familial renal cancer syndromes. Multifocal chromophobe renal cell carcinoma should prompt genetic test for BHD even in a patient without typical skin lesion. In our patient, c. 1557delT mutation was found out which has been first reported in BHD-associated renal cancers.

2753W

Cytosine Modification Signatures Distinguish Clinical Subtypes of Glioblastoma Stem Cells. *W. Zhang[1,2], T. Shi[3], Z. Zhang[1], C. Liu[4], C. Horbinski[5], S. Y. Cheng[2,3].* 1) Department of Preventive Medicine, Northwestern University , Chicago, IL; 2) The Robert H. Lurie Comprehensive Cancer Center; 3) The Ken & Ruth Davee Department of Neurology, Northwestern University, Chicago, IL; 4) Department of Bioengineering, University of Illinois, Chicago, IL; 5) Department of Neurological Surgery, Northwestern University, Chicago, IL.

Glioblastoma multiforme (GBM), the most malignant form of primary brain tumors, is diagnosed in over 18,000 new patients in the United States every year and kills ~13,000. Despite access to state-of-art modalities of therapy, GBM prognosis remains dismal. A major challenge for treating these patients is that GBM is in reality a heterogeneous group of neoplasms, wherein discrete subsets feature unique patterns of pathogenesis, cellular biology, and prognosis. Though specific molecular markers are of considerable value in clinical care, including mutations in IDH1/2 and 1p/19q, there is still a great deal of room for improvement in prognostic stratification and targeted therapeutics. In particular, aberrantly high tumor heterogeneity in intra- and inter tumors of high-grade glioma (HGG) is recognized by four clinically relevant GBM subtypes based on distinguished gene signatures. However, molecular signaling in glioma stem cells (GSCs) in individual HGG subtypes remains to be fully characterized. Given that alterations in DNA methylation, a type of cytosine modifications, have been associated with survival and response to treatments in GBM patients (e. g. , methylation of *MGMT* promoter and response to temozolomide), these epigenetic markers may have the potential to distinguish GBM subtypes. Specifically, our previous gene expression analysis distinguished two mutually exclusive GSC subtypes: proneural (PN) from mesenchymal (Mes) GSCs, and revealed a pronounced correlation with the corresponding PN or Mes HGGs. In this study, we profiled whole-genome cytosine modification levels of eight PN and four Mes GSCs using the Illumina 450K arrays. At a raw p value of 0. 005, we identified 171 differential CpGs in 172 genes between PN and Mes GSCs. A hierarchical clustering analysis based on these differential CpGs showed complete separation between PN and Mes GSCs. In addition, these differential CpG sites were enriched in canonical pathways and Gene Ontology biological processes such as “anti-apoptosis”. Notably, an enriched pathway - “proteoglycans in cancer”- is related to multiple oncogenic pathways in tumor cells, and promotes critical tumor-microenvironment interactions. In summary, cytosine modifications appear to play a critical role in determining the molecular characteristics of GSC heterogeneity, suggesting potential applications of methylomic markers to distinguish GBM subtypes for the ultimate goal of precision medicine in these patients.

2754T

Differential Expression of miR-139-5p as a Saliva Biomarker in Tongue Squamous Cell Carcinoma Patients. *M. B. Duz[1], O. F. KARATAS[2], E. GUZEL[1,5], N. F. TURGUT[3], M. YILMAZ[3], C. J. CREIGHTON[4], M. OZEN[1,5,6].* 1) Department of Medical Genetics, Cerrahpasa Medical School, Istanbul University , Istanbul, Turkey; 2) Molecular Biology and Genetics Department, Erzurum Technical University, Erzurum, Turkey; 3) Department of Otorhinolaryngology, Cerrahpasa Medical School, Istanbul University, Istanbul, Turkey; 4) Department of Medicine and Dan L. Duncan Cancer Center Division of Biostatistics, Baylor College of Medicine, Houston, Texas, USA; 5) Biruni University, Department of Molecular Biology and Genetics, Topkapi, Istanbul, Turkey; 6) Department of Pathology & Immunology Baylor College of Medicine, Houston, TX, USA.

41% of all oral carcinomas have been found to localize at tongue, where they were characterized as being aggressive and having capacity to locally invade and relapse frequently. Despite considerable enhancements in cancer diagnosis and treatment techniques, tongue squamous cell carcinoma (TSCC) still remains to be one of the most common and lethal cancer types in the head and neck region. In this study, we aimed to identify a signature of saliva-specific microRNAs (miRNAs) expression in TSCC patients. To explore putative diagnostic biomarkers, we compared the miRNA profiles of saliva samples obtained from 4 TSCC patients and 4 healthy control individuals using Agilent miRNA microarray (V19). Three of the differentially expressed miRNAs were selected for further validation with quantitative reverse-transcription PCR (qRT-PCR) using saliva samples of 25 TSCC patients and 25 healthy control individuals. Microarray analysis demonstrated that 419 miRNA probes were deregulated in TSCC patients when compared to control group and qRT-PCR results validated the reduced expression of miR-139-5p in TSCC saliva. Further analysis of post-operative saliva samples of TSCC patients revealed that miR-139-5p level elevated to normal level after surgery, pointing its diagnostic and prognostic power. As a conclusion, we propose saliva as a feasible source in routine patient examination for early diagnosis of TSCC patients, and our results suggest saliva miR-139-5p as a novel potential diagnostic marker.

2755W

HDACi-induced differentiation of myelogenous leukemia results in targeted chromatin accessibility changes. *C. L. Frank[1,2], D. S. Hsu[2], G. E. Crawford[2,3].* 1) Department of Molecular Genetics and Microbiology, Duke University, Durham, NC; 2) Center for Genomic and Computational Biology, Duke University, Durham, NC; 3) Department of Pediatrics - Division of Medical Genetics, Durham, NC.

Small molecule inhibitors of histone deacetylases (HDACi) serve as potent anticancer agents for particular malignancies. The exact mechanism by how HDACi work is unclear, but there is evidence these compounds either directly promote apoptosis or sensitize cancer cells by cell cycle arrest and differentiation, facilitating synergistic combinations with other compounds. For over 25 years it has been known the myelogenous leukemia cell line K562 can be differentiated in vitro by exposure to HDACi. Despite this, the precise genomic and epigenomic changes these cells must undergo to assume an erythrocytic state remain poorly defined. To investigate chromatin-based changes responsible for anti-proliferative differentiation, we treated K562 cells for 72 hours with sub-lethal concentrations of the HDACi sodium butyrate or SAHA, and assessed global chromatin accessibility and expression changes by DNase-seq and RNA-seq. Despite the potential for global hyperacetylation, we identified several thousand specific regulatory elements (<10% of total DHS sites) that become significantly more or less accessible following treatment. These regulatory elements are enriched for non-promoter regions of the genome, nearby to genes that change expression with HDACi exposure, and contain abundant motifs for key hematopoietic lineage-defining transcription factors. We verified by ChIP-seq that the pioneer factor PU. 1 increases binding at opened hypersensitive sites and likely contributes to active enhancer formation at these sites. Increased binding does not appear to be a result of PU. 1 induction following treatment. Luciferase assays demonstrate HDACi-opened sites containing PU. 1 motifs are sufficient for the transcriptional response. Ongoing and future work will examine the necessity of factors like PU. 1 for HDACi-induced differentiation of myelogenous leukemia and other cancer cell types.

2756T

Distinctive expression profiles of lncRNAs and lncRNA-related miRNAs in glioma subtypes. *D. Glavac[1], A. Matjašič[1], E. Boštjančič[1], M. Popovič[1], B. Matos[2].* 1) Department of Molecular Genetics, Institute of Pathology, Faculty of Medicine, University of Ljubljana, Slovenia; 2) Department of Neurosurgery, University Clinical Center, Ljubljana, Slovenia.

Glioblastoma the most common and most aggressive diffuse glioma is associated with short survival and characterized by morphological, genetic and gene-expression heterogeneity including long non-coding RNAs (lncRNAs) and microRNAs (miRNAs) expression. Due to the heterogeneity, most tumours develop resistance to treatment and shortly recurrence. Our purpose was to investigate differential expression profiles of lncRNAs in glioma and their potential association with different glioma subtypes as well as association to their co-regulated or co-expressed miRNAs. We used the qPCR approach in order to determine differentially expressed lncRNAs using LncRNA array profiler. A subset of differentially expressed lncRNAs and their potentially related miRNAs was further analyzed with alternative qPCR approach on a bigger cohort of glioma samples. We found measurable differences of lncRNA expression in glioma samples compared to brain reference RNA. Our results show downregulation of *MEG3*, *JPX* and *RNCR3* and upregulation of *HOTAIR* and *ZFAS1* in all glioma subtypes. *7SL* was downregulated in astrocytoma, primary and secondary glioblastoma and upregulated in oligodendroglioma and oligoastrocytoma. We have also observed a correlation in expression of *miR-124a* and *RNCR3*, *miR-196a* and *HOTAIR*, *miR-770* and *MEG3*. In a subcohort of our patients with recurrent tumour our data showed statistically significant change in expression of several miRNA (miR-7, miR-9, miR-21, miR-26b, mirR-124a, miR-199a and let-7f) in recurrent tumour compared to the first biopsy. Expression levels in recurrent tumour significantly changed for miR-15b, let-7d and in let-7f comparing treatment with radiotherapy and chemotherapy and with radiotherapy alone. Our data suggest that in the patients with glioblastoma time to tumour recurrence is longer when using radiotherapy and chemotherapy after surgery and that expression of certain miRNAs change based on therapy used. Wehave demonstrated distinctive glioma subtype lncRNA and miRNA -related expression patterns, which may contribute to more objective diagnosis and developing new therapeutical approaches.

2757W

SFRP1 promoter hypermethylation in white blood cells is associated with sprodic breast cancer in Iranians women. *Y. Kiumarsi[1], M. Norouzinia[2], M. Hasanzad[3], A. Pooladi[4], G. Khakpoor[5], A. Dehghani fard[6], M. Ahmadvand[6].* 1) Islamic Azad University, Medical Tehran branch, Tehran, Iran,MSc; 2) Department of Medical Genetics, Tarbiat Modares University, Tehran, Iran,Associate Professor; 3) Department of Medical Genetics Islamic Azad University, Medical Tehran Branch, Tehran,Iran, Associate Professor; 4) Department of Medical Genetics, Tarbiat Modares University, Tehran, Iran, Ph. D student; 5) Department of Medical Genetics, Tehran University, Tehran, Iran, Ph. D student; 6) Department of Hematology, TarbiatModares University, Tehran, Iran, Ph. D student.

Breast cancer is leading cause of cancer death in women worldwide but early detection is corelated free survival. Changes in DNA methylation in peripheral blood could be associated with cancer at early stage. Peripheral white blood cell DNA contains epigenetic marks that can be used as potential biomarker in order to detecting individuals predisposing to malignancies including breast cancer. We examined the association between promoter hypermethylation level of SFRP1 and CDH1 genes ,which are antagonist in WNT signaling and cell-cell adhesion respectively, in peripheral blood leukocyte DNA and breast cancer among Iranian women**.** *SFRP1 and CDH1* promoter methylation was examined by MeDIP-Realtime qPCR in WBC from 30 breast cancer patients aged less than 40 years with newly diagnosed sprodic breast cancer, no history of any therapy and 30 matched controls from medical checkup examinees in Imam khomayni hospital. Out of 30 breast cancer patients, 16 (53. 3%) manifested various levels of SFRP1(p-value=0. 04) and 1 (3. 3%) showed CDH1(p-value>0. 05) methylation. However, no methylation was found in 30 controls. The present results suggest that the SFRP1 promoter methylation level of peripheral blood leukocyte DNA is low in patients with breast cancer and may be a potential biomarker for breast cancer risk but highly methylated CDH1 promoter may be related to metastasing tumours.

2758T

Selection of Allelic Expression in lung adenocarcinoma through Genome-wide Allelic Expression Imbalance Analysis. *X. Kong, J. Huang, Y. Zhou, M. Wang, P. Zhang, M. Jin, Z. Xing.* Molec Gen, Inst Hlth Sci, Shanghai, China.

For heterozygous genes, the expression levels of paternal and maternal alleles should be the same if they are in the same environment. However, multiple factors such as *cis*-action, epigenetic regulation, and mRNA processing, can affect individual gene expression, resulting in the expression variations of alleles. Allele expression imbalance (AEI), defined as the quantitative difference of the degree in gene expression between two alleles, is a widespread phenomenon in human genome. In recent studies, growing evidences have proved the connections between the AEI status of oncogenes and cancer.

By analyzing RNA and DNA sequencing data of cancer and adjacent paired normal tissues obtained from 57 The Cancer Genome Atlas (TCGA) samples and RNA sequencing from 72 South Korean lung adenocarcinoma patients, we discovered that the allelic expression was under selection. In normal tissues, genes with synonymous Single Nucleotide Variants (SNVs) occupied higher ratios in AEI than those with nonsynonymous SNVs and the specific-expression genes had higher AEI frequencies than common-expression genes, which had no differences in tumor tissues. We screened genes which had high AEI frequencies in tumor tissues and low in normal tissues and genes with low AEI frequencies in tumor tissues and high in normal tissues. GO annotation of these two groups showed that genes with high frequencies in tumor tissues were significantly enriched in protein biogenesis related functions, while the other group of genes were enriched in immune related functions. More importantly, we found that in normal tissues, more conservative sites had lower AEI frequencies. On the contrary, more conservative sites had higher AEI frequencies in tumor tissues.

We conjecture that AEI works as modulating abnormalities of single allele caused by mutations and the loss of allele expression selection may contribute to tumorigenesis.

2759W

Tumor DNA methylation profiling in African American men: A high-risk prostate cancer population. *R. Rubicz[1], S. Zhao[1,2], J. L. Wright[1,3], S. Kolb[1], B. Klotzle[4], M. Bibikova[4], D. Troyer[5], R. Lance[6], E. A. Ostrander[7], Z. Feng[8], J. -B. Fan[4], J. L. Stanford[1].* 1) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC; 3) Department of Urology, University of Washington School of Medicine, Seattle, WA; 4) Illumina, Inc. , San Diego, CA; 5) Departments of Pathology and Microbiology and Molecular Cell Biology, Eastern Virginia Medical School, Norfolk, VA; 6) Department of Urology, Eastern Virginia Medical School, Norfolk, VA; 7) Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 8) Department of Biostatistics, University of Texas MD Anderson Cancer Center, Houston, TX.

Men of African ancestry have the highest incidence of prostate cancer world-wide, and they are at increased risk of dying from the disease compared to other populations. There are a number of factors that likely contribute to this health disparity, such as socioeconomic factors, access to healthcare, type of treatment, and biological factors including genetics and epigenetics. DNA methylation is a major epigenetic regulator of gene expression that has previously been associated with prostate cancer. Here we characterized DNA methylation profiles of 76 African American men with prostate cancer who had radical prostatectomy as primary treatment. Samples were from two institutions: Fred Hutchinson Cancer Research Center (n=44), where patients were enrolled in a population-based study; and Eastern Virginia Medical Center (n=32), which included samples from a highly selected group of patients with no evidence of recurrence or who had lethal prostate cancer. The Infinium HumanMethylation450 BeadChip array (Illumina) was used to determine methylation status for ~480,000 CpG sites in prostate cancer tissues. Age- and cohort-adjusted analyses based on comparisons of M-values (minfi, Bioconductor) showed that 18 CpGs had significant differences in methylation levels between men with vs without evidence of prostate cancer recurrence, and 5 CpGs were differentially methylated in men with lethal prostate cancer (FDR q≤0. 2). DNA methylation differences were also observed between men with local vs regional pathological stage (2 CpGs) and for men with a high composite aggressiveness score (1 CpG) that was based on recurrence status, Gleason score, and pathological stage. The recurrence-associated CpGs include CpGs in the transcription start sites of *NEUROG1* (previously associated with medulloblastoma and colorectal cancer) and *PRTFDC1* (a possible tumor-suppressor gene). These results indicate that differentially methylated sites may contribute to more aggressive tumor biology in African American prostate cancer cases.

2760T

Epigenome-wide profiling of DNA methylation in paired prostate cancer versus adjacent benign tissue. *J. L. Stanford[1,2], S. Zhao[3], C. Wong[1], M. Bibikova[4], B. Klotzle[4], M. Wu[1], E. A. Ostrander[5], J. Fan[4], Z. Feng[6], M. S. Geybels[1,7].* 1) Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA; 2) Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA; 3) National Institute of Environmental Health Sciences, Biostatistics & Computational Biology Branch, NC; 4) Illumina, Inc. , San Diego, CA; 5) Cancer Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 6) MD Anderson Cancer Center, Houston, TX; 7) Department of Epidemiology, GROW School for Oncology and Developmental Biology, Maastricht University, Maastricht, the Netherlands.

Aberrant DNA methylation may promote prostate carcinogenesis. We investigated epigenome-wide DNA methylation in prostate cancer compared to adjacent benign tissue to identify differentially methylated CpGs. The study included paired prostate cancer and adjacent benign tissue samples from 20 radical prostatectomy patients. Epigenetic profiling was done using the Infinium HumanMethylation450 BeadChip. Linear models that accounted for the paired study design and False Discovery Rate Q-values were used to evaluate differential CpG methylation. Differentially methylated regions (DMRs) were called using the Probe Lasso method. The top-ranked differentially methylated genes were analyzed in relation to mRNA expression in the same paired tissue samples. There were 2,040 differentially methylated CpGs in prostate cancer versus adjacent benign tissue (Q-value <0. 001), the majority of which were hypermethylated (n = 1,946). Twenty-seven hypermethylated CpGs had a mean methylation difference of at least 40% in prostate cancer versus benign tissue, which included 25 CpGs in 17 genes. Furthermore, for ten genes over 50% of promoter region CpGs were hypermethylated in prostate cancer. The top-ranked hypermethylated genes included three genes that were associated with both promoter hypermethylation and reduced gene expression: *SCGB3A1*, *HIF3A*, and *AOX1*. A secondary analysis of DMRs showed 20 hypermethylated regions (Q-value <0. 001), several of which were in or near genes of the *HLA* gene family on chromosome 6. This study of prostate cancer versus adjacent benign tissue identified many differentially methylated CpGs and regions in and outside gene promoter regions, whichmay potentially beused for the development of future methylation-based diagnostic tests or as therapeutic targets.

2761W

BCL2 : A Relevant Immunohistochemical Marker for Triple Negative Breast Cancer Patients. *S. Zubeda[1,2,4,5], M. Khaliq Mohiuddin[2], K. Prabhakar Rao[1,5], Q. Hasan[2,3,4].* 1) Department of Genetics, Osmania University, Hyderabad, Telengana, India; 2) Department of Genetics, Vasavi Medical and research centre, Hyderabad, Telengana, India; 3) Department of Genetics, Kamineni Hospital, Hyderabad, Telengana, India; 4) Department of Genetics, Hyderabad Science Society, Hyderabad, Telengana, India; 5) RAS life sciences, Hyderabad, Telengana, India.

Breast cancer is a heterogeneous disease at histopathological, clinical and molecular level. Currently, the outcome or treatment response in patients with breast cancer are based on tumour size, N-status, histological grade and hormonal receptor status (ER/PgR) and Her-2. However, these conventional clinico-pathological characteristics are of limited value. More accurate prognostication markers are needed to select patients, who are most likely to benefit from treatment. Dysregulation of normal programmed cell death mechanisms play an important role in the pathogenesis and progression of breast cancer. BCL2, an anti-apoptotic gene that appears to block a distal step in an evolutionarily conserved pathway plays a crucial role in apoptosis. In this study BCL2- P2 Promoter methylation is evaluated by Methylation Specific Restriction Assay in 300 FFPE breast tissue samples, which included 100 each of fibroadenoma, breast cancer and adjacent normal breast tissue. Immunohistochemical expression and RNA transcripts levels of Bcl-2 were assessedin all three types of tissue (i. e fibroadenoma, breast cancer and adjacent normal) by using Rabbit polyclonal antibody specific to Bcl-2 (*Abcam, USA* ,Catalogue # *ab59348*) and by RT PCR respectively. 65% of BC tumour samples showed hypomethylation where as only 35% showed the expected 312 bp band indicating methylation of the BCL2 gene. While 67 % of adjacent normal breast tissues and 41% of benign tumours showed methylation of the same region. An overall mean of Bcl2 expression in BC tumour sections was 1. 40 ±0. 87, which is two times [P = 0. 002] higher than the adjacent normal tissue sections that had a mean of 0. 7424 ±0. 63 . BCL2 expression showed statistically significant association with triple negative [ER-, PR-, Her-2 -] breast cancer tissues with a mean of 1. 84 ±0. 8 [P = 0. 0006] , whereas the mean of Bcl2 expression in triple positive is only 1. 1 ±0. 7 . BCL2 transcript levels were found to have 5 fold up regulation in malignant tumours compared to benign tumours. Hence two fold increased expression of BCL2 in BC tumour tissues when compared to adjacent normal tissues and statistically significant difference in its expression in triple negative BC indicates that it is a relevant immunohistochemical marker for triple negative breast cancer. **Acknowledgements**I thank *UGC MANF – AND - MUS - 4126* for providing me the financial assistance for carrying out the work..

2762T

Thyroid nodules and multinodular goiter associated with germline mutations in *DICER1*. *N. Khan[1], L. Doros[2], K-A. Schultz[3], G. Williams[4], A. Bauer[5], L. Dehner[6], Y. Messinger[3], D. A. Hill[2], D. R. Stewart[1].* 1) National Cancer Institute, Rockville, MD; 2) Children's National Medical Center, Washington, D. C; 3) Children's Hospitals and Clinics of Minnesota, Minneapolis, MN; 4) International PPB Registry, Minnesapolis, MN; 5) Children's Hospital of Philadelphia, Philadelphia. PA; 6) Washington University, St. Louis, MO.

DICER1-syndrome is a recently recognized pleiotropic tumor predisposition syndrome. Multiple case-reports and case-series suggest nodular thyroid hyperplasia, non-toxic multinodular goiter (MNG), and possibly thyroid cancer are features of *DICER1*-syndrome. We collected self-reported histories of MNG and thyroidectomies from 76 cases with *DICER1*-syndrome recruited for history non-thyroid *DICER1*-syndrome-associated tumors, and 52 of their unaffected family members, and prospectively evaluated the cohort at the Clinical Center of the National Institutes of Health. The risk of diagnosed MNG was higher in both women (P=0. 0003) and men (P=0. 011). Counting thyroid surgery as a censoring event, the actuarial (lifetime) risks of diagnosed MNG in women and men with *DICER1*-syndrome were 29% and 11% by age 20 years, respectively, compared to 0% in controls of either sex. By age 40 years, the risks in cases were 80% and 19% among females and males, respectively, compared to 27% and 0% among female and male controls, respectively. In total, 22 out of 26 adult women with *DICER1*-syndrome reported a history of either MNG or thyroid surgery, compared to 5 out of 17 women without mutations (P<0. 001). In women without a history of MNG or thyroid surgery, more thyroid lesions were detected in those with *DICER1*-syndrome than in women without *DICER1*-syndrome. Similarly, a greater proportion of men with *DICER1*-syndrome reported a history of either MNG or thyroid surgery (p<0. 05), and more lesions were detected in cases (median:4; range:0-9) than were detected in controls (median:0; range:0-4) (P<0. 001). We used generalized estimating equations to test for differences in the size of lesions in cases (n=149) and controls (n=54). Lesions found in cases tended to be larger than those found in controls (P=0. 034). The composition of lesions (solid, cystic, partial cystic/solid, indeterminate) was not different between cases and controls. A total of 15 (10%) lesions in cases harbored calcifications compared to only one lesion (2%) in controls (P=0. 075). From 109 cases with 3,063 person-years of risk, three cases of thyroid cancer in the cohort yielded an observed/expected ratio of 15. 6 (95%CI: 3. 22-45. 6), though one case arose after administration of chemotherapy for PPB type II. We confirm in our cohort a high risk of diagnosed MNG, particularly in women and those between ages 15 and 25 years, and a modest increase in the risk of thyroid cancer.

2763W

Clinical Utility Evaluation of a Multiple-Gene Sequencing Assessment in Chinese Hereditary Breast Cancer. *X. Ye[1], X. Zhang[2], X. Wang[1], B. Xia[2], J. Xie[2], S. Zhu[1], J. Wang[1], D. Pang[2,3].* 1) BGI-Shenzhen, Shenzhen, Guang Dong, China; 2) Department of Breast Surgery, Harbin Medical University Cancer Hospital, China; 3) Translational Medicine Research and Cooperation Center of Northern China , Heilongjiang Academy of Medical Sciences, China.

Background Currently, dozens of susceptibility genes have been defined for cancer, including the most classic breast cancer susceptibility genes, BRCA1 and BRCA2, in hereditary breast cancer. Though several researches have proved the clinical utility of multiple-gene sequencing, instead of only BRCA1/2 test, in assessment of people with high hereditary cancer risk, little clinical approach is available. Especially, there is limited such research in China, which country counts for approximately one fifth of world population. Here, we use a multiple-gene sequencing assessment to identify the variations spectrum and evaluate the clinical utility of this assessment in Chinese cohort with high hereditary breast cancer risk. **Methodology/Principal finding:** Breast cancer patients, who meet the enrollment criteria of NCCN genetic or familial high-risk assessment guideline were informed and involved in this research. DNA was extracted from participants' blood. The coding regions of 115 genes, which are reported associated with the susceptibility or potential risk of 38 types of cancers were captured and sequenced using next generation sequencing platform. Single Nucleotide Polymorphism (SNP) and Insertion/deletion (Indel) variants were defined and analyzed. Pathogenic variants and potentially actionable results were informed to participants according to their willingness. Totally, 424 high-risk breast cancer patients participated in the research. From which, 89 patients harbored pathogenic germline variants, including 60 patients (14. 15%) carrying BRCA1, BRCA2 or PALB2 pathogenic variants, as well as 29 patients (6. 84%) carrying pathogenic variants in other genes. 250 VUS spots from 27 breast cancer susceptibility genes were identified in Chinese cohort and revealed a significantly different spectrum compared with Caucasian data. Additionally, 14 non-BRCA1/BRCA2/PALB2 gene pathogenic variant carriers and their relatives were followed and informed to recruit for further particular precancerous lesions screening. Positive precancerous lesion results have been observed in 3 families. **Conclusion** Overall, based on this research, we have revealed a comprehensive susceptibility variations spectrum and firstly evaluated clinical utility of a multiple-gene sequencing assessment in high hereditary breast cancer risk Chinese cohort.

2764T

Germline mutations in *RECQL* genes identified in high-risk breast cancer families. *F. Fostira[1], I. S. Vlachos[1,2], P. Apostolou[1], A. Vratimos[1], V. Mollaki[1], I. Konstanta[1], A. Vagena[1], M. Papamentzelopoulou[1], I. Konstantopoulou[1], D. Yannoukakos[1].* 1) Molecular Diagnostics Laboratory, National Centre for Scientific Research 'Demokritos', Athens, Greece; 2) DIANA-Lab, Department of Electrical & Computer Engineering, University of Thessaly Volos, Greece.

Even though extensive research has been performed in the field of breast cancer genetics using cutting-edge next generation sequencing (NGS) techniques, till date only a small number of candidate genes has been identified, most of which showing incomplete penetrance. Analysis of families with breast cancer aggregation and no mutations in the known breast cancer genes through exome sequencing has revealed a new promising gene: *RECQL1*. *RECQL1* is one of the five members in the RecQ helicases family, whichplay a critical role in DNA unwinding during DNA repair. Mutations in both alleles of *RECQL2, RECQL3* and *RECQL4* genes are known to cause Werner, Bloom and Rothmund-Thomson syndromes respectively, while a role in cancer susceptibility has been reported but remains unclear.

One hundred and ninety eight individuals (n=198) with early onset (<40 y) breast cancer, male breast cancer or having at least three breast and/or ovarian cancer cases within their family have been sequenced by NGS using the Illumina TruSight Cancer panel. It is a DNA capture-based targeted re-sequencing panel covering more than 1,700 coding exons of 94 cancer-related genes. All individuals have been previously tested negative for *BRCA1* & *BRCA2* mutations. All identified mutations have been validated with Sanger sequencing.

In total, four loss-of-function mutations were identified in RecQ family genes. Specifically, two mutations (c. 2205_2006insT & p. Q548X) have been identified in *RECQL3,* while one mutation has been detected in each of the*RECQL2* (p. R1406X) and *RECQL4* (c. 2300delG) genes. None of the mutations carriers had a detectable mutation in any other breast cancer predisposition gene including *BRCA1, BRCA2, CHEK2, PALB2, PTEN* and *TP53.*

Our findings suggest that genes within the RecQ family, other than *RECQL1,* can be potential breast cancer gene candidates. Further, large-scale genotyping will enable the determination of these genes' prevalence.

.

2765W

Whole exome sequencing enhances our understanding of pituitary adenoma. *K. Reddy[1], M. Korbonits[2], A. K. Madugundu[i], S. Sekhar[1], K. K. Mukerjee[3], A. Bhansali[3], P. Gupta[3], A. Rai[3], H. Gowda[1], R. K. Vashista[3], B. D. Radotra[3], A. Pandey[j], P. Datta[3].* 1) Institute of Bioinformatics, Bangalore, Karnataka, India; 2) Barts Health NHS Trust, London, UK; 3) PGIMER, Chandigarh, India.

Acromegaly(in adults)orgigantism(in child) isusually caused by a growth hormone-secreting pituitary adenoma. Apart from the well-known *GNAS* mutations, somatic sequence changes leading to somatotrophinoma tumorigenesis are unknown. We performed whole exome sequencing (WES) on DNA from 9 somatotrophinoma/paired peripheral blood sample and on 2 further somatotrophinoma samples (Illumina-HiSeq 2500)at an average read depth of 60X. We used a140 gene panel associated with acromegaly or pituitary adenomas to screen the exome. In a familial case of mother and daughter, the daughter had a NM_002072 (GNAQ):c. 286A>T(p. T96S) somatic variant in the adenoma but not in peripheral blood. In a second familial case of first cousins,one had somatic NM_001077488. 2(GNAS):c. 604C>T (p. Arg202Cys); the other cousin had a TSHoma and a somatic deletion disrupting a splice site previously-described as a pathogenic change in *MEN1.* A 3y patient had a heterozygote germline stop gain NM_003977. 2 (AIP):c. C241T(p. R81*) in *AIP*,which was hemizygous in the tumor sample. The tumor of a 7y patient had a novel low level somatic NM_001077490(-GNAS):c. 1639G>A(p. A547T) variant. A recurrent adenoma had a known pathogenic somatic substitution NM_020975. 4(RET):c. 2556C>G(p. I852M) (HGMD citationCM012799, RCV000021869), the paired blood or first tumor sample were not available for investigation. An acromegaly patient with Ph+ chronic myeloid leukemia (CML) had a pathogenic NM_020630(RET):c. 1118C>T(p. A373V) (HGMD citation CM065437) somatic *RET*variant. Three other cases had somatic*GNAS* variants: two were c. 604C>T and one was a novel variant NM_001077488(GNAS): c. 1120C>T(p. R374C). In conclusion, the majority of acromegaly samples (10/11, 90. 1%) harbored suspicious changes using WES. In our cohort, 54% (6/11) of adenomas had a *GNAS* somatic variant,of which 4/6 (67%) were the well described c. 604C>T variant. Interestingly, somatic *MEN1*, *GNAQ* and *RET* variants were also identified.

2766T

Challenges in disease gene identification in complex diseases: example from familial melanoma. *A. M. Goldstein[1], M. C. Fraser[1], H. Bennett[1], M. Rotunno[1], L. Burdett[1,2], A. Vogt[1,2], B. Hicks[1,2], K. Jones[1,2], S. J. Chanock[1], M. Yeager[1,2], X. R. Yang[1], M. A. Tucker[1], NCI DCEG Cancer Sequencing Working Group, NCI DCEG Cancer Genomics Research Laboratory.* 1) Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Bethesda, MD, USA; 2) Cancer Genomics Research Laboratory, Leidos Biomedical Research, Inc. , Frederick National Laboratory for Cancer Research, Frederick, MD, USA.

Although next generation sequencing (NGS) studies have resulted in the identification of a number of novel disease predisposition genes, disease gene discovery has been more limited for many cancers and other complex diseases. To investigate this issue further, we evaluated malignant melanoma patients from families with germline mutations in *CDKN2A/ARF*, *CDK4*, and *POT1* that are part of an ongoing single institution familial melanoma genetics study. Germline mutations in *CDKN2A*, the major known high-risk melanoma susceptibility gene, have been described in approximately 20% of familial melanoma kindreds. In contrast, germline mutations in *CDK4*, *ARF*, and *POT1*, occur in less than 1-2% of melanoma-prone families and are thus considered “rare” high-risk melanoma susceptibility genes. In our ongoing familial melanoma genetics study, we previously identified germline mutations in 28 melanoma-prone families including 23 in *CDKN2A*, 2 in *CDK4*, 1 in *POT1*, and 2 large deletions in *ARF*. Overall, 6% of melanoma patients (11/175) were sporadic, defined as not carrying their family’s high-risk melanoma-related mutation. Families with mutations in the rare melanoma predisposition genes (*CDK4*, *POT1*, and *ARF*) were significantly more likely to include sporadic patients (4/5) than families with *CDKN2A* mutations (4/23), p=0. 015. In addition, although early age at cancer diagnosis is a hallmark of genetic susceptibility, age at melanoma diagnosis in the sporadic patients (n=7) was below the United States median for 6 of the 8 families. Further, the sporadic patients in two families, both with mutations in rare melanoma predisposition genes, were among the youngest cases in their families with age at diagnosis six and fifteen years below the family median, respectively. Similarly, the sporadic patients were phenotypically indistinguishable from their mutation-carrying melanoma patient relatives. Therefore, selection of the youngest melanoma patient(s) for NGS or other types of genomic studies would have precluded disease gene identification in these families. Although based on a single genetics study at one institution, these findings illustrate some of the challenges in disease gene discovery in complex diseases. As investigators search for rarer disease predisposition genes, recognizing potential complexities of the diseases, studying multiple patients, and investigating larger numbers of families, most likely through collaborations, will be required.

2767W

Germline *BAP1* mutations misreported as somatic based on tumor-only testing. *M. H. Abdel-Rahman[1,2], K. Rai[2], C. M. Cebulla[1], R. Pilarski[2].* 1) Ophthalmology, The Ohio State University, Columbus, OH; 2) Division of Human Genetics, The Ohio State University, Columbus, OH.

Objective: There is an increasing interest in genomic-based management of tumors with several clinical laboratories offering molecular genetic testing of tumors using an extended panel of cancer genes as well as exome sequencing. Several of these laboratories utilize tumor tissue without matching germline DNA for their assessment. Results are based on in-house developed calling algorithms based on tumor cellularity and minor allele frequency (MAF) of the mutation. Here we present three patients reported with somatic mutations of *BAP1* in various tumors. Upon further assessment in our laboratory we determined these mutations to be germline heterozygous mutations. Methods: Germline *BAP1* mutations were evaluated by direct single-site sequencing of the reported somatic mutation in *BAP1* in germline DNA extracted from peripheral blood leukocytes. Cancer family histories of the patients were also collected. Results: Two of the mutations were reported by one laboratory while the third was reported by another laboratory. The three patients had strong personal and/or family histories suggestive of *BAP1*-Tumor Predisposition Syndrome. The tumor from the first patient was a metastatic adenocarcinoma of unknown origin; immunohistochemistry and imaging suggested a cholangiocarcinoma. Family history of pancreatic, ovarian, non-melanoma skin cancers, and mesothelioma was reported. The second patient presented with metastatic uveal melanoma with family history of uveal melanoma and a cancer of unknown origin. The third patient presented with personal history of invasive breast cancer and family history of mesothelioma, renal cell carcinoma, uveal melanoma, and several other cancers. Tumor tissue and germline DNA was tested in all three patients and heterozygous mutations were identified in both tumor and germline DNA in all cases. The somatic and germline mutation allelic ratios were similar in all cases (about 50:50). Conclusion: Sequencing of matching germline and tumor DNA is crucial for proper assessment of somatic versus germline mutation status. In patients with reported somatic mutation from laboratories carrying out tumor-only genomic testing, the possibility that a variant may be a germline mutation is warranted especially if personal and/or family history suggest hereditary cancer predisposition.

2768T

The association between RAD51 135G>C polymorphism and breast cancer risk. . *A. Mr. Alblihy.* Prince mohammad bin Naif medical centre, Riyadh, Saudi Arabia.

Breast cancer is the most frequent type of cancer diagnosed among women worldwide. Many dominantly inherited breast cancers result from mutations in the tumour suppressor genes *BRCA1* and *BRCA2* responsible for repairing DNA double strand breaks. A DNA double strand break (DSB) is the most dangerous type of DNA damage. *RAD51* located on chromosome 15q15. 1 and encodes for a protein plays a vital role in repairing DNA double strand breaks, particularly through a homologous recombination mechanism. A single nucleotide polymorphism in the 5' untranslated region of *RAD51*135G>C (rs1801320) has been identified to modify the risk of breast cancer among *BRCA1/2* mutation carriers. However, the results of studies that investigate the role of this polymorphism in the risk of breast cancer have been controversial. This report reviewed these studies in order to identify possible reasons for the discrepancies in results. It was found that there are many possible reasons could play roles in the inconclusive results. The effect of *RAD51* polymorphism on gene expression is still unclear. Thus, This report recommend to measure the gene expression in order to have a comprehensive understanding of the impact of this polymorphism in the *RAD51* function of homologous recombination (HR).

2769W

Genetics of Breast Cancer in a Highly Consanguineous Population, Challenges and lessons learned so far. *A. AlSaegh.* Genetics Department, Sultan Qaboos University, Sultanate of Oman.

Cancer genetics service was officially established in Sultanate of Oman in August 1st 2012 at the Genetics and Developmental Medicine Clinic in Sultan Qaboos University Hospital. Sultanate of Oman is situated in the southeast part of the Arabian Peninsula and it's characterized by high rate of consanguinity (52%) Individuals with moderate to high risk were referred mainly from the multidisciplinary breast clinic and other tertiary hospitals in the country. From August 2012 to May 2015, a total of 61 patients were referred, of which 36 individuals underwent molecular testing. 9/36 (25%) were found to haveBRCA mutations, 2/36(5. 6%) had PTEN mutations, 2/36 (5. 6%)had PMS2 mutations and 1/36 (2. 8%) had BRIP1 mutation . There were a number of challenges that were identified: lack of confirmatory dataon other affacted individuals of the family , concealing family history due to cultural believes, fear of stigma and some patients were treated outside the country . In addition, one major difficulty was selecting the right test when medical and pathological data was missing on other affacted members of the family. Conclusions: BRCA genes are to be considered in the Omani population. In some families, the predisposiposition to cancer was illustrating an autosomal recessive mode of inheritance, these families are to be approached differently toreveal their molecular etiology.

2770T

Next Generation Cancer Gene Testing in a Large HMO: the First 10 Months' Experience. *M. Alvarado[1], G. E. Tiller[1], J. Goff[1], R. Haque[2].* 1) Dept Genetics, Kaiser Permanente, Pasadena, CA; 2) Dept Research & Evaluation, Kaiser Permanente, Pasadena, CA.

Kaiser Permanente provides medical services to over 4 million members in southern California. Since July 2014, 1,343 patients (1,251 women and 92 men) who were referred to Genetics were tested via a multigene cancer panel, when the patient's clinical presentation and/or family history suggested the possibility of more than one cancer syndrome. Patient demographics: age 22-81 years; race/ethnicity: European 27%, Latino/Hispanic 25%, Asian 11%, African-American 7%, other 29%. The custom-designed panel included 20 cancer susceptibility genes: APC, ATM, BMPR1A, BRCA1, BRCA2, CDH1, CDKN2, CHEK2, EPCAM, MLH1, MSH2, MSH6, MUTYH, PALB2, PMS2, PTEN, SMAD4, STK11, TP53, and VHL. All tests were performed using NGS as well as CGH or MLPA. We identified 184 pathogenic/likely pathogenic mutations among 176 patients (13%) in the following genes: BRCA1 (42), BRCA2 (31), MUTYH (28; only one biallelic carrier), CHEK2 (22), ATM (18), PALB2 (13), MSH6 (6), TP53 (6), MSH2 (3), PMS2 (3), MLH1(2), CDKN2 (2) APC (2) VHL (2), and one mutation in each of the following genes: BMPR1, CDH1, and PTEN. Of the 176 patients with pathogenic mutations, 128 patients (73%) had both a personal and family history of cancer and/or polyps. Among the 176 patients with pathogenic mutations, 6 were found to carry more than one pathogenic mutation. No mutations were detected in 733 patients (55%); at least one variant of unknown clinical significance (VUS) was detected in 432 patients (32%); variants in ATM, APC, MSH6, BRCA2 and CHEK2 made up over 55% of all VUS detected. The overrepresentation of BRCA1 & 2 among all mutations identified (39%) likely reflected the significantly higher proportion of female patients referred to our department whose personal and/or family history included breast and/or ovarian cancer, compared to other cancer types. The identification of a significant percentage (37%) of dominant-acting gene mutations other than BRCA1 & 2 and those encoding MMR proteins provides further justification for offering a cancer gene panel to selected patients with a personal and/or family history of cancer. As more natural history data are compiled in patients with mutations in less well-characterized cancer genes, we anticipate a) expanding the scope of cancers associated with specific genes, b) encountering fewer VUS, and c) developing additional gene-specific surveillance protocols to enable early cancer detection in these high-risk patients and their family members.

2771W

Expression of A-kinase anchor protein 4 (AKAP4) in breast cancer. *S. Amini[1], E. Sakhinia[2], A. Abak[2].* 1) human genetic, tabriz university of medical science, tabriz, Iran; 2) biotechnology research center, tabriz, iran.

Introduction: Breast cancer is a common malignancy and a leading cause of cancer related death among woman. About 1. 3 million women are diagnosed with breast cancer each year over the world. A-Kinase anchor protein 4 (AKAP4) is a novel cancer testis antigen and it is a member of the AKAP family. AKAP4 is a part of protein kinase A super pathway. AKAPs are a group of proteins, which have function of binding to the regulatory subunit of protein kinase A. Early diagnosis contribute to improved survival rate and clinical outcome in breast cancer. AKAP4 gene are aberrant expressed by various malignancies but not by normal tissues with the exception of testis. The aim of this study was to investigate the association of AKAP4 with breast cancer and its relationship with the clinicopathological features of the patients. Materials and Methods: total RNA was isolated from 40 pairs of breast cancer tissues and adjacent non-cancerous tissues. CDNAs were synthesized. the expression level of AKAP4 gene was quantified by real time PCR. the correlation between the expression level of AKAP4 and clinicopathological features was studied. Results: AKAP4 was expressed in 87% (35/40) of breast cancer tissues but not in any of adjacent non cancerous tissues. and There was no significant relationship between AKAP4 mRNA expression and clinicopathological features of patients. Conclusions: This finding suggest that A-kinase anchor protein 4 can serve as a biomarker and novel prognostic indicator in breast cancer and may be a potential target for immunotherapy.

2772T

Systematic pan-cancer analysis of tumor purity. *D. Aran, M. Sirota, A. Butte.* UCSF, San Francisco, CA.

The tumor microenvironment describes the non-cancerous cells present in and around the tumor, including mainly immune cells, but also fibroblasts and cells that comprise supporting blood vessels. These non-cancer components of the tumor may play an important role in cancer biology. However, they also have a strong influence on the analysis of tumor samples by genomic approaches, and accordingly may alter the biological interpretation of the results of such analyses. Here we present a systematic analysis using different measurement modalities of tumor purity in more than 10,000 samples across 21 cancer types from The Cancer Genome Atlas (TCGA). We stratify patients according to clinical features to detect differences in purity levels, and demonstrate the confounding effect of tumor purity on correlating and clustering tumors with genomic data. Finally, using a differential expression method that accounts for tumor purity, we reveal an immunotherapy gene signature in several cancer types that cannot be detected by traditional differential expression analysis.

2773W

Large scale rare variation case-control studies. Comparing rare variation landscapes in uveal and cutaneous melanoma. *M. Artomov[1,4,6], I. Kim[2], A. Stratigos[3], M. Daly[1,4], H. Tsao[5].* 1) ATGU, Massachusetts General Hospital, Boston, MA; 2) Massachusetts Eye and Ear Infirm, Boston, MA; 3) Department of Dermatology, University of Athens Medical School, Athens, Greece; 4) Broad Institute, Cambridge, MA; 5) Department of Dermatology, Massachusetts General Hospital, Boston, MA; 6) Chemistry and Chemical Biology Department, Harvard University, Cambridge, MA.

Recent progress in sequencing data processing pipelines development allows creation of unprecedentedly large data sets. Here we report analysis involving almost 10000 full exome samples processed in a single batch using the most recent techniques for DNA variant calling. Ocular melanoma (OM) [OMIM 155720] is the most common primary intraocular cancer in adults. Unlike cutaneous melanoma (CM) [OMIM 155600], both environmental and genetic risk factors for OM are still largely unclear. It has long been recognized that OM and CM can co-exist in the same kindred though germline CDKN2A mutations, which occur in 10-40% of hereditary CM kindreds, account for a small fraction (<1%) of OM/CM families. Recently, a subset of hereditary OM families were found to harbor germline mutations in the BAP1 gene. Members of these BAP1 kindreds frequently develop CM, rather than OM, along with a host of other internal malignancies. These findings suggest that both ocular and cutaneous melanomas may share common risk pathways. Here we present our results of a large-scale rare variants association study. We assembled a full exome sequencing dataset of 400 cutaneous and ocular melanoma cases matched to 8000 non-cancer controls. We identified 19 loci that showed some evidence of association with OM risk though only BAP1 reached genome-wide significance. This analysis of an uncommon cancer phenotype (i. e. hereditary ocular melanoma) is important in establishing the dominance of BAP1 in the risk landscape of OM and in building a foundation for performing RVAS in the future. Literature reports that BAP1 protein is a member of highly cooperative BAP1 protein complex. Unlike in OM cases of CM do not show any enrichment of BAP1 mutations, however they show association signal for other members of BAP1 complex (ASXL1, ASXL2) allowing the discrimination between the two melanoma phenotypes based on the germline mutation landscape. As well this clarifies validity and value of rare variants association studies for various applications.

2774T

Report of two patients with unclassified variants in *BRCA2* gene. How to perform the genetic counseling? *M. D. F. Carvalho[1, 2,3, 4], K. M. Carvalho[1,2,3], E. D. F. Carvalho[1,2,3,4].* 1) Medical Genetics, Unichristus, Fortaleza, Ceara, Brazil; 2) Renorbio, Universidade Estadual do Ceará, Fortaleza, Ceara, Brazil; 3) Medical Genetics, Genpharma, Fortaleza, Ceara, Brazil; 4) Medical Genetics, APAE-CE, Fortaleza, Ceara, Brazil.

Hereditary breast and ovarian cancer syndrome, caused by a germline mutation in*BRCA1*or*BRCA2*, is characterized by an increased risk for breast, ovarian, prostatic and others cancers. An increased likelihood of a *BRCA1* or *BRCA2* mutation is suspected on the basis of certain personal and family history characteristics and various clinical criteria. The diagnosis is made by molecular genetic testing in an individual or family with a germline*BRCA1*or*BRCA2* mutation. DNA variants of uncertain significance (VUS) are common outcomes of clinical genetic testing for susceptibility to cancer. We report two patients with DNA VUS at *BRCA2* gene. Case 1 is a woman with a previous diagnosis of breast cancer before age 40. Her maternal grandmother and her father had breast and kidney cancer, respectively. The sequencing of the BRCA2 gene revealed the probably pathogenic heterozygous variant: c. 4804delA, not reported previously. This variant creates a frameshift in exon 11, resulting in a stop codon after 13 subsequent amino acids. There are no data in the literature regarding this particular variant, however, all frameshifts resulting in early stop codon in this exon are considered pathogenic causing loss of protein interaction domains. Case 2 is a woman without cancer diagnosis but with a family history of ovarian cancer in her mother and maternal haunt. She came to our ambulatory of medical genetics to perform genetic counseling. The sequencing of the BRCA2 gene revealed the heterozygous variant: c. A5096G, not reported previously. This is a rare variant with often less than 0. 1% of the population. Literature data are controversial as the role of variant in breast or ovarian cancer onset in people with personal or family history of breast cancer, it is not possible classifies it as pathogenic or benign. Bioinformatics analysis of this variant with pathogenicity prediction software and conservation show that the exchange occurred in the region not conserved among different species (low scores in PhyloP and GERP), and the SIFT and Polyphen 2 tools predict the exchange as not pathogenic. The LTR and Mutation Taster software ranked as neutral variant. Unclassified variants with unknown clinical significance in *BRCA1* and *BRCA2* genes pose a problem in genetic counseling, as their impact on risk of breast and ovarian cancer is still unclear. We encourage the development of international VUS-related guidelines for genetic counselors.

2775W

Characterization of Cycle-regulated genes in a cancer cell line by single cell genomics. *O. De la Cruz Cabrera[1], D. Mikkola[2,3], S. Edelheit[4], S. Markowitz[2,3].* 1) Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH; 2) Department of Medicine, Case Western Reserve University and Case Medical Center, Cleveland, OH; 3) Case Comprehensive Cancer Center, Case Western Reserve University and Case Medical Center, Cleveland, OH; 4) Department of Genetics and Genome Science, Case Western Reserve University, Cleveland, OH.

We used single-cell RNA-seq data to characterize cycle-regulated genes in a rapidly dividing cancer cell line derived from an aggressive liver metastasis of a colon cancer. Integration of the static information from the 70 individual cells into dynamic information about the cell cycle was achieved by use of the CRESCA (Cycle-regulated Expression via Single Cell Analysis) method. Single cell sequencing of mRNA was performed using microfluidic capture and paired-end sequencing of barcoded libraries. CRESCA uses tools from manifold regression and statistical learning to estimate the position of the each cell along the cell cycle. These positions, called artificial time stamps, are then used as a synthetic time-course experiment to scan all genes for cyclical expression. This powerful combination of methodologies enables analysis of cell cycle gene expression in populations of actively dividing cells. In particular, it holds great promise in the genetic dissection of proliferation in cancer.

2776T

The power of next generation sequencing in the detection of breast and ovarian cancer susceptibility genes other than *BRCA*. *M. Eliade[1], J. Skrypski[2], A. Baurand[1,2], C. Jacquot[1,2], G. Bertolone[1], C. Loustalot[2], C. Coutant[2], F. Guy[2], P. Fumoleau[2], J. -B. Rivière[1], Y. Duffourd[1], F. Ghiringhelli[2], F. Végran[2], R. Boidot[2], S. Lizard[2], L. Faivre[1,2].* 1) CHU Dijon, Dijon, Cote d'Or, France; 2) Centre Georges-François Leclerc, Dijon, France.

Background : Most cases of breast and ovarian cancer susceptibility remains unexplained. Testing multiple genes in one go with next generation sequencing is then an asset with the recent discovery of new genes involved in breast and ovarian cancer susceptibility. Methods : We studied 457 patients originated from Burgundy (France) fulfilling the criteria for *BRCA1* and *2* testing using a next generation sequencing 25-gene panel including 17 genes of predisposition for breast and/or ovarian cancer (*ATM, BARD1, BRCA 1/ 2, BRIP1, CHEK2, PALB2, RAD51C, TP53, PTEN, RAD50, MRE11, MLH1, MSH2, MSH6, PMS2, STK11*). Results : A pathogenic *BRCA1/2* mutation was found in 8% (n=37) of patients. Besides, we found 39 deleterious or probably deleterious mutations in 13 different genes. The most frequently mutated genes were *CHEK2* (n=10 ; 2. 1 %), *ATM* (n=9 ; 2 %), and *PALB2* (n=4; 0. 9%). One patient had deleterious mutations in both *TP53* and *PALB2,* and another one had deleterious mutations in both *BRCA2* and *CHEK2.* The mutation could explain the phenotype in the majority of cases, but a pathogenic mutation was found in a different predisposing gene in 7 patients, and could be considered as incidental findings with the currently published spectrum of cancer locations. Conclusion : Besides *BRCA1/2* mutations, that remain the most frequent susceptibility genes for breast and ovarian cancer, gene panels remain a powerful tool for identifying the other less frequent susceptibility genes. The penetrance and spectrum of cancer associated to these other genes remain sometimes undefined, and further collaborative work is crucially needed to address this question. The possibility of double hits should led to careful genetic counseling.

2777W

Mosaic loss of chromosome Y (LOY) in peripheral blood is associated with smoking, shorter survival and increased risk of cancer in men. *L. A. Forsberg[1, 2], C. Rasi[1, 2], N. Malmqvist[1, 2, 3], H. Davies[1, 2], S. Pasupulati[1, 2], G. Pakalapati[1, 2], J. Sandgren[4], T. Diaz de Ståhl[4], A. Zaghlool[1, 2], V. Giedraitis[5], L. Lannfelt[5], J. Score[6], N. C. P. Cross[6], D. Absher[7], E. Tiensuu Janson[3], C. M. Lindgren[8, 9], A. P. Morris[8], E. Ingelsson[2, 3], L. Lind[3], J. P. Dumanski[1, 2].* 1) Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden; 2) Science for Life Laboratory, Uppsala University, Uppsala, Sweden; 3) Department of Medical Sciences, Uppsala University, Uppsala, Sweden; 4) Department of Oncology-Pathology, Cancer Center Karolinska, Karolinska Institutet, Stockholm, Sweden; 5) Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden; 6) Faculty of Medicine, University of Southampton, Southampton, UK; 7) HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA; 8) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK; 9) Broad Institute of MIT and Harvard University, Cambridge, Massachusetts, USA.

We have shown that a male specific genetic risk factor, lifetime-acquired mosaic loss of chromosome Y (LOY) in non-cancerous blood cells, is associated with an increased risk of non-hematological tumors (Forsberg et al. 2014 Nature Genetics, PMID: 24777449). Analyzes performed in the Swedish ULSAM-cohort with >1100 participants analyzed on 2. 5M Illumina SNP-array indicated that LOY in peripheral blood could be associated with all-cause mortality (hazards ratio (HR) = 1. 91, 95% confidence interval (CI) = 1. 17–3. 13; 637 events) and non-hematological cancer mortality (HR = 3. 62, 95% CI = 1. 56–8. 41; 132 events). The median survival time after sampling in men affected with LOY was half, i. e. 5. 5 years shorter. The association of LOY with risk of all-cause mortality was validated in the independent PIVUS-cohort (HR = 3. 66, 95% CI = 1. 27–10. 54; 59 events). A frequent loss of the Y chromosome in the haematopoietic system was first described more than 50 years ago but its phenotypic consequences have been elusive. At least 15% of men above 70 years of age in our so far analyzed cohorts have LOY in a substantial degree of the blood cells. This discovery could help explain why males have a higher incidence and mortality from most sex-unspecific cancers. Smoking killed about 100 million people during the 20th century and is projected to kill one billion people during this century. We have recently shown (Dumanski et al. 2015 Science, PMID: 25477213), that having blood cells without a Y chromosome (LOY) is associated with smoking status using three independent cohorts (TwinGene: OR=4. 3, 95% CI =2. 8-6. 7; ULSAM: OR=2. 4, 95% CI=1. 6-3. 6; and PIVUS: OR=3. 5, 95% CI=1. 4-8. 4). A transient and dose-dependent mutagenic effect from smoking on LOY-status was observed. Lung cancer is the prime cause of cancer associated death in relation to smoking but it is less well appreciated that smoking also causes tumors outside the respiratory tract, which are predominant in men, and cumulatively roughly as common as lung cancer. The mutagenic effect from smoking on LOY could help explain why smoking seems a greater risk factor of cancer in smoking men compared to smoking women. It is well known that men have a shorter life expectancy compared to women in the entire word, but the underlying cause of this sex-difference was previously unknown. LOY in blood might help explain the shorter lives of men and could also become a new predictive biomarker of male carcinogenesis.

2778T

Germline PRKCD mutation linked to childhood EBV+ Hodgkin Lymphoma. *M. Ivanova[1], S. Lipinski[1], D. Loizou[2], R. Saadeh-Haddad[4], O. Alpan[2, 3], O. Goker-Alpan1[1, 2].* 1) Translation Medicine, Lysosomal & Rare disorders research treatment center, Fairfax, VA; 2) O and O Alpan, LLC, Fairfax, VA; 3) AmerImmune, Fairfax VA; 4) Medstar, Georgetown University Hospital, Washington, DC.

Hodgkin lymphoma (HL) comprises 6% of childhood cancers. While little is known about the molecular mechanisms, 40% of HL occurs after an infection by Epstein-Barr virus (EBV). It is suggested that B cells accumulate somatic mutations and DNA repair mechanism become nonfunctional. Chromosomal and genetic instability are typical features in the development of Hodgkin and Reed-Sternberg cells from B and, to a lesser extent, T cells. *Case report and results:* A five-year-old boy, born to non-consanguineous parents of Afghani descent, was evaluated for macrocephaly, developmental delay and hearing loss. At age 4, he was diagnosed with HL after the biopsy of a nasopharyngeal mass. He was (EBV+) and while he responded well to chemotherapy initially, he had recurrences in parallel with increasing EBV titers. Flow cytometric analysis of PBMCs showed the expansion of immature B (CD21 Dim 23. 7%) and absent memory B -CD20/CD27 cells(14, normal 22-134). B cells immunoglobulin expression were decreased, G (0. 5%) (normal 1. 4-12. 5%) and A (0. 3%) (1. 1-6. 0%), indicating a class switch defect. 98. 3% of B cells were CD5+ dim (normal average 1-9%). T helper cells showed slight lymphopenia (26%, normal 26-57%). The patient had normal NK and NKT populations (4. 6 and 3. 5%) (normal 3. 4-24% and 0. 6-8. 2%). Whole exome sequencing identified a homozygous **G3149D** variant of unknown significance in *PRKDC* (NM_006904. 6). EBV selectively inhibits CD5+ cells, and thus can stay latent by preventing apoptosis of the transformed cells. B cell impairments and CD5+ cell proliferation could be attributed to alterations in *PRKDC,* a DNA-dependent protein kinase catalytic subunit, acting as a molecular sensor for DNA damage, and a scaffold protein to deliver repair proteins to the site of DNA damage. While missense *PRKDC* mutations are associated with Severe Combined Immune Deficiency, with absent B and T cells, and additional dysmorphic features, CNS structural abnormalities, the somatic mutation G3149D is listed as a factor contributing to oncogenesis. GGC>GAC (c. 9446 G>A) in exon 68 leads to a replacement of glycine by aspartic acid, likely to impact the secondary protein structure and compromise the protein-protein binding ability by changing the polarity and charge. While *PRKDC* knockout models are used to study lymphoproliferative disorders and EBV infection, this is the first case in humans linking *PRKDC* to EBV associated *HL.* .

2779W

Non-random occurrence and early age of onset of diverse lymphoid cancers in families supports the existence of genetic risk factors for multiple lymphoid cancers. *S. Jones[1,2], J. Voong[3], R. Thomas[2], A. English[2], J. Schuetz[2], G. W. Slack[4], J. M. Connors[4], A. Brooks-Wilson[2,5].* 1) Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada; 2) Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver BC; 3) Department of Statistics and Actuarial Science, Simon Fraser University, Vancouver BC; 4) Centre for Lymphoid Cancers, BC Cancer Agency, Vancouver BC; 5) Department of Biomedical Physiology and Kinesiology, Simon Fraser University, Vancouver BC.

Lymphoid cancers are a biologically diverse group of neoplasms, yet clustering of different lymphoproliferative disorders within families has been observed, suggesting the existence of shared genetic risk factors. We have established, in Metro Vancouver, British Columbia, Canada, a collection of 140 families with 2 or more members with lymphoid cancer. We hypothesized that lymphoid cancers arising in families as a result of genetic factors that predispose to multiple types of lymphoid cancers would have: 1) an earlier median age of onset, and 2) a different pattern of co-occurrence in families than that predicted based on population frequencies. The median age at diagnosis was 9, 9, 6 and 10 years earlier for Hodgkin lymphoma (HL), non-Hodgkin lymphoma (NHL), myeloma (MM) and chronic lymphocytic leukemia (CLL) cases in these families, respectively, than for SEER cases (Howlander et. al. , 2011, SEER cancer statistics review). This corresponds to the familial HL, NHL, MM and CLL cases being at the 32nd, 29th, 30th and 25th percentile for age of onset. We also observed earlier age of onset in later generations, a phenomenon referred to as anticipation. Apparent anticipation was seen in NHL ($P < 0.0001$), HD ($P < 0.0001$), CLL ($P < 0.0013$) and all lymphoid cancers together ($P < 0.0001$). Canadian population frequencies of HL, NHL, MM and CLL were used to derive expected frequencies for pairwise combinations of lymphoid cancer types. Observed pairwise frequencies in the families were compared to expected frequencies and significance of the difference determined by permutation. Familial co-occurrences differed from those expected based on population frequencies ($P < 0.0001$). The difference is accentuated in families with more affected members (2 cases, $P < 0.0001$; 3 cases, $P < 0.0001$; 4 or more cases, $P < 0.0001$). Non-random patterns of co-occurrence, and earlier age of onset of familial cases, underscore a difference between these familial cases versus population cases, which are mainly sporadic. These observations support the idea that lymphoid cancer families are not all collections of sporadic cases that arose in the same family by chance. These observations provide evidence for the existence of inherited genetic factors that significantly increase the risk of developing several types of lymphoid cancers, and justify the use of family-based genomic methods to identify lymphoid cancer susceptibility genes.

2780T

Feasibility of cascade screening in a randomly selected sample of families at risk for hereditary breast cancer from a statewide cancer registry. *M. C. Katapodi[1,2], D. Duquette[3], K. Mendelsohn-Victor[2], B. Anderson[3], J. J. Jang[2], L. L Northouse[2], S. Duffy[4], K. J. Milliron[5], D. Ronis[2], S. D. Merajver[5,6], N. K. Janz[7], S. J. Roberts[7].* 1) Institute of Nursing Science, University of Basel, Basel, Basel-Stadt, Switzerland; 2) University of Michigan School of Nursing, Ann Arbor, MI, USA; 3) Michigan Department of Health and Human Services, Lansing, MI, USA; 4) Ohio State University School of Nursing, Columbus, OH; 5) University of Michigan Comprehensive Cancer Center, Ann Arbor, MI; 6) University of Michigan School of Medicine, Ann Arbor, MI; 7) University of Michigan School of Public Health, Ann Arbor, MI.

Genetic counseling is recommended for women diagnosed with breast cancer at a young age and their close relatives. Cascade screening is a new public health intervention that promotes systematic familial testing amongrelatives of individuals with known mutations. In an effort to assess the feasibility of statewide cascade screening, we examined the willingness of young breast cancer survivors (YBCS) to invite their high-risk relatives in such a program. A randomized trial aiming to increase awareness of cancer genetics, identified a random sample of 3000 YBCS, diagnosed 25-45 years old, and stratified by race (Black/Other), through the Michigan Cancer Surveillance Program. YBCS' responses helped identify high-risk, unaffected, female relatives. A linear regression analysis, guided by the Theory of Planned Behavior, examined baseline demographic and clinical characteristics, access and attitudes towards healthcare services, family history, and family environment as predictors of the proportion of eligible relatives YBCS were willing to contact (number willing to contact/ number of all high-risk relatives). Responses from 859 eligible YBCS (33%) helped identify 1842 high-risk relatives. YBCS were willing to contact 1271 (69%) relatives. Response rate among invited relatives was 52%. The study recruited a final sample of n=431 eligible high-risk relatives with a mean age of 43±12 years old. Willingness to contact a greater number of relatives (b=26. 03), and higher perceived risk for another breast cancer (b=0. 82) were associated with contacting a greater proportion of relatives. Having a larger family (b=-18. 92) was negatively associated with the proportion of relatives contacted. Relatives were more likely to be invited in the study by the YBCS if they were living withing 50 miles distance. Most YBCS were willing to contact relatives for cascade screening, supporting the feasibility of this approach. Results suggest barriers and facilitators in disseminating information about hereditary breast/ovarian cancer among families. Recruitment from a statewide cancer registry helped target a large number of YBCS, with multiple "red flags" in their personal and family health history, and should be considered as a systematic approach for trials involving hereditary and familiar forms of common cancers.

2781W

Retaining familial cancer cases and families in a genetic research study on lung cancer. *D. Mandal[1], A. Bencaz[1], J. Chambliss[1], J. Thompson[1], J. E. Bailey-Wilson[2].* 1) Genetics, LSU Health Sciences Center, New Orleans, LA; 2) NHGRI/NIH, Baltimore, MD.

About 10% of lung cancer cases (22,000 cases per year in the U. S.) have at least one first-degree relative affected with lung cancer, and 25% of cases at least one first- or second-degree affected relative, indicating that family history is a significant risk factor. In order to identify susceptibility genes for lung cancer, a genetic research study was initiated in southern Louisiana in 1996. Families with 3 or more cases of lung cancer were enrolled for the genetic linkage study to identify rare variants with large effects. Affected individuals and unaffected relatives of individuals who were informative for genetic linkage analysis were recruited. In 2008, we also initiated recruiting familial cases with at least two cases of lung cancer in families to identify common variants with small effects. Recruitment for familial lung cancer study is challenging due to the poor prognosis once diagnosed. In addition, the family involvement required for linkage studies makes enrollment of complete families even more challenging. The objective of the present study is to report various retention strategies used in our research study that can be utilized in the retention of any familial cancer cases/families. Annual follow up phone calls and other communication methods were implemented to maintain retention of study subjects, since family health history status may change over the years and it is important to update any such information in genetic studies. To date, 2147 individuals with lung cancer have been interviewed and 742 of them reported having a family history of lung cancer. Of them, 19 high-risk families (3 or more affected cases per family) for the linkage study have been recruited. In addition, 147 familial cases (2 or more affected cases per family) have been recruited for the genome-wide association study. Identification of susceptibility genes for familial lung cancer is underway using whole genome linkage analysis, genome wide association studies and exome sequencing using next generation strategies in collaboration with GELCC (Genetic Epidemiology of Lung Cancer Consortium), which is the only familial lung cancer research group in the world. Retention strategies used to receive continuous participation of lung cancer cases/families will be discussed in detail.

2782T

A prospective study of mitochondrial DNA copy number and the risk of prostate cancer. *A. Moore[1], Q. Lan[1], J. N. Hofmann[1], C. S. Liu[2], W. L. Cheng[2], T. T. Lin[2], S. I. Berndt[1].* 1) National Cancer Institute, Bethesda, MD; 2) Vascular and Genomic Research Center, Changhua Christian Hospital, Taiwan.

Mitochondrial DNA (mtDNA) copy number may increase in response to DNA damage as a compensatory mechanism. Increased mtDNA copy number has been observed in prostate cancer (PCa) cells, suggesting a role in PCa development, but this association has not yet been investigated prospectively. We conducted a nested case-control study (793 cases and 790 controls) of men randomized to the screening arm of the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial (PLCO) to evaluate the association between pre-diagnosis mtDNA copy number, measured in peripheral blood leukocytes, and the risk of PCa. We used logistic regression to estimate odds ratios (OR) and 95% confidence intervals (CI) and polytomous logistic regression to analyze differences in associations with non-aggressive (Stage I/II AND Gleason grade <8) and aggressive (Stage III/IV OR Gleason grade ≥8) PCa. Although mtDNA copy number was not significantly associated with PCa odds overall (per 1-unit increase on the natural log scale, OR = 1. 23, 95% CI 0. 97-1. 55, p = 0. 089), increasing mtDNA copy number was associated with an increased odds of non-aggressive PCa (OR = 1. 29, 95% CI = 1. 01-1. 65, p = 0. 044) but not aggressive PCa (OR = 1. 02, 95% CI = 0. 64-1. 63, p = 0. 933) compared to controls. We report that increasing mtDNA copy number is associated with an increased odds of non-aggressive PCa, suggesting that high leukocyte mtDNA copy number is a biomarker of risk associated with non-aggressive PCa development but not advanced PCa.

2783W

Exploring the regulatory roles of common variants associated with lung cancer subtypes. *T. O'Brien[1,2], P. Jia[2], Z. Zhao[2,3,4].* 1) Vanderbilt Genetics Institute, Vanderbilt University School of Medicine, Nashville, TN; 2) Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN; 3) Department of Psychiatry, Vanderbilt University School of Medicine, Nashville, TN; 4) Department of Cancer Biology, Vanderbilt University School of Medicine, Nashville, TN.

Lung cancer is classified into two main types: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). NSCLC has several subtypes, the two most common being lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC). These three subtypes of lung cancer have distinct physiological, clinical, and genetic signatures. Previous research has identified common variants (i. e. SNPs) associated with lung cancer subtypes; however, these common variants are often located in the non-coding regions and their functional roles in each subtype remain largely unknown. We hypothesize that these non-coding SNPs function in common or distinct regulatory mechanisms for each subtype. Thus, we explored the potential regulatory roles for common variants found to be associated with each lung cancer subtype from genome-wide association studies (GWAS) of European samples. We first obtained the SNPs that are associated with each lung cancer subtype by $p < 10\text{-}5$. Then, we expanded this list to include all SNPs in linkage disequilibrium (LD) with $r2 > 0.8$ using the 1000 Genomes Project data. This expansion resulted in 1427, 1788, and 1059 SNPs for LUAD, LUSC, and SCLC, respectively. Next, we searched for enrichment of the lung cancer SNPs in expression quantitative trait loci (eQTLs) for lung and an additional eight tissues using the data from the Genotype-Tissue Expression (GTEx) Project. Although we found that several SNPs were eQTLs, there was no statistical enrichment compared to random SNPs selected from the GWAS platform. This lack of enrichment in eQTLs was confirmed in another lung tissue-specific eQTL dataset with a larger sample size. We further searched for regulatory enrichment using the data from five lung-related tissues and cell lines generated in the Roadmap Epigenomics Project. We found very distinct regulatory signatures that differed by lung cancer subtype and lung tissue/cell type, such as SNPs within enhancer regions in LUSC and SCLC in adult lung tissue, but not in LUAD. We also found similarities among subtypes such as the highest proportion of SNPs located in an active state identified by transcription signatures at the 3' and 5' end of genes (the "TxFlnk" class in Roadmap Epigenomics) in the A549 lung cancer cell line. These preliminary results suggest that genomic regions containing common variants found in three major subtypes of lung cancer may share both common and distinct regulatory chromatin states.

2784T

Correlation between BRAF and MLH1 hypermethylation testing in a Lynch syndrome universal screening program. *L. H. Rodgers[1], K. M. Shannon[1], A. J. Iafrate[2], G. Y. Lauwers[2], A. R. Mattia[2, 3], D. C. Chung[1].* 1) Center for Cancer Risk Assessment, Mass General Hospital, Boston, MA; 2) Department of Pathology, Mass General Hospital, Boston, MA; 3) Department of Pathology, North Shore Medical Center, Danvers, MA.

Universal MSI testing of all colorectal cancers is an effective strategy to detect new cases of Lynch syndrome (LS). Immunohistochemical (IHC) staining for the mismatch repair (MMR) genes MLH1, MSH2, MSH6 and PMS2 can be performed on all resected colorectal cancer specimens. The absence of one or more MMR proteins raises the suspicion of LS and further risk assessment is indicated. However, there is a significant subset of cases with MLH1 loss due to somatic MLH1 promoter hypermethylation. These cases are considered sporadic and do not require genetic evaluation. The presence of a BRAF mutation has been proposed as a surrogate marker for somatic MLH1 hypermethylation. We sought to determine the performance characteristics of BRAF mutation testing in comparison to MLH1 promoter hypermethylation testing in a large universal screening program established at our institution. In tumors in which MLH1 protein was absent by IHC, BRAF mutation analysis (V600E) and MLH1 promoter hypermethylation testing were both performed. Of the 1011 colorectal tumors that underwent IHC testing between January 2011 and December 2014, 148 (15%) exhibited loss of the MLH1 protein. 124 specimens were available for BRAF/MLH1 hypermethylation reflex testing. 84 (68%) were positive for MLH1 hypermethylation and a BRAF mutation and these were thus classified as sporadic tumors. Sixteen (13%) were negative for both the BRAF mutation and MLH1 hypermethylation and those individuals were invited for further genetic workup of suspected LS. The remaining 24 tumors (19%) exhibited discordant results: 23 samples did not exhibit the BRAF mutation but were MLH1 hypermethylated, while 1 exhibited the BRAF mutation but was not MLH1 hypermethylated. Although all of these 24 cases were invited in for additional risk assessment, only 9 (38%) participated in genetic counseling. The average age at diagnosis was 66 years (range 50-87), 0 met Amsterdam II criteria, and none had more than one LS associated tumor. 6/9 underwent germline genetic testing and no pathogenic mutations were identified. In conclusion, BRAF testing correlates with MLH1 hypermethylation results in 81% of colon cancer cases with MLH1 loss by IHC. Using BRAF testing as a sole strategy would result in excess referrals for genetic counseling and testing. No germline MLH1 mutations were identified in the group with discordant BRAF/MLH1 hypermethylation results but the rate of uptake for genetic counseling was low.

2785W

GSK3β is a pivotal regulator in cervical cancer development and a potential target of therapeutics. *A. Saeidian.* Department of Molecular and Human Genetics, Banaras Hindu University, Varanasi, Uttar Pradesh , India.

PI3K/AKT pathway is a crucial regulator for many cellular processes such as cell proliferation, cell survival, apoptosis, and cell cycle regulation and its deregulation leads to the growth of tumor. In this study, our objectives were to investigate the role of AKT1/GSK3b signaling in cell cycle regulation, especially the role of GSK3b, cyclin D1 and p27 in cervical cancer progression and we assessed the effect of drug SB216761 alone and in combination with MG132 on cervical cancer. Our western blots data showed the over expression of AKT1 and p-AKT1 (Thr-308) in both primary cervical tumor and in cervical cancer cell lines. Similarly, RT-PCR and western blotting results were indicative of the up-regulation of GSK3b at transcript level and the p-GSK3b (Ser-9) at protein level along with up-regulation of cyclin D1 protein that were suggestive of their critical role in cell cycle progression and cervical cancer development. Also, our results showed that the up regulation of GSK3b, cyclin D1 and inactivation of GSK3b through phosphorylation at Ser-9 via AKT1 signaling axis is important in cervical cancer development. Furthermore, our western blot analysis demonstrated the down-regulation of p27KIP1 and p-Cyclin D1 that leads to G1 progression of the cell cycle in cervical cancer. The treatment of cell lines such as SiHa, Hela, CaSKi, Me-180 in specific dose-and time-dependent manner with SB216761 alone and together with SB216763+MG132 induces GSK3b inactivation and also leads to rapid degradation of cyclin D1 through GSK3b as an important mediator in cervical cancer cell lines. Similarly, knockdown of GSK3b with siRNA resulted in the degradation of GSK3b and phospho-GSK3b (p-Thr-286). Inhibited GSK3b induced proteasomal degradation of cyclin D1 by reducing Thr286-phosphorylated cyclin D1 levels. FACS result also showed inhibited progression of cell cycle from G0/G1 phase to S phase in the presence of the particular dosage of MG132+SB2167663 which indicate the degradation of cyclin D1. Drug induced cyclin D1 ablation inhibits cancer cell proliferation and thus presents an important promising therapeutic strategy for treating and/or preventing cervical cancer.

2786T

Validation of copy number variations in hereditary prostate cancer families using Droplet Digital PCR. *K. Wood[1], R. Gambhira[2], E. Ledet[2], O. Sartor[2], D. Mandal[1].* 1) Department of Genetics, Louisiana State University Health Sciences Center, New Orleans, LA; 2) Tulane Cancer Center, New Orleans, LA.

Prostate cancer (PCa) is the second leading cause of death in American men, and approximately 27,540 men will die from the disease in the US this year. Family history is one of the most prominent predictors of PCa risk, and approximately 10% of PCa cases are attributable to inheritable genetic factors. However, identifying PCa genetic susceptibility has been extremely difficult due to genetic heterogeneity. Genomic copy number variations (CNVs) have been detected in prostate tumors and several other cancers, but changes in copy number have not been studied extensively in germ-line DNAs from hereditary PCa (HPC) cases. In our preliminary study, we have identified CNVs in germline DNAs from 7 HPC families of European ancestry by using array comparative genomic hybridization (aCGH) technology. The regions identified were 2q22, 11q22, and 16q23, and these regions were found to be duplicated in all 7 of the HPC cases analyzed. In addition, a 68 kb duplication was observed at 2q14 in all 7 HPC cases of European ancestry, and this region was further identified by using linkage analysis in 4 of the 7 HPC families (heterogeneity LOD-score = 1. 94). The goal of the present study is to validate the CNVs identified by aCGH using other quantitative methods in our HPC families. Droplet Digital Polymerase Chain Reaction (ddPCR) provides absolute quantification of target DNA copies. This cutting edge technology can measure small fold differences in target DNA copy numbers. In the literature, the region 2q14 has been reported to harbor structural variations in prostate cancer tumors. We are validating the duplication observed within this region by performing ddPCR in the germline DNA samples from the HPC families. Each ddPCR reaction uses duplex TaqMan assay reagents for the target region and the reference gene. Analysis of ddPCR data will be conducted using QuantaSoft Analysis software (Bio-rad). The absolute quantification of copy number differences using ddPCR will provide fresh insight into HPC families. Validation of the variation in copy numbers and analysis of our results in HPC families using other affected and unaffected male family members will be presented.

2787W

Multi-gene panel testing in breast, ovarian, and pancreatic cancer cases: prevalence and unintended screening recommendations. *E. L. Young[1], D. W. Neklason[1,2], A. K. Snow[1], B. A. Thompson[1], M. A. Firpo[7], S. J. Mulvihill[7], T. L. Werner[8], R. Bell[1], K. R. Smith[5], A. Fraser[1], W. K. Kohlmann[1], L. A. Neumayer[6], L. A. Cannon-Albright[2,3], S. V. Tavtigian[1,4].* 1) Huntsman Cancer Institute, University of Utah; 2) Division of Genetic Epidemiology, Department of Medicine, University of Utah; 3) Research and Development Service, George E. Wahlen Department of Veterans Affairs Medical Center; 4) Department of Oncological Sciences, University of Utah School of Medicine; 5) Population Sciences, Huntsman Cancer Institute, University of Utah; 6) Department of Surgery and Arizona Cancer Center, University of Arizona, Tucson; 7) Department of Surgery, University of Utah School of Medicine; 8) Division of Oncology, Department of Medicine, University of Utah.

Multi-gene testing allows for the identification of variants from a list of genes in a single test. Multi-gene testing for hereditary breast and ovarian cancer (HBOC) is beneficial in instances where the suspected predisposition gene is ambiguous or when multiple variants are segregating in a pedigree. Results from multi-gene testing are used as rationale for increased screening or other preventive measures in individuals carrying pathogenic variants and their biological relatives. Consideration for HBOC germline testing is currently based predominately on family history criteria, which limits the availability of genetic testing of women with limited or unknown family histories. In an effort to expand the criterion for genetic testing, we examined the prevalence of medically actionable variants in breast (n=55), ovarian (n=25), and pancreatic (n=66) cancer cases, unselected for family history, with a custom 34-gene panel consisting of genes known or suspected to underlie familial cancer risk, including the pathways involved in DNA homologous recombination repair (HRR), mismatch repair (MMR), etc. We found that 12. 7% of our breast cancer cases, 36% of our ovarian cancer cases, and 13. 6% of our pancreatic cancer cases carried either a known pathogenic variant or a variant of uncertain significance (VUS) predicted to be deleterious based on priors found in publically available databases: InSIGHT for the MMR genes, or the HCI BRCA1/2 Prior Probabilities Database for BRCA1 and BRCA2. When such a database was not available, an in silico consensus prediction from CADD and Polyphen 2 was used. Of the pathogenic or predicted deleterious VUS findings, 72% occurred in genes with established screening recommendations. The combination of our ascertainment procedure and mutation screening results now put us in the position to use the Utah Population Database (UPDB), a comprehensive population-based resource linking genealogies with medical records, to ask whether the relatives of mutation-carrying ovarian cancer or pancreatic cancer cases are at increased risk of a screen detectable cancer. The answer to this question has important implications for the clinical utility of applying panel testing to a wide spectrum of incident cancer cases.

2788T

A new method to estimate clonal composition of solid tumors from whole whole-genome copy number profiles. *L. Wu[1], K. Sakai[2], K. Ueshima[3], M. Ukita[4], M. Mandai[4], M. Kudo[3], A. Roter[1], K. Nishio[2], J. Schmidt[1].* 1) Affymetrix Inc. , Santa Clara, CA; 2) Department of Genome Biology, Kinki University Faculty of Medicine; 3) Department of Gastroenterology and Hepatology, Kinki University Faculty of Medicine; 4) Department of Obstetrics and Gynecology, Kinki University Faculty of Medicine.

Background: Estimation of clonal composition or degree of heterogeneity in solid tumors is a challenging and important problem that has been linked to treatment choices and clinical outcome. The problem is particularly difficult because samples are typically admixtures of both aberrant tumor cells and normal tissue cells. We propose a novel method to estimate clonal composition in solid tumors from whole-genome copy number profiles obtained from OncoScan® FFPE Assay Kit from Affymetrix. Methods: Data from OncoScan FFPE Kit provides genome-wide segmentation, where each segment is associated with a Log2ratio (Log2R) (=Log2([observed signal]/[normal signal])) and B-allele frequency (BAF). Due to the admixture, Log2R is a function of the percentage of aberrant tumor cells (%AC) that contribute to the aberration and the true copy number (CN) in the cancer genome. The computation of BAF uses the notion of mA and TA (number of minor and total alleles at a heterozygous site, respectively). BAF of a segment is defined as mA/TA for the heterozygous sites within the segment (e. g. , for a segment with CN = 1, mA = 0, and TA = 1). Due to admixture, the interpretation of BAF values is more complicated as they too are a function of %AC. A key observation is that for fixed mA and %AC, Log2R is a linear function of log2(BAF), denoted by Log2B, and hence the (Log2B,Log2R) values of segments with same %AC and mA align in a straight line. Values corresponding to mA = 0 are of particular interest as they are sparse and well separated for different %AC. The (Log2B,Log2R) values of the segments of a cancer sample are clustered via an iterative clustering approach and mapped onto a 2D plot labeled with the theoretical (Log2B,Log2R) points and their associated mA and %AC values. Results: We processed more than 100 samples from a variety of cancer tissues and mapped them using this method. A significant number, including samples with hundreds of segments, clearly fall into 1–3 distinct %AC points, with each %AC corresponding to a distinct subclone. The method borrows many of the ideas brought forward by previous approaches, such as Genome Alternation Print (GAP) and Allele-Specific Copy Number Analysis of Tumors (ASCAT), but it provides two clear advances: 1) transparent and informative graphic presentation of clonality; 2) visual and analytical estimation of the clonality of a tumor sample.

2789W

Cost-effectiveness of routine screening for Lynch syndrome in endometrial cancer patients up to 70 years of age. *A. Goverde[1,2], M. C. W. Spaander[2], H. C. van Doorn[3], H. J. Dubbink[4], A. M. W. van den Ouweland[1], C. M. Tops[5], S. G. Kooi[6], J. de Waard[7], R. F. Hoedemaker[8], M. J. Bruno[2], R. M. W. Hofstra[1], E. W. de Bekker-Grob[9], W. N. M. Dinjens[4], E. W. Steyerberg[9], A. Wagner[1] on behalf of the LIMO study group.* 1) Dept. of Clinical Genetics, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands; 2) Dept. of Gastroenterology and Hepatology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands; 3) Dept. of Gynaecology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands; 4) Dept. of Pathology, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands; 5) Dept. of Clinical Genetics, Leiden University Medical Center, Leiden, Netherlands; 6) Dept. of Gynaecology, Albert Schweitzer Hospital, Dordrecht, Netherlands; 7) Dept. of Gynaecology, Sint Franciscus Gasthuis, Rotterdam, Netherlands; 8) Pathology laboratory Pathan, Rotterdam, Netherlands; 9) Dept. of Public Health, Erasmus MC, University Medical Center Rotterdam, Rotterdam, Netherlands.

Introduction Lynch syndrome (LS) is a hereditary cancer syndrome characterized by early onset of colorectal cancer (CRC), endometrial cancer (EC) and other cancers. Identifying LS carriers is important, since surveillance programs can reduce morbidity and mortality. Although routine screening for LS is recommended among CRC patients ≤70 years of age, guidelines for LS screening among EC patients remain ambiguous. We aimed to assess the cost-effectiveness of routine molecular screening for LS in EC patients ≤70 years of age. **Methods** Consecutive EC patients ≤70 years of age were routinely screened for LS by analysis of microsatellite instability, immunohistochemistry and *MLH1* hypermethylation, followed by germline mutation analysis if indicated. Life years gained (LYG) were based on the number of LS carriers detected among EC patients and their relatives. Total costs consisted of LS diagnostics and LS surveillance, including gynaecological surveillance and prophylactic surgery in relatives. We calculated the incremental cost-effectiveness ratios (ICERs) for LS screening among EC patients ≤70 years compared with those ≤50 years of age and compared with LS screening according to the revised Bethesda guidelines. **Results** Screening of 179 EC patients identified 7 LS carriers, only one of whom was ≤50 years of age. Genetic testing of relatives identified 27 LS carriers. In total, benefit from LS screening amounted to 26,5 LYG; 16,3 LYG for LS screening among EC patients ≤50 years and 10,2 LYG for screening EC patients 51-70 years of age. Total costs resulting from LS screening increased from kλ154 (λ8,125 per LS carrier detected) for screening among EC patients ≤50 years of age to kλ308 (λ9,073 per LS carrier detected) for screening among EC patients ≤70 years of age. The ICER for LS screening in EC patients ≤70 years compared with ≤50 years was λ15,056/LYG. The revised Bethesda guidelines identified 3/7 (43%) LS carriers among EC patients and 22/27 (81%) LS carriers among relatives. The ICER for routine LS screening in EC patients ≤70 years of age compared with the revised Bethesda guidelines was λ19,204/LYG. Both ICERs remained <λ25,000/LYG (<$30,000/LYG) in sensitivity analyses. **Conclusion** Routine screening for LS by analysis of microsatellite instability, immunohistochemistry and *MLH1* hypermethylation in EC patients ≤70 years is a cost-effective strategy compared with screening those ≤50 years of age and screening according to the revised Bethesda guidelines.

2790T

Genotype-phenotype analysis of Von Hippel Lindau syndrome in Korean families. *JS. Lee[1], JH. Lee[3], KE. Lee[2], JH. Kim[3], JM. Hong[1], EK. Ra[1], SH. Seo[1], MJ. Kim[1], SS. Park[1], MW. Seong[1].* 1) Departments of Laboratory Medicine,Seoul National University College of Medicine, Seoul, Korea; 2) Departments of Surgery, Seoul National University College of Medicine, Seoul, Korea; 3) Departments of Internal Medicine, Seoul National University College of Medicine, Seoul, Korea.

Background: Von Hippel-Lindau(VHL) disease is a rare hereditary tumor syndrome, caused by *VHL* gene mutations and characterized by heterogeneous phenotypes: benign/malignant tumors of CNS, retina, kidney, adrenal gland and pancreas. The genotype-phenotype correlation has been studied in different ethnic groups. However, it has not been studied well in Korean population so far. The aim of this study is to evaluate Korean *VHL* mutation spectrum and genotype-phenotype correlations. **Methods:** 13 unrelated subects with *VHL* mutations were included. After *VHL* mutation was studied, clinical manifestations and family histories were consequently evaluated. **Results:** We identified 10 different *VHL* mutations. 2 frameshift mutations were novel: c. 592delC and c. 160_161delAT. Missense mutations were clustered in two domains(α in exon 1 and ß in exon 3). CNS hemangioblastoma(CHB)(n=10), Retinal hemangioblastoma(n=6), Renal cell carcinoma(n=5) and pheochromocytoma(n=2) were present in 13 subjects. Mild phenotypes were observed in possibly *de novo* patients(n=5)(*P*=0. 007). Our unique finding is that age specific risk for CHB is significantly higher in subjects carrying mutations within HIF-alpha binding site(*P*<0. 05). **Conclusions:** We report identified that HIF-a binding site mutations elevate the age-related risk of CHB. The results of this study also provide Korean VHL mutation spectrum and its correlation with phenotypes.

2791W

Majority of familial colorectal cancer unexplained by known cancer susceptibility genes. *B. A. Thompson[1], D. W Neklason[1,2], A. K. Snow[1], E. L Young[1], R. Bell[1], N. R. Sargent[1], W. K. Kohlmann[1], L. A. Cannon-Albright[2,3], S. V. Tavtigian[1,4].* 1) Huntsman Cancer Institute, Salt Lake City, UT; 2) Division of Genetic Epidemiology, Department of Internal Medicine, University of Utah, Salt Lake City, UT; 3) Research and Development Service, George E. Wahlen Department of Veterans Affairs Medical Center, Salt Lake City, UT; 4) Department of Oncological Sciences, University of Utah School of Medicine, Salt Lake City, UT.

It is estimated that ~30% of all colorectal cancer (CRC) cases diagnosed have a familial basis, ~5% of which are caused by high-risk germline mutations in *APC*, *MUTYH* and the DNA mismatch repair (MMR) genes (*MLH1*, *MSH2*, *MSH6*, and *PMS2*). Identifying new genetic risk factors will likely contribute to our understanding of CRC susceptibility, initiation, and progression. Using germline DNA from 52 first or second-degree cousin target pairs from high-risk CRC families (n=101), we performed targeted capture sequencing with a 33-gene cancer susceptibility panel. The family collections from which the targets pairs are drawn had been sporadically screened for MMR gene mutations. We evaluated the germline alterations using multiple filters to categorize them as pathogenic, variants of uncertain significance (VUS), or likely benign. Likely pathogenic sequence variants were then tested for segregation in the families in which they were observed. There were 294 variants identified in the 33 genes, three of which are pathogenic mutations in high-risk genes. These included two *MSH2* truncating mutations: one found in both members of a cousin pair, while the other was found in a single case from a target pair. The third was a *BRCA2* frameshift mutation that segregated with breast cancer in a different branch of the family than that represented by the CRC target pair. Additionally, four monoallelic *MUTYH* mutation carriers and a case with the moderate risk *CHEK2* c. 1100delC variant were identified. There were 69 VUS, six of which were shared by a target pair (*BARD1*, *RAD50*, *PKHD1*, *PMS2*, *BRCA1*, *BRIP1*). In this screen of target cousin pairs with unexplained CRC, we identified two high-risk CRC, one high-risk breast cancer, and one moderate risk breast cancer susceptibility variants. Furthermore, at least two of the large CRC families in this study could have mutations in multiple genes contributing to cancer in different branches. Inactivating mutations in other known/novel cancer susceptibility genes may cause the CRC in the majority of these families.

2792T

TruSeq®FFPE DNA Library Prep, a reliable method for preparing high quality Illumina® Whole Genome Sequencing (WGS) libraries from challenging FFPE samples. *R. Sanches-Kuiper, S. MacArthur, A. Jackson, T. James, C. Haskins, H. Sansom, V. Smith.* Illumina Ltd, Chesterford Research Park, North Saffron Walden, Essex, United Kingdom.

Formalin-fixed, paraffin embedded (FFPE) samples are one of the most important sample types used in large scale Whole Genome Sequencing (WGS) studies, particularly those focused on cancer. Simple, flexible and reliable methods for the preparation of high quality WGS libraries from these samples are therefore required. Performing WGS studies using FFPE samples is challenging for a number of reasons: (i) DNA quality is highly variable, and challenging to predict based only on sample age and tissue type; (ii) DNA is often highly fragmented, limiting library yield, complexity and useful read-length; (iii) Chemically modified bases are prevalent, inhibiting library construction processes and producing false positives; (iv) Aggressive DNA extraction methods can result in coverage bias & alignment artifacts. The challenge is increased further because the quantity of FFPE DNA available is often limited, and because there is a requirement for libraries of high complexity to support deep sequencing to detect low-frequency somatic variants. Many methods currently used for the preparation of WGS libraries from FFPE samples fail to address the issues identified above and produce libraries with substantial compositional bias, artifacts which produce many false-positive variant calls and/or insufficient complexity to support deep sequencing. To enable routine use of FFPE DNA libraries in large-scale WGS studies we have developed an easy to use method, TruSeq® FFPE DNA Library Prep, that address the issues highlighted above. Here we show that TruSeq® FFPE DNA Library Prep enables reliable generation of high quality WGS libraries, supporting deep sequencing (>60x coverage) of tumor samples, from as little as 200ng of DNA. Sample quality is determined before the sample is committed using a qPCR-based QC assay. This enables DNA input to be adjusted for maximum assay performance and/or maximum preservation of scarce material. The method consists of several steps designed to maximize library complexity, minimize FFPE-specific artifacts and produce even coverage. Furthermore, we demonstrate that the method can be used to perform deep WGS of some tumor samples using as little as 20ng of input material. Genome wide analysis of copy number variation can be performed using as little as 1ng DNA. Additionally, when comparing TruSeq® FFPE DNA Library Prep with other commercially available protocols we found that it produces substantially fewer FFPE-specific false positive variant calls.

2793W

Using a Costello Syndrome cohort as a syndromic model to better understand embryonal rhabdomyosarcoma at the molecular, cytogenetic, and functional level. *K. M. Robbins[1,3], D. L. Stabley[1], J. Holbrook[1], R. Sahraoui[1,3], K. W. Gripp[2], K. Sol-Church[1].* 1) Center for Pediatric Research, Al duPont Hospital for Children Wilmington , DE; 2) Medical Genetics, Al duPont Hospital for Children, Wilmington, DE; 3) University of Delaware Biological Sciences, Newark, DE.

Costello Syndrome (CS) is a rare disorder, and patients have failure-to-thrive, coarse facial features, short stature, skeletal abnormalities, and a range of intellectual disabilities. CS is the result of heterozygous germline mutations in the proto-oncogene *HRAS*. CS patients have a predisposition for malignancies, specifically embryonal rhabdomyosarcoma (ERMS). We have one of the largest cohorts of Costello Syndrome patients in the world. Therefore, we have the unique opportunity to use this cohort as a syndromic model to study ERMS. ERMS is one of the most common pediatric soft-tissue malignancies in the general population. These sporadic ERMS tumors have rarely been associated with *HRAS* mutations, and instead are characterized by loss of heterozygosity (LOH) at 11p15. 5. We studied syndromic ERMS, in our unique cohort of CS patients, in addition to sporadic ERMS tumors and cell lines. We have ten ERMS tumors from seven unrelated CS patients. In addition, we have three sporadic ERMS cell lines and eight FFPE tumor samples. The goal of this study was to further characterize ERMS at the molecular, cytogenetic, and functional levels. First, we molecularly characterized the RAS mutation status in all ERMS through PCR and Sanger sequencing. Next, we cytogenetically characterized all ERMS using chromosome 11 microsatellite markers. Finally, we aimed to identify important signaling pathways involved in ERMS tumorigenesis. Differential gene expression analysis was performed using Affymetrix Human 1. 0 ST arrays. CS and sporadic ERMS arrays were normalized and compared to 6 CS patient derived fibroblast lines and 6 control fibroblast lines in order to identify ERMS specific genes. Real-time PCR and western blot analysis validated the ERMS specific genes/pathways. We discovered a higher frequency of RAS mutations in our sporadic ERMS cohort than previously reported. An unexpected novel finding was complete LOH of chromosome 11 due to uniparental disomy for the paternal allele within our syndromic and sporadic ERMS cohort. We determined ERMS specific genes contributing to tumorigenesis. This study demonstrated that using CS as a syndromic model for ERMS uncovered novel findings furthering our understanding of ERMS tumorigenesis. This knowledge will ultimately guide drug development and precision medicine based drug selection.

2794T

Variant of Unknown Significance Rates Vary by Ethnicity and Genes Analyzed. *L. Panos, J. Thompson, V. Speare, J. Dolinsky, K. Panchani, H. LaDuca.* Ambry Genetics, Aliso Viejo, CA.

As the uptake of multi-gene panel testing (MGPT) continues to increase worldwide, a primary concern of clinicians and patients is the possibility of variants of unknown significance (VUS). The VUS rate increases with expanding genomic content and ethnic diversity. We aim to define the likelihood of identifying a VUS on MGPT across five major ethnic groups: Caucasian, Ashkenazi Jewish, Asian, Hispanic, and African American. A retrospective review of over 40,000 patients undergoing MGPT for hereditary predisposition to cancer was performed at a single diagnostic laboratory. Ethnicity was indicated on test requisition forms. The rate of identifying one or more VUS was calculated for 10 different MGPTs ranging in size from 5-42 genes. Individuals with one or more VUS, including those who carried a pathogenic or likely pathogenic mutation were included. The number of individuals within each ethnic group varied from 1,873 Asians to 37,151 Caucasians. The likelihood of a VUS was consistently lowest for Caucasian and Ashkenazi Jewish individuals across all panels (Table 1). Asians showed the highest MGPT VUS rates on all panels except PancNext. The VUS rate for Lynch syndrome genes (*MLH1, MSH2, MSH6, PMS2, EPCAM*) across all panels was 6. 9% in Caucasians and 16. 4% in Asians. In *BRCA1* and *BRCA2*, the VUS rate for Asians was 5. 7%, while the VUS rate for Caucasians was 2. 1%. Our results demonstrate that VUS rates for MGPT vary based on the patient's ethnic background and the number of genes analyzed. The specific genes analyzed also play a role, as genes differ in size and some tolerate more variability than others. For example, BreastNext has higher overall VUS rates than RenalNext despite both panels containing 17 genes. Availability of population frequency data also contributes to VUS rate variation by ethnicity, although these gaps will continue to improve with the use of public population frequency databases such as the Exome Aggregation Consortium (ExAC).

Table 1: VUS Rates (%) by Ethnicity and MGPT

MGPT	# of Genes	Caucasian	African American	Ashkenazi Jewish	Asian	Hispanic
BRCAplus	5	4. 4	8. 9	2. 6	13. 7	8. 0
GYNplus	9	10. 8	17. 0	15. 6	24. 5	15. 2
PGLNext	10	10. 9	33. 3	0. 0	28. 6	16. 0
PancNext	13	18. 0	39. 4	17. 1	27. 8	29. 4
BreastNext	17	21. 5	37. 1	24. 7	42. 2	29. 0
ColoNext	14	16. 1	21. 6	16. 6	35. 7	29. 6
RenalNext	17	18. 0	30. 0	16. 7	47. 6	27. 3
OvaNext	23	27. 4	43. 0	27. 6	54. 3	37. 2
CancerNext	28	29. 9	48. 3	33. 1	59. 1	36. 8
CancerNext-**Expanded**	42	39. 4	57. 1	41. 7	74. 4	52. 7

2795W

Phenotypic and molecular characterization of the SCA28 knockin mouse model harboring the *Afg3l2* p. M665R mutation. *C. Mancini[1], E. Hoxha[2,3], L. Iommarini[4], C. Cagnoli[1], E. Turco[5], F. Altruda[5], E. Giorgio[1], A. Brussino[1], S. Cavalieri[6], E. Di Gregorio[6], G. Gasparre[7], A. M. Porcelli[4], F. Tempia[2,3], A. Brusco[1,6].* 1) Department of Medical Sciences, University of Torino, Torino, Italy; 2) Neuroscience Intitute Cavalieri Ottolenghi; 3) Deptartment of Neuroscience, University of Torino (Italy); 4) Department of Pharmacy and Biotechnologies (FABIT), University of Bologna, Bologna, Italy; 5) Department of Molecular Biotechnology and Health Sciences, University of Torino, Torino, Italy; 6) Medical Genetics Unit, "Città della Salute e della Scienza" Hospital, Torino, Italy; 7) Department Medical and Surgical Sciences, Medical Genetics, University of Bologna, Bologna, Italy.

SpinoCerebellar Ataxia type 28 (SCA28, OMIM#610246) is one of the 41 known subtypes of autosomal dominant cerebellar ataxias (ADCA). SCA28 is caused by mutations in the*AFG3L2*gene, encoding for m-AAA metalloprotease (mitochondrial ATPases Associated with a variety of cellular Activities). Most data regarding SCA28 pathogenetic mechanism rely on yeast complementation experiments and on studies of*Afg3l2*-knockout heterozygous mice. Almost all known*AFG3L2*mutations are missense in exon 16, and their effect as haploinsufficiency or dominantnegative mechanism is debated. We generated a knockin (KI) mouse carrying the c. 1993A>G, p. M665R mutation (corresponding to the human p. M666R SCA28 mutation). Homozygous KI mice died perinatally; heterozygous KI mice showed a significant motor impairment in the beam balance test (P<0. 05) at 18 months. In the repeated Rotarod test their performance was indistinguishable from wildtype littermates. Morphological analysis revealed a preserved cytoarchitecture of cerebellar lobules, with unaltered thickness of the molecular and granular layers and a conserved number of Purkinje cells (PC). Electron microscopy showed moderate mitochondrial damage in PC. The spontaneous action potential firing of PC was significantly increased (P<0. 05), measured-*invitro*in cell-attached configuration. We evaluated mitochondrial morphology and function, studying both homo- and heterozygous Mouse Embryonic Fibroblasts (MEF). We showed an impairment in mitochondrial dynamics, detected as the complete loss of long OPA1 forms in homozygous KI MEF, with a fragmented mitochondrial network (MitoTracker Red staining). Homozygous KI MEF showed also a reduced uptake of the MitoTracker probe that might be suggestive of a decreased mitochondrial membrane potential. Bioenergetics analysis revealeda 25% reduction of Complex III activity in homozygous KI (P<0. 05), detected by spectrophotometric analysis of the respiratory chain. Homo-, heterozygous and wildtype MEF showed comparable citrate synthase activity, and the SDS–PAGE of complex subunits at steady-state level appeared normal: these data indicate that a compensatory event that increase the mitochondrial mass is unlikely. These results suggest that PC of SCA28-KI mice at 18 months display functional alterations associated with an initial mitochondrial damage, and that mutations hitting the peptidase M-41 domain negatively impact on m-AAA complex function, probably acting as hypomorphic.

2796T

***Rai1* haploinsufficient mice exhibit abnormal social behavior.** *K. Walz, C. Abad, N. Rao, I. C. Perez, J. I. Young.* Hussman Inst Human Genomics, Univ Miami, Miami, FL.

Smith-Magenis syndrome (SMS) is a genetic condition associated with a deletion of ~ 3. 7Mb in chromosome 17p11. 2. SMS is characterized by neurobehavioral abnormalities such as intellectual disability, developmental delay, sleep disturbance and self-injurious behaviors. Recently, SMS patients were reported to score within the autism spectrum range for social responsiveness and communication. Mutations in a single gene, the Retinoic Acid Induced 1 (*RAI1*), are found in ~30% of SMS patients. The objective of this study is to assess autistic features in the *Rai1*+/- mouse model in an attempt to determine whether *Rai1* is the dosage sensitive gene responsible for abnormal social interactions in SMS. We evaluated adult *Rai1*+/- mice (in a *C57BL/6-Tyrc-Brd* genetic background, *Rai1*+/- N=12; wildtype N=15) with a set of tests that inform on social behaviors: self-grooming in a novel environment (repetitive behavior), three chamber test (sociability and preference for social novelty), tube test (social behavior), and olfactory habituation/dishabituation (communication). All animal testing were approved by the IACUC at the Miller School of Medicine, UM. *Rai1*+/- mice displayed a significant increase in their preference for social novelty when compared to wildtype (p<0. 05). The tube test paradigm, confirmed the social behavior abnormalities of the *Rai1*+/- mice; when confronted to wildtype mice, they backed out of the tube significantly more times than expected by chance (83% of the times). No significant differences were observed between *Rai1*+/- and wild type mice in response to non-social or social odors, suggesting that the identification of social cues was not affected in the mutant mice. Repetitive behaviors were not augmented in the *Rai1*+/- mice. These data showing abnormal responses of the *Rai1*+/- mice in the three-chamber and tube tests suggest that Rai1 is involved in the regulation of a specific domain of social behavior.

2797F

Assessing Behavior and Anxiety in the *Dhcr7*$^{\Delta 3\text{-}5/T93M}$ mouse model of Smith-Lemli-Opitz Syndrome. *J. L. Cross, M. F Keil, F. D Porter.* NICHD, National Institutes of Health, Bethesda, MD.

Smith-Lemli-Opitz Syndrome (SLOS) is an autosomal recessive inborn error of cholesterol synthesis caused by mutation of the 7-dehydrocholesterol reductase (*DHCR7*) gene. This results in abnormal sterol levels; increased 7-dehydrocholesterol and typically decreased cholesterol. Although SLOS has a characteristic physical phenotype, with the most common finding being 2-3 toe syndactyly, there are also multiple behavioral abnormalities. These include cognitive deficits, anxiety, hyper-activity, sleep cycle disturbance, language impairment and autism spectrum behaviors. There are currently two main mouse models of SLOS; a homozygous null, *Dhcr7Δ3-5/Δ3-5*, and a model combining the null mutation and the common p. T93M missense mutation, *Dhcr7Δ3-5/T93M.* Unlike the null model, the hypomorphic *Dhcr7Δ3-5/T93M* mice can live to adulthood and are therefore suitable for behavioral studies. Two frequently used behavioral protocols, studying burrowing and nesting behavior, can be used to assess hippocampal function as it has been shown that hippocampal-lesioned mice engage in less burrowing and nest building than healthy controls. The burrowing test can additionally indicate anxiety as rodents exhibit defensive behavior by kicking the contents of the burrow to deflect intruders. In a pilot study, 3 to 5 month old *Dhcr7Δ3-5/T93M* mice were compared to age-matched controls for these two paradigms. It was observed that, on average, *Dhcr7Δ3-5/T93M* mice burrowed and engaged in less nest building than controls. Some variation was present but as only 75% of *Dhcr7Δ3-5/T93M* are found to have communicating hydrocephalus, structural analysis of the brain upon completion of the behavioral testing period may explain this variation. Future study will include use of the Elevated Plus Maze and Open Field protocols for further analysis of anxiety. These studies will provide insight into the mechanisms driving the behavioral phenotype in SLOS.

2798W

Mouse with substitution of type I collagen 3-hydroxylation site has altered ECM but does not recapitulate the bone dysplasia of types VII/VIII Osteogenesis Imperfecta. *J. C. Marini[1], N. Fratzl-Zelman[2], J. E. Perosky[3], A. Alimasa[3], R. Harris[3], P. S. Backlund[4], P. Roschger[2], K. Klaushofer[2], A. Forlino[5], K. M. Kozloff[3], W. A. Cabral[1].* 1) Bone and Extracellular Matrix Branch, NICHD/NIH, Bethesda, MD; 2) Ludwig Boltzmann Institute of Osteology at Hanusch Hospital of WGKK and AUVA Trauma Centre Meidling, 1st Med. Dept. Hanusch Hospital, Vienna, Austria; 3) Orthopaedic Research Laboratories, Department of Orthopaedic Surgery, University of Michigan, Ann Arbor, MI; 4) Biomedical Mass Spectrometry Facility, NICHD, NIH, Bethesda, MD; 5) Department of Molecular Medicine, Biochemistry Unit, University of Pavia, Pavia, Italy.

Recessive types VII and VIII osteogenesis imperfecta (OI) are severe bone dysplasias caused by null mutations in the genes that encode prolyl 3-hydroxylase 1 (*P3H1*) or cartilage-associated protein (*CRTAP*), two mutually supportive components of the prolyl 3-hydroxylase (P3H) complex expressed in bone and cartilage. Individuals with types VII or VIII OI lack both the P3H complex components and post-translational hydroxylation of its collagen substrate at residue α1(I)P986. This primary modification defect leads to increased post-translational lysine hydroxylation of the type I collagen helix, consistent with delayed folding and loss of complex chaperone function. Several potential functions of the substrate modification have been proposed, including modulating collagen helical stability, fine-tuning the alignment of collagen monomers within fibrils, and regulating bone matrix mineralization. We therefore sought to clarify the role of the P986 substrate modification in bone dysplasia and collagen overmodification by generating a knock-in mouse model with an α1(I)P986A substitution that cannot be 3-hydroxylated. Exclusive expression of α1(I)A986 collagen in murine-derived tissues and cell transcripts and protein was confirmed by cDNA sequencing and mass spectrometry. Neither heterozygous (986P/A) nor homozygous (986A/A) mice recapitulate critical features of types VII and VIII OI. Mutant mice exhibited normal growth rates, femoral mechanical properties (stiffness, ultimate load, brittleness) and collagen folding kinetics. However, the P986A substitution affected collagen biochemistry, matrix organization and mineralization. Despite normal P3H complex levels in cell lysates, 986P/A and 986A/A osteoblast type I collagen was moderately overmodified on PAGE analysis. Dermal fibrils of 986A/A displayed decreased diameters and heterogeneity. Skeletal staining and radiographs revealed flared rib cages, delayed calvarial mineralization and kyphosis by 2 months in 986A/A pups. Interestingly, although femoral aBMD and TMD was reduced in 986A/A versus WT femora, bone matrix mineral density distribution assessed by qBEI was normal. Together, these data suggest that 3-hydroxylation of α1(I)P986 is important for regulating type I collagen modification, crosslinking and mineral organization in bone, but does not cause the severe bone pathology of collagen 3-hydroxylation defects, which likely result from absence of the ER complex and cartilage α1(II)P986 3-hydroxylation.

2799T

Defects of lipid synthesis underlie the age dependent demyelination caused by lamin B1 over expression. *Q. S. Padiath[1], H. Rolyan[1], Y. Tyurina[2], M. Hernandez[3], A. A. Amoscato[2], L. J. Sparvero[2], B. S. Nmezi[1], Y. Lu[4], M. R. H. Estécio[4], K. Lin[4], J. Chen[1], R. He[2], P. Gong[2], L. H. Rigatti[5], J. Dupree[6], H. Bayir[2], V. E. Kagan[2], P. Casaccia[3].* 1) Dept. of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Environmental and Occupational Health, University of Pittsburgh, Pittsburgh, PA; 3) Department of Neuroscience, The Icahn School of Medicine at Mount Sinai, New York, NY; 4) Department of Epigenetics and Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Houston, TX; 5) Division of Laboratory Animal Resources, University of Pittsburgh School of Medicine, Pittsburgh, PA; 6) Department of Anatomy and Neurobiology, Virginia Commonwealth University, Richmond, VA.

Lamin B1 is a component of the nuclear lamina and plays a critical role in maintaining nuclear architecture, regulating gene expression and modulating chromatin positioning. We have previously shown that *LMNB1* gene duplications cause Autosomal Dominant Leukodystrophy (ADLD), a fatal adult onset demyelinating disease. The mechanisms by which increased Lamin B1 levels cause ADLD are unclear. To address this, we studied a transgenic mouse model over-expressing Lamin B1. These mice showed severe age-dependent vacuolar degeneration of the spinal cord white matter together with secondary axonal damage and loss of neuronal cell bodies. Oligodendrocytes in the transgenic mice revealed alterations in histone modifications favoring a transcriptionally repressed state. Chromatin changes were accompanied by reduced expression of genes primarily involved in lipid synthesis pathways, many of which are known to play important roles in myelin regulation. Decreased lipogenic gene expression resulted in a significant reduction in multiple classes of lipids involved in myelin formation. Many of these gene expression changes and lipid alterations were observed even before the onset of the phenotype, suggesting a causal role. Our results have important implications for ADLD disease pathogenesis and establish, for the first time, a link between Lamin B1 and lipid synthesis in oligodendrocytes. They provide a mechanistic framework to explain the age dependence and white matter involvement of the disease phenotype and sheds light on a critical role for the nuclear lamina at the intersection of pathways involving aging,epigenetics,demeylination, and lipid metabolism.

2800F

Knockout CERKL gene causes retinal degeneration in zebrafish. *S. Yu, M. Liu.* Huazhong University of Science and Technology, Wuhan, China.

AbstractCERKL mutations are associated with severe retinal degeneration. Despite several CERKL knockout mice as well as CERKL knockdown zebrafish have been used to characterize the underlying mechanism in pathophysiology for this retinal disease, the function of CERKL remains unknown. In this study, we generated a CERKL knockout zebrafish model by TALEN technology. The targeted CERKL deletion resulted in progressive degeneration of both rod and cone photoreceptor outer segments (OS) and increased apoptosis of retinal cells, accompanying with oxidative damage. The electroretinogram assay for the CERKL-/- zebrafish showed that photoreceptor function was affected as early as 5-7 days before photoreceptor degeneration. Several key proteins involving in phototransduction cascade, including rhodopsin, GNAT1, PDE6B and GRK1, were found to decrease in various degrees in the CERKL-null fishes before the onset of photoreceptor degeneration. Collectively, our results showed that knockout of CERKL in zebrafish caused photoreceptor degeneration, with significant decreased expression of photoreceptor cascade proteins. These findings would help to understand the molecular pathogenesis of retinal degeneration caused by CERKL mutations.

2801W

A mouse modifier study using the Nphp10 (Sdccag8$^{Tn(sb-Tyr)2161B. CA1Cove}$) model to identify a modifier locus of ciliopathy-related phenotypes. *K. Weihbrecht[1,2], W. A. Goar[1,2], C. C. Searby[1,3], T. E. Scheetz[2], V. C. Sheffield[1,3], S. Seo[2].* 1) Department of Pediatrics, Univ. of Iowa; 2) Department of Ophthalmology, Univ. of Iowa; 3) Howard Hughes Medical Institute, Univ. of Iowa.

Serologically defined colon cancer antigen 8 (SDCCAG8), later defined as NPHP10, is a nephronophthisis-related ciliopathy gene, mutations in which cause Senior-Loken syndrome and Bardet Biedl syndrome. The mouse model, Sdccag8Tn(sb-Tyr)2161B. CA1COve, was generated by a Sleeping Beauty Transposon (SBT) insertion between exons 12 and 13 of Nphp10 leading to a deletion of Nphp10 exons 13-18 as well as exons 2-13 of the neighboring gene, Akt3. This mouse model was originally available on the FVB/NJ background and homozygous mutant mice (Nphp10 -/-) show a phenotype of death before P1 at 100% penetrance. Previous studies have shown that some genes show differential expression on different backgrounds and other studies have shown that background strain can have an effect on some phenotypes. For example, Bbs4 knockout animals are found at weaning near Mendelian ratios (25%) on a 129/SvEvTac background, but survival of knockouts on a FVB/NJ background is greatly decreased (5%). The purpose of this study was to determine if a modifier of Nphp10 -/- survival could be identified. Pure strain FVB/NJ Nphp10 +/- mice were crossed with 129/SvEvTac Nphp10 +/+ to generate Nphp10 heterozygotes on a 50% mixed background. These mice were then intercrossed and DNA was collected from Nphp10 -/- mice. Twenty samples were submitted to Jackson Lab for a custom 150 SNP panel: 10 survivors and 10 non-survivors. Chi-Square analysis with Bonferroni correction identified two SNPs of interest on chromosome 11: rs3023266 (p=0. 018) and rs3714172 (p=0. 030). The distance between these two SNPs is 15 Mb and contains hundreds of genes. However, multiple additional SNPs fall within this region that we were able to use to narrow the region of interest. We performed sequencing of these two SNPS plus SNPs falling between them on the original 20 samples plus an additional 26 survivors and 2 non-survivors (final ratio of 3:1 survivors to non-survivors) to determine if the region of interest could be narrowed. The addition of sequence data from these SNPs has narrowed our region of interest to 1. 2 Mb between rs3023266 and rs26968358, which contains ~30 genes of both known and unknown function. A few of these genes are of interest because of their role in cell death, such as tumor necrosis factor superfamily, member 13b (Tnfrsf13b) and ubiquitin specific peptidase 22 (Usp22). Further SNP analysis and additional crosses will allow us to determine a smaller region containing a specific modifier gene.

2802T

Interaction of *MAB21L2* with BMP signaling and *PAX6* in ocular development. *B. Deml[1,2], S. Muheisen[2], E. Semina[1,2].* 1) Cell Biology, Neurobiology and Anatomy, Medical College of Wisconsin, Milwaukee, WI; 2) Pediatrics and Children's Research Institute, Children's Hospital of Wisconsin, Milwaukee, WI.

Coloboma is defined as the incomplete closure of the optic fissure and is often associated with additional ocular or systemic anomalies. Previously our lab identified a pathogenic variant, p. (Arg51Gly), in*MAB21L2*[MIM 604357]in a family with dominant ocular coloboma, microcornea, cataracts and skeletal dysplasia. To investigate this further, we created zebrafish*mab21l2*genetic lines carrying either a frameshift truncation (*mab21l2Q48Sfs*5*) or an in-frame deletion (*mab21l2R51_F52del*) allele and demonstrated the presence of variable severe ocular phenotypes in these mutants. Human*MAB21L*genes belong to the mab-21 family; first discovered in*C. elegans*, *mab-21* mutants display thinning of ray 6 and frequent fusion of rays 4 and 6. The molecular function of*MAB21L2*is unknown and its role in development is not well understood. Double mutant analysis in*C. elegans*indicates that *mab-21* is downstream of SMADs and is negatively regulated by BMPs, placing *mab-21* in the TGF-b/BMP signaling cascade. Additionally, *mab-18* (PAX6 [MIM 607108]ortholog) mutants demonstrate a phenotype that is highly similar to *mab-21C. elegans*mutants. Mutations in *PAX6*, considered a "master regulator" of eye development, *BMP4*[MIM 112262] and *BMP7*[MIM 112267]all cause ocular phenotypes, including coloboma. Taken together these data demonstrate the importance of the PAX6 pathway and TGF-b/BMP signaling cascade in ocular development and possible interaction of *MAB21L2*with them. To investigate these interactions, we crossed*mab21l2*mutants onto a destabilized eGFP (GFP) reporter line driven by Smad-mediated BMP signaling and a*pax6b*mutant line. In 24 hours post fertilization (hpf) wild type embryos, GFP expression is seen in the dorsal eye and lens. In both*mab21l2*mutants GFP levelsare reduced; additionally, GFP expression inhomozygous*mab21l2R51_F52del*embryos was found to be ectopically expressed in the posterior part of the eye. Mutants carrying double homozygous mutations in *mab21l2Q48Sfs*5*and*pax6b*demonstrated small eyes completely lacking the lens and ventral retina, which isa more pronounced phenotype in comparison tothe single mutants for either gene and thus suggests a synergistic effect for*mab21l2*and*pax6b*mutations in zebrafish. These findings reveal interaction of MAB21L2/mab21l2 with the TGF-b/BMP signaling cascade and PAX6/pax6 pathway during vertebrate eye development. Additional studies are underway to further define these relationships.

2803F

Ubiquitous expression of the Proteus mutation *Akt1* c. 49G>A, p. Glu17Lys causes embryonic lethality in mice. *M. J. Lindhurst[1], M. R. Yourick[1], C. Rivas[2], G. Elliott[2], J. Gomez-Rodriguez[2], L. Garrett[2], P. L. Schwartzberg[2], D. Bodine[3], L. G. Biesecker[1].* 1) MGMGB, NHGRI/NIH, Bethesda, MD; 2) GDRB, NHGRI/NIH, Bethesda, MD; 3) GMBB, NHGRI/NIH, Bethesda, MD.

Proteus syndrome is a rare disorder characterized by progressive, sporadic, segmental overgrowth that can affect any tissue in the body. Common manifestations include distorting bony overgrowth, cerebriform connective tissue nevi, epidermal nevi, dysregulated adipose tissue, and vascular malformations. It is clinically and molecularly distinct from other segmental overgrowth disorders and is caused in all patients evaluated to date, by the same c. 49G>A, p. E17K mosaic activating mutation in *AKT1*. This affirms the hypothesis of Happle (1986) who hypothesized that Proteus syndrome was due to a dominant mutation that would be lethal if present in the germ line. For many reasons, it is impossible to study the mechanism of lethality in humans. Therefore, we set out to test the Happle hypothesis using an inducible mouse model for the Proteus syndrome mutation. The conditional knock-in allele (*Akt1cond*) contains the c. 49G>A mutation and a STOP cassette inserted immediately upstream of the mutation and silences the allele until it is removed by Cre recombinase. To activate the mutation early in development, mice heterozygous for *Akt1cond* were crossed to mice expressing Cre using the β-actin promoter. No live born mutant (*Akt1mut*) offspring were identified (n=32). Timed matings were performed and embryos were harvested between E10. 5 and E17. 5 and assessed for appearance, heartbeat, and genotype. Before E14. 5, the phenotype of *Akt1mut* embryos ranged from normal with heartbeats to dead and being resorbed. By E14. 5, all *Akt1mut* embryos had a visible phenotype that included scattered hemorrhages throughout the embryo, fewer blood vessels and pale yolk sacs, although some of these embryos were still alive. Colony forming assays are being conducted to assess whether fetal livers from mutant E13. 5 embryos have similar numbers of myeloid progenitor cells and immunohistochemistry and perfusion studies are underway to examine the location, structure, and integrity of the embryonic vasculature. These experiments will contribute to our understanding why the mosaicism for the c. 49G>A mutation is rare in peripheral blood in humans with Proteus syndrome and test the Happle hypothesis. These data suggest that AKT1 plays a critical role in the development of the murine circulatory system.

2804W

Involvement of GTF2IRD1 in the control of facial skin features and patterning of Williams-Beuren Syndrome. *C. P. Canales[1], S. Corley[2], P. Carmona-Mora[1], F. Tomasetig[1], A. Beverdam[1], P. Kaur[3], I. Smyth[4], M. Wilkins[2], E. C. Hardeman[1], S. J. Palmer[1].* 1) Cellular and Genetic Medicine Unit, UNSW Australia, Kensington, Sydney, New South Wales, Australia; 2) School of Biotechnology and Biomolecular Sciences, the New South Wales Systems Biology Initiative, UNSW Australia, Sydney NSW, Australia; 3) Epithelial Stem Cell Biology Laboratory, Peter MacCallum Cancer Centre, Melbourne VIC, Australia; 4) Dept of Biochemistry and Molecular Biology, Monash University, Melbourne VIC, Australia.

GTF2IRD1 is an important transcriptional regulator located within the Williams-Beuren Syndrome (WBS) common microdeletion region in 7q11. 23, which involves up to 28 genes. Deletion of *GTF2IRD1*, discovered in our laboratory, and the evolutionary-related *GTF2I*, which encodes a protein (TFII-I) with overlapping functional properties, account for the major aspects of the WBS cognitive profile. Analysis of *Gtf2ird1* knockout mice recently allowed us to demonstrate that GTF2IRD1 has an important role in the auditory pathology of WBS. Genotype-phenotype correlations, drawn from patients carrying atypical deletions or duplications of the WBS critical region, have identified GTF2IRD1 and TFII-I as prime candidates for the cause of the distinctive WBS facial dysmorphology. By careful analysis of *Gtf2ird1* knockout mouse models, we have investigated the role of this transcriptional regulator in the control of epidermal proliferation and differentiation and its possible involvement in the causation of the physical WBS craniofacial features. Facial MRI analysis showed that the lack of *Gtf2ird1* produced striking similarities to aspects of the human disease, with soft tissue enlargement due to an extreme epidermal thickening in restricted regions of the face. However, CT imaging of this knockout model did not indicate any abnormalities of the skull. At the cellular level, keratinocyte proliferation was found to be augmented, keratinized epidermal layers were disorganized and the barrier function of the skin was compromised in the affected facial zones of *Gtf2ird1-/-* mice. Analysis of *Gtf2ird1* expression during development showed that transcripts from this gene are most abundant in regions of the skin where the phenotype is observed. In-depth RNA-Seq analyses of affected skin regions identified an abnormal FGF7 expression profile and associated molecular pathways. These data indicate that GTF2IRD1 is a crucial component of the transcriptional regulation machinery in specific regions of the skin and plays an important role in facial skin patterning, and keratinocyte proliferation and differentiation. Furthermore, these data also provide an insight into some of the molecular pathways involved in the craniofacial abnormalities of WBS.

2805T

Loss of function mutation of EGFR in compound heterozygous state causes severe skin defect, gastrointestinal dysfunction and electrolyte imbalance. *T. Yokoi[1,3], C. Hatano[1], Y. Tsurusaki[1], Y. Enomoto[1], T. Naruto[2], S. Hayashi[3], M. Kobayashi[3], H. Ida[3], K. Kurosawa[1].* 1) Division of Medical Genetics, Kanagawa Children's Medical Center, Kanagawa, Japan; 2) Department of Stress Science, Tokushima University Faculty of Medicine, Tokushima, Japan; 3) Department of Pediatrics, The Jikei University of School of Medicine, Tokyo, Japan.

<Background>Epidermal Growth Factor Receptor; EGFR is a receptor which recognizes EGF and is very important regulator for cell proliferation and differentiation. In this report we experienced a case which has compound heterozygous loss of function (LoF) mutations in *EGFR*. <Case report>He was born at 32 weeks' gestation, weighing 1089 g. He had defect of scalp hair and ventricular septal defect but no other morphological abnormality at birth. He had eczema, anhidrosis, hyperalgesia, milk allergy and electrolyte imbalance. He died from septic shock at day 134. <Method>Using the patient's DNA from peripheral blood, we performed Mendelian disorders related gene-based Next Generation Sequencing (NGS) (TruSight One, 4813 genes) on the Illumina Miseq. <Results> We identified two LoF mutations in *EGFR*; R98X/I365N. The pathogenicity of I365N is severely deleterious by ANNOVAR. Sanger sequencing demonstrated that each mutation was derived from each parent. <Discussion> Campbell et al. [2014] reported a male infant who had homozygous missense mutation of *EGFR* and had respiratory symptom add to cutaneous and digestive symptoms. The *EGFR* knock-out mice are not embryonic lethal but have the common phenotype in human as shown in the patients. These results suggest that EGFR is not necessarily essential for period of organogenesis, but is considered to be important for organs which has tissue-specific immune system, such as lung, digestive organs and skin after birth. NGS has been useful for detecting new pathogenic genes and new pathogenic variants in known genes. Further, rare complete knock-out in patients with characteristic and severe defect phenotype highlights the crucial roles of the genes in human development. NGS makes it possible to analyze unrecognizable function in known gene and establish new causes of diseases.

2806F

Lipoid Proteinosis (LP), a complex neuro-cutaneous disorder: Novel and recurrent mutations in the *ECM1* gene, and consequences of *Ecm1* knock-down in zebrafish. *L. Youssefian[1], H. Vahidnezhad[1,2,3], P. Mansouri[2], MA. Barzegar[4], M. Daneshpazhouh[2], H. R. Mahmoudi[2], R. Mobasher[2], S. Zeinali[3], Q. Li[1], J. Uitto[1].* 1) Thomas Jefferson University, Philadelphia, PA; 2) Tehran University of Medical Sciences, Tehran, Iran; 3) Pasteur Institute of Iran, Tehran, Iran; 4) Shahid Beheshti University of Medical Sciences, Tehran, Iran.

LP is a rare autosomal recessive disorder with cutaneous and neurologic manifestations. Specifically, the skin findings consist of infiltration of the skin with formation of verrucous plaques on the elbows and pearly eyelid papules. Neurologic manifestations include epilepsy and convulsions associated in most cases with mineralization of the brain. LP is caused by mutations in the *ECM1* gene which results in extensive phenotypic variability even in individuals with the same mutation. In this study, we have examined an Iranian cohort of 13 families comprised of 32 affected individuals with LP, all families being consanguineous. LP was diagnosed with characteristic mucocutaneous findings, and neurologic manifestations were present in half of the patients. Mutation analysis by sequencing of all 10 exons of *ECM1* in these families revealed homozygous loss-of-function mutations, 12 of them being nonsense mutations or small deletions resulting in frameshift of translation. Three of these mutations, c. 477delC, IVS3-2G>A, and p. W237R, were previously unreported. One missense mutation (p. W237R) was predicted to be pathogenic by informatics program (PolyPhen -2) and by conservation of W237 through evolution. Among the mutations, two of them, c. 507delT and c. 735delTG, were recurrent and were found independently in four separate families. Haplotype analysis with intragenic and flanking SNPs on chromosomal locus 1q21 identified a 2 Mb conserved region in families with c. 507delT, suggesting that this mutation is a result of founder effect, rather than being a hotspot mutation. The pathogenic role of *ECM1* mutations was confirmed in a zebrafish model using morpholino-mediated knock-down approach. A splice-site morpholino was designed at the exon 3/intron 3 junction of *Ecm1* to prevent pre-mRNA splicing, which decreased the corresponding mRNA levels by >95%. The knock-down induced a visible skin phenotype consisting of altered surface contour with micropapules and loss of microridges, noticeable at two days post-fertilization. Transmission electron microscopy of the morphants demonstrated apparent reduplication of basement membrane and accumulation of pinocytotic vesicles in the periphery of fibroblastic cells within the dermis, similar to findings seen in patients with LP. The knock-down phenotype was partially rescued with full-length human *ECM1a* mRNA. These results suggest zebrafish is a suitable model system to test the pathogenicity of missense mutations by mRNA rescue.

2807W

Zebrafish genetics and disease models: an emerging model for human disease. *S. A. Hutchinson, J. J. Dowling.* Genetics and Genome Biology, Hospital for Sick Chlidren, Toronto, Ontario, Canada.

The advance of next generation sequencing has resulted in the identification of many new disease associated gene mutations. However, validating the function of these variants, particularly in rare diseases, is still a challenge. Recent advances in targeted mutagenesis technology have led to the development of zebrafish models of human disease. Zebrafish have rapid early development, transparent embryos, and large clutch sizes making them suitable for high throughput drug screening and quick, *in vivo* testing of the consequences of newly identified gene mutations. Our goal is to use the expertise and infrastructure at the Hospital for Sick Children to establish a Zebrafish Genetics and Disease Models Facility, accessible to researchers worldwide, that provides all the services required to generate and analyze zebrafish models of human disease at the highest quality, most affordable prices and most efficient time frame. We provide targeted mutant generation and phenotype analysis services for a reasonable fee as part of our cost-recovery business plan. We are currently developing models for a diverse set of diseases including, but not limited to, inflammatory bowel disease, epilepsy, cardiac arrhythmia and congential myopathy. Zebrafish models of congenital myopathy led to both validation of human disease associated mutations, and the discovery of a novel potential therapeutic drug. For example, *CCDC78* was identified as a disease associated gene in a dominantly inherited congenital myopathy. Next, a zebrafish model of the disease resulted in fish with altered motor function and abnormal muscle ultrastructure that mirrored the human disease confirming *CCDC78* as a causative mutation1. Another example is the use of zebrafish dystrophin mutants, a gene known to be mutated Duchenne muscular dystrophy, in a non-biased drug screen looking for rescue of movement defects. A selective serotonin reuptake inhibitor called fluoexetine was identified as a potential therapeutic and the serotonin pathway as a target for further drug development2. The ability to accelerate the pipeline from identification of human mutations to drug discovery is a strength of the zebrafish model system. We provide access to this powerful model system by using our experience in zebrafish mutant generation, characterization and drug discovery to develop these models as a fee-for-service cost. Examples of our initial zebrafish mutants created to model and study human diseases will be presented.

2808T

Facilitating access to human disease models. *S. Rockwood, M. Sasner, The Genetic Resource Science Team.* The Jackson Laboratory, Bar Harbor, ME.

The mouse continues to be a widely used model organism, especially well-suited for the study of mammalian development and physiology, and more importantly, for elucidating details of disease mechanisms. Recent advances in genetic engineering have expanded the options available to researchers to precisely generate specific disease models. The Jackson Laboratory (JAX) Mouse Repository, as part of its mission, facilitates the access of these valuable mouse strains to the international biomedical research community. The JAX Mouse Repository has served as a centralized resource for the development, distribution, and cryopreservation of high health status mouse models for over 60 years. A wide selection of models are now available that have applications as hosts for cancer xenograft modeling, stem cell biology, and infectious disease research. Strains are increasingly specialized for specific purposes: *optimized* support of human and murine hematopoietic cell engraftment, *reduced* xenogeneic graft-versus-host disease response, engraftment of human hematopoietic stem cells *without* irradiation and immunological *tolerance* to firefly luciferase and enhanced green fluorescent protein. Mouse strains that recapitulate aspects of specific diseases and disorders, such as Alzheimer's Disease or Chordoma are available as are those that create conditions that facilitate disease states, such as increased oxidative stress. In response to the promising potential of CRISPR/Cas9 technology to be instrumental in the generation of novel mouse models of disease, a variety of Cas9-expressing lines are now available, including constitutive expressing and Cre-inducible models. Cas9 strains are available on multiple genetic backgrounds. In order to make the vast number of Repository models more accessible, we have recently developed a new search interface that facilitates queries by disease term as well as by keyword and strain, allele, and gene name or synonym. Donating a strain to the Jax Mouse Repository is an easy way to fulfill the NIH's requirements for sharing mice. Researchers wishing to have strains considered for inclusion in the Repository are encouraged to submit their strains at: www. jax. org/donate-a-mouse. This work is supported by NIH, The Howard Hughes Medical Institute, and several private charitable foundations.

2809F

Analysis of zebrafish orthologs of the human Peters Plus syndrome gene, *b3glcta* and *b3glctb*, using morpholino knockdown and TALEN-mediated genetic knockout assays. *E. R. Weh[1], N. Mlodik[1], S. Muheisen[1], E. V. Semina[1,2].* 1) Department of Pediatrics, Medical College of Wisconsin, Milwaukee, WI; 2) Department of Cell Biology, Neurobiology, and Anatomy, Medical College of Wisconsin, Milwaukee, WI.

Human *B3GLCT* deficiency results in a complex phenotype known as Peters Plus Syndrome (PPS). Patients with PPS display variable features, however three are almost invariably present: Peters anomaly (or other forms of anterior segment dysgenesis), short stature and brachydactyly. *B3GLCT* encodes for β-1,3-glucosyltransferase which is expressed throughout the developing embryo and adult tissues. Recently it was found that loss of B3GLCT disrupts a non-canonical ER quality control mechanism, however it is not known how loss of this apparently ubiquitously expressed enzyme can lead to tissue specific defects observed in patients with PPS. In this study we aimed to characterize the two zebrafish orthologs of *B3GLCT*, *b3glcta* and *b3glctb*, and to create a model of PPS using both morpholino (MO) mediated knockdown and TALEN genome editing. We identified the full coding region of both orthologs via 5' RACE and defined their expression patterns during embryonic development. Embryos were injected with two independent splice interrupting MOs for each ortholog and complete knockdown was confirmed using RT-PCR. The affected mRNA transcripts were predicted to encode a truncated protein product lacking the entire catalytic domain. We found a robust, repeatable phenotype correlating with the degree of knockdown of either transcript. Morphant embryos displayed shorter body length, disrupted craniofacial cartilage development, ocular defects and short, stubby fins; the phenotype associated with complete *b3glcta* deficiency was more severe than *b3glctb* deficiency. Following this, genetic knockout lines for both *b3glct* genes were generated using TALEN technology: for *b3gclta*, four frameshift mutations in exon 1 and a 1-bp deletion in exon 12 (c. 1013delG;p. (S338Ifs*9)) were generated; for *b3gcltb,* a 22-bp deletion in exon 12 (c. 1005_1026del22;p. (A336Cfs*2)) was created. The p. (S338Ifs*9) and p. (A336Cfs*2) mutations are expected to disrupt the required catalytic DXD residues located at positions 350-352 and 347-349 for b3gclta and b3gcltb, respectively. Contrary to MO data, both single and double homozygous *b3glcta:b3glctb* mutants showed no observable embryonic or adult phenotype. Our data indicate that, despite significant sequence conservation, the zebrafish orthologs of *B3GLCT* are not required for normal embryonic development. Additionally, the discrepancy between phenotypic effects produced by MO and genome editing approaches suggest caution in using MOs to model human disease.

2810W

A heterozygous mutation of *MOV10L1* in a pedigree of spermatogenic failure. *J. Lu[1], Y. Shao[1], J. Xiao[1], Q. Wang[1], X. Zhang[1], J. Zhang[1], C. Li[1], Z. Li[2], F. Zhang[1].* 1) Fudan university, ShangHai, China; 2) ShangHai General Hospital, ShangHai, China.

Piwi-interactive RNA (piRNA) is essential for silencing of transposon and genome integrity in the germline. Previous experimental evidence using mouse models has showed that the homozygous null mutations of piRNA biogenesis genes can result in spermatogenic defects and male infertility. However, the spermatogenic roles of piRNA biogenesis gene mutations have not been elucidated in human. Here we identified a heterozygous missense mutation in *MOV10L1*, a gene involved in piRNA biogenesis, in a pedigree of spermatogenic failure. It suggested that the inheritance pattern is autosomal dominant. The female carriers of this heterozygous *MOV10L1* mutation showed no defects in fertility. The knock-out (KO) mouse model has been generated using the CRISPR-Cas9 technology for further functional assays of this *MOV10L1* mutation.

2811T

A new mouse model of type 2 Gaucher disease. *A. Gonzalez[1], R. Grey[1], B. Berhe[1], E. I. Ginns[2], D. Ory[3], P. Elias[4], E. Sidransky[1], N. Tayebi[1].* 1) Molecular Neurogenetics section, MGB/ NHGRI, National Institutes of Health, Bethesda, MD; 2) Lysosomal Disorders Treatment and Research Program, Clinical Labs, University of Massachusetts Medical School, Worcester, MA 01545, USA; 3) Diabetic Cardiovascular Disease Center, Washington University School of Medicine, St. Louis, Missouri 63110, USA; 4) Dermatology Service, Department of Veterans Affairs Medical Center, and Department of Dermatology, University of California, San Francisco, CA USA.

Gaucher disease (GD) is a recessive lysosomal storage disorder caused by mutations in *GBA1*, the gene coding for the hydrolase glucocerebrosidase (GCase). GD is classically divided into three types based on the presence and progression of neurological symptoms. Type 1 GD is the non-neuronopathic form and presents with broad phenotypic heterogeneity. Types 2 and 3 GD are the less common acute and chronic neuronopathic forms, respectively. Type 2, the most severe form of GD, presents with symptoms either prenatally or during infancy, and patients usually die before the age of 3 years. In order to study the pathogenesis of neuronopathic type 2 GD we have generated a new mouse model by crossing two previously described mouse lines: the knock-in point mutation Leu444Pro (c. 1448T>G) and an effectively null *gba* allele created through the targeted disruption of mouse GCase (ϑMGC). A GCase null /Leu444Pro genotype is common among patients with type 2 GD. The Leu444Pro/ϑMGC mice were non-viable, dying within the first few hours after birth. GCase protein and enzymatic activity levels in different tissues from Leu444Pro/ϑMGC were similar to levels in homozygous null *gba* mice. However, Leu444Pro/ϑMGC mice do not show the lipid storage observed in the homozygous null *gba* mice, and thus the cause of death remains unknown. While brain LIMP-2 levels were unaltered in these mice, there is some evidence of enhanced neuroinflammation, shown by increased levels of one GFAP isoform in the brains from the Leu444Pro/ϑMGC mice compared to wild type controls. To identify factors contributing to the lethality of these mice, ultrastructural studies of skin are being performed and other indicators of neuroinflammation are being evaluated. These mice may offer a new model in which to study the pathogenesis of neuronopathic GD as well as other possible functions of GCase.

2812F

Modeling DPMS function in zebrafish to explore the interface of dystroglycanopathies and congenital disorders of glycosylation. *M. Manzini[1], D. von Alpen[1], H. Issa[1], H. Pond[1], C. Ardiccioni[2], O. Clarke[2], F. Mancia[2].* 1) Pharmacology and Physiology, George Washington University, Washington, DC; 2) Physiology and Cellular Biophysics, Columbia University, New York, NY.

Dolichol-phosphate mannose synthase (DPMS) catalyzes the synthesis of dolichol-P-mannose, an important donor substrate for multiple glycosylation processes in the endoplasmic reticulum. Autosomal recessive mutations in the DPMS subunits *DPM1*, *2* and *3* cause congenital disorder of glycosylation 1e (CDG1e), a complex disorder characterized by developmental delay, hypotonia, dysmorphisms, microcephaly and seizures. In some cases a muscle biopsy showing muscular dystrophy and loss of dystroglycan glycosylation has been observed, suggesting an overlap with dystroglycanopathies, the most severe forms of congenital muscular dystrophy. To better understand the function of DPMS we have used the crystal structure of GtrB, a bacterial homolog of DPMS, to map the residues mutated in humans. In addition, we have developed zebrafish models of *dpm1*, *2* and *3* loss-of-function, which recapitulate phenotypes found in humans such as developmental delay, microcephaly, muscle weakness and vascular abnormalities, providing a good vertebrate model for these disorders. We then performed structure-function studies in the zebrafish by injecting human mRNA carrying the human mutations to study how different mutations affect enzyme function. With this approach we found striking correlation between enzyme function and phenotype. We are now exploring how the muscle, the brain and other tissues are affected. In summary, we have developed a strategy combining structural biology and animal models to determine the catalytic mechanisms of DPMS and the function of human *DPM1, 2* and *3* mutations in order to study the role of these proteins in the pathogenesis of CDGs and dystroglycanopathies.

2813W

Functional characterization of deleterious alleles in the vitamin B12 receptor, *CD320*, in humans and mice. *D. J. Bernard[1], F. J. Pangilinan[1], J. Cheng[2], J. L. Mills[3], D. M. Kay[4], C. Horan[5], A. M. Molloy[6], L. C. Brody[1].* 1) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, NIH, Bethesda, MD; 2) Transgenic Mouse Core, National Human Genome Research Institute, NIH, Bethesda, MD; 3) Division of Intramural Public Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, Bethesda, MD; 4) New York State Department of Health, Division of Genetics, Albany, NY; 5) Institute of Molecular Medicine, Trinity Centre for Health Sciences, St. James's Hospital, Dublin, Ireland; 6) Department of Clinical Medicine, School of Medicine, Trinity College, Dublin, Ireland.

Inherited and acquired deficiencies of cobalamin (vitamin B12) can cause megaloblastic anemia (OMIM#261100) and serious neurologic disease. Low maternal levels of vitamin B12 increase the risk of pregnancy complicated by neural tube defects. Poor nutrition, malabsorption and genetic variation in the vitamin B12 transport and metabolic pathways can all produce cobalamin deficiency. Cobalamin is transported from the circulation into tissues via the vitamin B12 receptor, TCblR, encoded by the *CD320* gene. We hypothesized that variants in *CD320* would produce a metabolic picture consistent with cobalamin deficiency. Newborn screening programs measure C3 (propionylcarnitine) to detect a number of defects including cobalamin deficiency. We evaluated newborns with elevated and normal C3, testing for a 3bp deletion in *CD320* (c. 262_264delGAG, rs150384171, MAF ~0. 01) previously observed with transiently elevated levels of methylmalonic acid (MMA) in newborns (OMIM#606475). Significantly, 7 of 351 newborn cases with elevated C3 and none of 388 controls were found to be homozygous for this deletion ($p<6x10-6$). To evaluate the impact of *CD320* on cobalamin metabolism, we created *Cd320* null knockout mice. *Cd320* knockouts are healthy, viable and fertile. These mice present with a metabolic phenotype consistent with intracellular cobalamin deficiency: low tissue cobalamin and elevated plasma MMA and homocysteine. Of note, these animals resemble the biochemical abnormalities observed in humans with a compromised ability to absorb cobalamin. Dietary restriction of vitamin B12 in *Cd320* knockout mice produces infertility in females but not males. Experiments to determine the cause of the infertility indicate that although conception occurs, the embryos perish around the implantation stage. A complete absence of intracellular cobalamin is lethal. Ourresults with *Cd320* null animals maintained on diet containing ample vitamin B12 suggest that an alternative vitamin B12 transport mechanism exists in mice, and perhaps humans. The biochemical phenotype of *Cd320* null animals raised on vitamin B12 deficient chow is consistent with historical and epidemiological data suggesting that low vitamin B12 levels in humans are associated with infertility and developmental abnormalities. Genetic variation in the cobalamin receptor may play a role in health and disease throughout the life cycle.

2814T

Strategy for generating and characterizing a zebrafish model of spondylometaphyseal dysplasia with cone-rod dystrophy. *J. Jurgens, C. Woods, P. W. Hook, N. Sobreira, J. Hoover-Fong, A. S. McCallion, D. Valle.* Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD.

Spondylometaphyseal dysplasia with cone-rod dystrophy (SMD-CRD) is a rare, autosomal recessive disorder caused by partial or complete loss-of-function variants in the phosphatidylcholine biosynthesis gene *PCYT1A* [Hoover-Fong et al. , 2014; Yamamoto et al, 2014]. Clinical characteristics of SMD-CRD include progressive, early-onset degeneration of cone and rod photoreceptors as well as short stature, bowing of the long bones, metaphyseal flaring, rhizomelic shortening, platyspondyly, and scoliosis. The mechanisms underlying this phenotype are not well understood, in part because its molecular basis was recently identified and there are no existing animal models. Here we describe a strategy for generating and characterizing a zebrafish knockout model of SMD-CRD. We first analyzed mRNA expression patterns of the two zebrafish orthologs of *PCYT1A*, *pcyt1aa* and *pcyt1ab*, in fish embryos at 24, 48, and 72 hours post-fertilization (h. p. f.) by *in situ* hybridization. Both paralogs showed broad expression at 24 h. p. f. and acquired more anterior localization, including retinal expression, at the two later timepoints. Next, we targeted *pcyt1aa* and *pcyt1ab* for knockout using CRISPR/Cas9 genome editing. This system generates a double-stranded DNA break which is then repaired by nonhomologous end joining (NHEJ). Since NHEJ is an error-prone mechanism, it generates small insertions or deletions at the site of repair, often resulting in gene knockout. We exploited this system by microinjecting Cas9 RNA and guide RNAs directed against *pcyt1aa* and *pcyt1ab* into zebrafish embryos at the single-cell stage, resulting in fish mosaic for a heterozygous null allele in the corresponding paralog. These fish were then outcrossed to generate germline heterozygous knockouts for each of the two paralogs, which were screened by Surveyor assay followed by sequence verification. As in humans, the heterozygotes are asymptomatic. We are breeding these animals to generate individual and double homozygous knockouts for *pcyt1aa* and *pcyt1ab*, which we will examine for relevant phenotypes with tests including length measurement, bone and cartilage staining, optomotor response assays, and prey capture assays. We anticipate that these models will provide a resource for screening potential small molecule and/or nutritional therapies. We suggest that creating and characterizing a zebrafish model of SMD-CRD will facilitate greater understanding of this phenotype and perhaps other bone or retinal dysplasias.

2815F

Development of a zebrafish model for ATP7A-related motor neuron disease. *D. Martinelli[1], B. Feldman[2], L. Yi[1], V. Shram[3], S. G. Kaler[1].* 1) Section on Translational Neuroscience, Molecular Medicine Program, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, Bethesda, MD; 2) Zebrafish Core, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, Bethesda, MD; 3) Section on Cell Biology and Signal Transduction, Eunice Kennedy Shriver National Institute of Child Health and Human Development, NIH, Bethesda, MD.

Distal hereditary motor neuropathies comprise a heterogeneous group of diseases characterized by a length-dependent axonal neuropathy. The ATPase ATP7A regulates cellular copper homeostasis and is associated with Menkes disease, a severe pediatric neurocutaneous disorder. Recently, two novel *ATP7A* missense mutations (T994I and P1386S), have been implicated in a late-onset distal motor neuropathy without overt copper metabolic abnormalities or other clinical abnormalities. Understanding the precise functional role of ATP7A in the peripheral nervous system and the neurodegenerative effects of these two mutations remains elusive. *Danio rerio* (zebrafish) has emerged as a useful model of Menkes disease: a mutant harboring a mutation in the zebrafish homolog of ATP7A, *atp7avu69* (*calamity*), shows defects in pigmentation and hindbrain development, and early mortality (within 1 week). We are using *calamity* as a model system in which to study the effects of the human ATP7A motor neuropathy inducing alleles. We crossed *calamity*+/- heterozygotes with transgenic zebrafish that express green fluorescent protein (GFP) driven by a motor neuron-specific promoter, Hb9. This breeding scheme resulted in adult fish carrying one *atp7avu69* allele and one *Tg(Hb9-GFP)* allele. We crossed these double heterozygote fish (*Hb9-GFP*+/- /*calamity*+/-) to obtain *calamity*-/- mutants that selectively express GFP in motor neurons. Using time-lapse confocal imaging, we evaluated motor neuron development *in vivo* from 24 to 72 hours post fertilization (hpf). We observed motor neuron degeneration starting as early as 48 hpf in *calamity*, associated with progressive paralysis, as assessed by neurobehavioral testing (active motility; touch, light and sound sensitivity). Mutant embryos also showed disorganized motor neurons cell bodies and axonal distribution, and reduced GFP signal compared to wild type, suggesting an essential role of *atp7a* in normal zebrafish motor neuron development. We are currently exploiting this model system to dissect human disease mechanism(s) by microinjection of the P1386S and T994I human ATP7A RNAs into *calamity* embryos. We predict that these hypomorphic alleles will correct the early mortality and enable longer survival, allowing evaluation of the specific motor neuron degenerative effects of P1386S and T994I.

2816W

Differentially expressed striatal transcripts in the OVT73 ovine model of Huntington 's Disease. *R. G. Snell[1], S. J Reid[1], R. R Handley[1], R. Brauning[2], P. Maclean[3], S. Patassini[1], P. Tsai[4], H. J Waldvogel[1], J. F Gusella[5], M. E MacDonald[5], R. L. M Faull[1], H. D HDSCRG[6].* 1) Center for brain reserach University of Auckland, Auckland, Auckland, New Zealand; 2) AgResearch, Invermay Agricultural Centre, Mosgiel, New Zealand; 3) AgResearch, Ruakura Research Centre, Hamilton, New Zealand; 4) School of Biological Sciences, The University of Auckland, Auckland; 5) Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston; 6) Huntington's Disease Sheep Collaborative Research Group.

We have made considerable progress in the characterisation of our Huntington's Disease, OVT73 transgenic sheep model and have identified time dependent progressive histopathological, behavioural and molecular changes. To progress beyond confirming that this model accurately represents the pre-symptomatic phase of HD we have undertaken a deep RNAseq experiment to identify transgene effects on transcription. RNA-seq data was collected from ribosomal depleted total RNA isolated from anterior striatum tissue of twelve 5-year old OVT73 sheep (6 transgenics, 6 controls). Specific differentially expressed transcripts in OVT73 were identified and 25 selected for further analysis. Nanostring nCounter quantification confirmed thirteen of these genes as differentially expressed in various sub-regions of the striatum. The validated transcripts will be presented; these include a urea transporter (SLC14A1) and a choline transporter (SLC5A7), both up-regulated in OVT73 striatum. The OVT73 line animals express the full length human huntingtin protein with 73 glutamine repeats[1]. We have shown there is reduced striatum levels of GABAA α1 receptor, cortical huntingtin aggregates [2] and an early circadian abnormality [3]. The sheep are healthy and visibly indistinguishable from control South Australian Merino. They present an opportunity to better understand and utilise the pre-symptomatic phase of Huntington's Disease. 1. Jacobsen, J. C. , et al. , *An ovine transgenic Huntington's disease model.* Hum Mol Genet, 2010. **19**(10): p. 1873-82. 2. HDSCRG, et al. , *Further Molecular Characterisation of the OVT73 Transgenic Sheep Model of Huntington's Disease Identifies Cortical Aggregates.* J Huntingtons Dis, 2013. **2**(3): p. 279-295. 3. Morton, A. J. , et al. , *Early and progressive circadian abnormalities in Huntington's disease sheep are unmasked by social environment.* Hum Mol Genet, 2014.

2817T

Functional dissection of *PITX2* in zebrafish. *K. Hendee[1,2], E. Sorokina[2], E. Semina[1,2].* 1) Cell Biology, Neurobiology and Anatomy, Medical College of Wisconsin, Milwaukee, WI; 2) Pediatrics and Children's Research Institute, Children's Hospital of Wisconsin, Milwaukee, WI.

Paired-like homeodomain 2 (*PITX2* [MIM 601542]) is a transcription factor that has been associated with Axenfeld-Rieger syndrome (ARS [MIM 180500]). The pathway by which *PITX2* causes the disease phenotype has yet to be elucidated. In order to identify the downstream targets of PITX2, we generated a zebrafish *pitx2* knockout line and performed comparative transcriptome analyses of wild-type and mutant tissues. TALEN-mediated genome editing produced an 8-bp deletion in the *pitx2* homeobox domain, resulting in a premature stop codon (c. 565_572del8 (p. (M64*)). The mutation is predicted to produce a truncated protein lacking a functional DNA-binding domain. Homozygous *pitx2M64** embryos display a severely underdeveloped anterior chamber beginning at 96 hours post fertilization (hpf); additionally, many embryos exhibit smaller eyes and craniofacial abnormalities, whereas heterozygous *pitx2M64**embryos appear phenotypically normal. Adult *pitx2M64** heterozygous fish were bred with a *Tg(-2. 6pitx2-CE4:GFP)* reporter line that expresses GFP under control of the *CE4* upstream regulatory element of *pitx2* in periocular mesenchyme*.* Head tissues from a mix of (*CE4:GFP):pitx2M64** heterozygous and homozygous embryos along with (*CE4:GFP)* control embryos were collected at 24 hpf, and GFP-positive and negative cell populations were isolated using fluorescence-activated cell sorting. RNA extracted from the collected cells was submitted for Affymetrix microarray analysis. Transcriptome comparison between mutant and control GFP-positive populations identified 96 significantly downregulated and 131 upregulated genes. The top targets of interest were notum pectinacetylesterase homolog 1b (Drosophila) (*notum1b*) and elastin microfibril interfacer 2a (*emilin2a*), both of which have human orthologues (MIM 609847 and MIM 608928, correspondingly). Downregulation of both *notum1b* and *emilin2a* transcripts in *pitx2* mutant tissues was independently confirmed by qPCR with transcript-specific primers. *In situ* hybridization analysis identified broad embryonic expression for both genes, with *emilin2a* showing some enrichment in the periocular mesenchyme. Further analysis is underway to explore developmental roles of *notum1b* and *emilin2a* and their functions in the *pitx2* pathway. Identifying downstream targets of *PITX2/pitx2* will allow for better understanding of mechanisms of normal ocular development and disease processes.

2818F

Postnatal excision of a Bardet-Biedl Syndrome gene results in leptin resistance, leading to obesity. *J. E. Garrison[1], Y. Hsu[1], D. Y. Nishimura[1], K. E. Bugge[1], C. C. Searby[1], V. C. Sheffield[1,2].* 1) Department of Pediatrics, University of Iowa, Iowa City, IA; 2) Howard Hughes Medical Institute.

Bardet-Biedl Syndrome (BBS) is a rare, autosomal recessive disorder that is commonly associated with obesity, retinitis pigmentosa, postaxial polydactyly, renal abnormalities, urogenital defects, and intellectual disability as well as increased susceptibility to hypertension, heart disease, and diabetes mellitus. Although BBS is rare in the general population, identifying the underlying molecular mechanisms of the disorders associated with this syndrome has garnered great interest due to the fact that components of the phenotype including obesity, hypertension and diabetes are common. Patients with BBS usually present with obesity (BMI >30), and have increased visceral fat and increased adiposity compared to BMI-matched controls. BBS mouse models consistently develop obesity. Previous data indicate that aspects of leptin signaling do not function optimally in congenital BBS congenital knockout mice, even when the mice have been calorie restricted to prevent obesity. This indicates that leptin resistance is intrinsic and primary to obesity. However, controversy remains as to whether leptin resistance is intrinsic or secondary to obesity. In order to understand the development of leptin resistance and obesity in BBS, we developed a conditional genetic model in which we can temporally eliminate expression of *Bbs8*, a component of the BBSome. As such, we can examine the role of BBS8 in the development or amelioration of phenotypes in adult mice and determine the critical time points for BBSome function in the pathophysiology of obesity. Using this system, we are able to achieve efficient removal of *Bbs8* in adult mice at the DNA and protein levels. Tamoxifen-induction of Cre recombinase leads to obesity within one to two months after treatment. Plasma leptin levels in treated mice are comparable to control littermates until the mice become obese, suggesting that plasma leptin levels rise in response to increased adiposity. Some mice in which *Bbs8*-excision has been induced are resistant to interperitoneal administration of leptin; these mice quickly become obese. Other treated mice are more leptin sensitive; these mice are much slower to develop obesity. We hypothesize that this difference results from can variable excision efficiency of *Bbs8* in the hypothalamus. Further understanding of the development and progression of leptin resistance in BBS will result in better management of BBS-related obesity and the accompanying phenotypes of diabetes and hypertension.

2819W

A Knock-out Mouse Model of CMAMMA (Acsf3 Deficiency) Displays Neurological Phenotype and Methylmalonic Acidemia. *M. W. Epping[1], C. X. Wang[1], I. Manoli[1], J. L. Fraser[1], J. Gomez-Rodriguez[2], G. Elliot[3], J. L. Sloan[1], V. Hoffmann[4], P. L. Schwartzberg[2], C. P. Venditti[1].* 1) Genetics and Molecular Biology Branch, NHGRI, NIH, Bethesda, MD; 2) Genetics Disease Research Branch, NHGRI, NIH, Bethesda, MD; 3) Embryonic Stem Cells and Transgenic Mouse Core, NHGRI, NIH, Bethesda, MD; 4) Office of Research Services, Division of Veterinary Resources, NIH, Bethesda, MD, 20892.

Combined malonic and methylmalonic aciduria (CMAMMA [MIM 614265]) is a rare, autosomal recessive metabolic disorder resulting from deficient activity of ACSF3, a mitochondrial acyl-CoA synthetase with high specificity for malonate and methylmalonate. ACSF3 may catalyze the first step in the mitochondrial *de novo* fatty acid synthesis pathway, which has recently been identified in mammals. The highly variable clinical phenotypes reported continue to generate debate about whether patients exhibit clinical symptoms beyond the biochemical perturbations. While the full spectrum of the clinical phenotype caused by ACSF3 deficiency remains to be fully defined, some patients have manifested adult-onset neurological disease. In order to define disease pathophysiology, a deletion allele that removed the first coding exon of *Acsf3* was generated using homologous recombination in B6/129 ES cells. *Acsf3*-/- mice were born in Mendelian proportions and exhibit normal survival. No immunoreactive Acsf3 was detected in the liver, kidney, or brain of *Acsf3*-/- mice. When reared on regular diet, *Acsf3*-/- mice displayed the same growth pattern and weight as their heterozygous and wild type littermates. Plasma methylmalonic acid concentrations in the mutants were significantly higher than wild type mice (p=0. 0008), ranging from 25. 40 to 63. 16 μM, with heterozygous mice also displaying slightly elevated levels (p=0. 0165). To help address whether patients with CMAMMA need dietary restriction, *Acsf3*-/-mice were fed a high protein diet. The mutants maintained weights comparable to wild type littermates following two and four months of high protein challenge but did not display significant elevations of plasma methylmalonic concentrations compared to mice maintained on a regular chow diet. Vitamin B12 responsiveness was also assayed in the mice with minimal or no change in methylmalonic acid levels after administration. *Acsf3*-/-adult mice exhibit a clinical neurological syndrome of abnormal motor planning, coordination, and gait, hindlimb drag, tail ruddering. On the accelerating rotarod, *Acsf3*-/-mice demonstrate significantly impaired performance (p=0. 03, n=12 per genotype), with gait anomalies as described above. Exploration of post-translational modifications in the mice are ongoing and will help to elucidate the role of Acsf3. This novel mouse model serves to advance our understanding of CMAMMA disease pathophysiology, as well as a platform for developing therapeutic approaches.

2820T

Phenotypic characterization of a zebrafish model of Smith-Lemli-Opitz syndrome. *C. V. M. Cluzeau[1], K. M. Tabor[2], E. Leff[1], J. Picache[1], K. Burkert[1], C. A. Wassif[1], B. Feldman[2], H. A. Burgess[2], F. D. Porter[1].* 1) Program in Developmental Endocrinology and Genetics, NICHD, NIH, Bethesda, MD; 2) Program in Genomics of Differentiation, NICHD, NIH, Bethesda, MD.

Smith-Lemli-Opitz syndrome (SLOS) is an autosomal recessive disorder with a broad phenotypic spectrum. It is characterized by multiple malformations, cognitive impairment and abnormal behavior including autistic traits. Mutations of the 7-dehydrocholesterol reductase (*DHCR7*) gene, encoding the penultimate enzyme in the cholesterol biosynthetic pathway, result in decreased cholesterol and accumulation of the precursor 7-dehydrocholesterol (7DHC). Major questions still remain as to the pathophysiological mechanisms, mainly the contribution of toxic effects due to 7-dehydrocholesterol accumulation versus the effect of cholesterol deficiency, and the contribution of potential developmental versus physiological defects to the neurological function. Several rodent models have proven useful for elucidating some aspects of this syndrome, but they have limitations. To gain additional insights into the pathophysiological mechanisms and produce a model allowing rapid *in vivo* drug screening, we generated a zebrafish model of SLOS using Transcription Activator-like Effector Nucleases to disrupt *dhcr7* gene. We showed that mutant fish present with the characteristic accumulation of 7DHC and decreased cholesterol in both liver and brain from 2 weeks of age, with variable severity depending on the specific mutation. We further characterized the phenotype of fish carrying a frameshift mutation (NM_201330:c. 371_387delinsAAAT). *dhcr7*-deficient fish growth is delayed from 3 weeks of age. Mutant fish have a decreased ability to breed compared to controls. The yolk from eggs produced by mutant females contains high levels of 7DHC. Compared to the progeny of heterozygous individuals, the mutant progeny of the mutant females (maternal zygotic, MZ fish) have high levels of 7DHC during their first week of life. Interestingly, these MZ mutant fish present an abnormal acoustic startle response at 1 week of age compared to control fish, with decreased responsiveness and altered movement kinematics for non-Mauthner escapes. Additionally, adult mutant fish demonstrate decreased anxiety in the novel tank diving test. We will pursue the characterization of the motor and cognitive functions as well as the social behavior of adult *dhcr7*-deficient fish. Future work will focus on the neuroanatomical characterization of *dhcr7*-deficient fish brain to identify the causes of the abnormal behavior in the mutant fish.

2821F

Motor-phenotypical characterization and neuronal analysis of a transgenic rat model for DYT1 dystonia focusing on age dependent and brain region specific pathological key features from date of birth to old age of rats. *V. Gaiser[1], L. Lotzer[1], T. Roenisch[1], B. Fabry[1], Q. Welniarz[2], M. Walter[1], J. Huebener[1], M. Schaller[3], O. Rieb[1], T. Ott[1], K. Grundmann-Hauser[1].* 1) Dept. of Medical Genetics and Applied Genomics, Tuebingen, Germany; 2) Institut de Biologie Paris Seine, Paris; 3) Department of Dermatology, University of Tuebingen, Tuebingen, Germany.

Background: DYT-1-dystonia is an inherited autosomal-dominant disease characterized by involuntary movements due to a dysfunction of the CNS and is mainly caused by a deletion (ΔGAG) in the Tor1A gene. The pathophysiological features of this mutation are still not well understood. To analyze the cellular mechanism underlying this disease a transgenic rat model harboring the full length human mutant and wild type TorsinA gene including promoter and regulatory elements was analyzed. **Methods:** We performed motor phenotypical characterization with as well as analysis of different brain regions by immunohistochemistry and electron microscopy. Brains of young (postnatal day 0–1 month), middle (2-8 months) and old aged rats (12-20 months) where analyzed to follow up the brain region specific and age dependent key features of pathology. **Results:** We were able to determine the onset of nuclear envelope alteration in hΔETorA transgenic rat brains. Interaction of TorsinA with INM (inner nuclear membrane) compartments can be seen especially during INM assembly as well as membrane bleb formation could be observed already at early stages. Furthermore, we found manifestation of the nuclear envelope disintegrity of middle aged transgenic rats increasing with age in different hΔETorA transgenic rat lines. Only selective neuronal populations of brain regions with specific expression patterns of Torsine A and B seem to be affected by this pathology. We have evidence that a cellular shift of TorsinA towards tips of neurites discovered in rats of higher age does not take place in brain regions with early onset of NE pathological key features like in cell subtypes of striatum. Motor phenotypical testing revealed impairment of hΔETorA transgenic rats in the performance of simple motor reflexes and in learning of new complex motor tasks at old age of rats. **Conclusions:** We suggest that the mutated TorsinA causes NE pathology at early postnatal stages with a progressive cellular subtype specific phenotype and negative influence of mutated TorsinA to a molecular transport mechanism. Thus it appears that alterations of neurites in hΔETorA transgenic rats with increased age and brain regions with crucial functions for movement control like striatum and cerebellum being affected, leads to motor disability and highlights the importance of this model system for pharmaceutical approaches.

2822W

Mouse Models of Human Disease: How mouse data provides mechanistic insight into disease etiology and developmental therapeutics. *C. L. Smith, S. M. Bello, J. T. Eppig, The Mouse Genome Informatics Staff.* Mouse Genome Informatics, The Jackson Laboratory, Bar Harbor, ME.

Large scale forward and reverse mouse mutagenesis projects, the vast history of published mouse genetics and genomics from laboratory research and the genome-wide phenotyping efforts of the International Mouse Phenotyping Consortium (IMPC) have yielded an immense amount of data aimed at creating translational mouse models of human disease. Mouse Genome Informatics (MGI, www. informatics. jax. org) provides these integrated mouse phenotype data sets to enable complete correlation of known mouse phenotypes with human clinical signs and symptoms. MGI catalogs all known published and submitted mouse mutant mutations with genotypes annotated to phenotype and OMIM disease model descriptions, and includes links to other supporting gene information including sequence, polymorphism, spatiotemporal expression, genomic location, biochemical function and process, sub-cellular topology, and mammalian gene homology. Standardization of nomenclature and application of bio-ontologies, including the Mammalian Phenotype (MP) Ontology, ensure that data are consistently annotated, making robust data mining possible. Exome sequence data from patients with diseases with unknown etiology have provided several hundred new candidate genes for many human diseases. In addition, comparison of gene, phenotype and genetic disease data provided by mouse models aids in the refinement of potential therapeutic targets. We will show examples of MGI phenotype and disease data queries and results utilizing comparative mouse data that assist in identifying novel gene mutations and variations that are causative for human disease. We will also provide examples in which the biochemical data generated from mouse models and available at MGI has been used to infer gene function of human disease genes and define the gene's role as a therapeutic target.

2823T

Genomic copy number alterations in non-syndromic hearing loss. *R. Mingroni-Netto[1], E. L. Freitas[1], D. T. Uehara[1], M. T. B. M. Auricchio[1], J. Oiticica[2], A. G. Silva[1], C. Rosenberg[1], A. C. Krepischi[1].* 1) Dept of Genetics and Evolutionary Biology, Institute of Bioscience, University of Sao Paulo, Sao Paulo, Brazil; 2) Dept. of Otorhinolaryngology, Medical School, University of Sao Paulo , Sao Paulo, Brazil.

Due to tremendous genetic heterogeneity, the identification of genes that are related to hearing impairment has been challenging. In the absence of a clear etiology for the phenotype, recurrence risk estimates are often based on family segregation, and may be imprecise. We investigated by oligonucleotide array-CGH 100 patients presenting non-syndromic deafness, with either autosomic recessive (n=50) or autosomic dominant (n=50) presumptive patterns of inheritance. Rare copy number variants (CNV) were detected in 12 probands; four of the detected CNVs comprised genes previously associated with hearing loss (*POU4F3*, *EYA1*, *USH2A*, *BCAP31*), and were considered as causative, stressing the contribution of genomic imbalances to non-syndromic deafness. In six cases, segregation of the CNVs in pedigrees excluded them as causative. In one case, segregation could not be investigated and in other, a point mutation probably explains the phenotype. The findings showed that the presumptive patterns of inheritance were incorrect in at least two cases, impacting genetic counseling. Additionally, we report the first duplication reciprocal to the rare *ABCD1*, *BCAP31*, and *SLC6A8* contiguos deletion syndrome; as most microduplication syndromes, the associated phenotype is much milder than the respective microdeletion and, in this case, restricted to hearing impairment.

2824F

Sporadic Hidradenitis Suppurativa (acne inversa) and acne conglobata are not associated with gamma-secretase gene mutations. *N. M. Saiyed[1], U. Ratnamala[2,3], N. K. Jain[3], D. Jhala[3], M. Raveendrababu[3], B. C. Gorijala[4], R. S. Rawal[5], T. Y. Mehta[6], S. Nair[7], N. K. Chandramohan[8], D. G. Saple[9], F. M. Al-Ali[10], L. I. Al-Otaibi[10], EAL. Mohamed[10], M. Naveed[11], S. R. Bakshi[1], S. K. Nath[12], U. Radhakrishna[13].* 1) Biotechnology, Institute of Science, Nirma University, Ahmedabad, Gujarat, India; 2) Department of Pharmacology, Creighton University, Omaha, NE, USA; 3) Human Cytogenetics, Department of Zoology, School of Sciences, Gujarat University Ahmedabad, India; 4) Oncology Department, Krishna Institute of Medical Sciences, Secunderabad, India; 5) Department of dermatovenereology, Vadilal Sarabhai Hospital, Smt. Nathiba Hargovandas Lakhmichand (NHL) Municipal Medical College (NHLMMC), Ahmedabad, India; 6) Samarpan Medical & Research Organization on Skin, Modasa, India; 7) Department of Fetal Medicine, Lifeline Genetics and Research Centre, Lifeline Hospital, Adoor, Kerala, India; 8) Division of Surgical Oncology, Regional Cancer Centre, Medical College, Thiruvananthapuram, India; 9) Department of dermatovenereology, Brichkandi Hospital and Trust, Mumbai, India; 10) Dermatology Centre, Rashid Hospital, Dubai Health Authority (DHA), Dubai, United Arab Emirates; 11) H. H Sheikh Sultan Bin Khalifa Al Nahyan Humanitarian & Scientific Foundation, Fairmont, Dubai, United Arab Emirates; 12) Arthritis and Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, USA; 13) Department of Obstetrics and Gynecology, Oakland University William Beaumont School of Medicine, Royal Oak, MI, USA.

Hidradenitis suppurativa (HS) also known as acne inversa (AI) (MIM 142690), is a chronic, inflammatory and painful skin condition, characterized by swollen or inflamed lesions in the apocrine gland-bearing areas of the body which include: the buttocks, thighs, scalp, neck, nipples, armpits, groin, and anogenital region. Occurrence of HS is generally isolated to single patients; however a familial history has been reported in 30–40 % of patients revealing an autosomal dominant form of inheritance of the disease. Genetic predisposition plays a key role in HS and recently some progress has been made in identifying mutations of *NCSTN*, *PSEN1* and *PSENEN* genes on 1q23. 3, 19q13. 12 and 14q24. 2, respectively. To date, 13 gamma-secretase gene mutations have been identified including two recently identified *NCSTN* gene mutations by our group, accounting for almost 75% of Asian and European HS families. In contrast, only 1 *PSEN1* and 6 *PSENEN* mutations have been reported, suggesting that *NCSTN* plays a significant role in HS development. Remarkably, all of the variants have been identified in patients with a family history of the disease; however no mutations have been reported in sporadic patients. Genomic DNA from 94 unrelated HS patients, 1 with acne conglobata and 100 healthy sex-age-matched subjects with the same ethnic background were analyzed. Direct sequence analysis of amplified PCR fragments for all regions covering the entire coding sequence and splice junctions of the genes *NCSTN, PSENEN, PSEN1* and *PSEN2* with an automated DNA Analyzer (3730xl; AB) revealed 2 nucleotide substitutions. Nucleotide variation (Ile69Lys) was the only substitution present in the coding region of the *PSENEN* gene in one HS patient. In conclusion, the present study demonstrates genetic heterogeneity in the HS disorder and suggests the involvement of novel genes. It also suggests that gamma-secretase gene mutations are not a common cause of sporadic patients with HS and or acne conglobate; otherwise, it is possible the disease is associated with variations in promoters.

2825W

UNIPARENTAL ORIGIN IN OCUOCUTANEOUS ALBINISM. *W. Cardenas, M. Lattig.* Biological Sciences, Centro de Investigación en Enfermedades Genéticas Humanas, Bogota, Colombia.

Oculocutaneous albinism (OCA) is a group of autosomal recessive disorders of the melanin biosynthesis pathway. OCA has been classified molecularly in seven different types OCA1 to OCA7. In previous results our group reported a high frequency of the p. G47D (rs61753180) mutation in the TYR gene (MIM*606933) in our colombian population with OCA1 disorder (MIM #203100). This mutation has been reported mainly in Israel and Puerto Rico populations and a migration pattern has been postulated for the p. G47D from Sephardic Jews that migrated to Marruecos, Canary Island and Puerto Rico. Our objective was to assess if the individuals with the p. G47D mutation in our country have a Jewish ancestry. We evaluated the mitochondrial DNA haplogroups and Y-STR haplogroups in nine individuals with the p. G47D mutation in the TYR gene. We amplified and sequence the D-loop of the mitochondrial DNA in all individuals and genotyped the Y-filer STR systems in five male individuals. We found that our entire sample belong to the Amerindian groups in the mitochondrial DNA, represented by the A, C, D and X haplogroups. We did not find any of the haplogroups reported for Jewish ancestry. However, we did find four different Y-STR haplogroups the R1a-M17, G2a, R1b1a-M269 and one African haplogroup. The possible origin of the R1a-M17 was in the Northeast Europe and in the Mediterranean. The G2a has been described in recent Caucasian and European individuals and was derived from Africa. To date the R1b1a is distributed in Western Europe and is the most common form in Europe closely associated with the diffusion of Indo-European language, but nothing was described with a Jewish origin. Our results show that our samples are in accordance with known history of migrations and admixing among European colonizers and Amerindian groups along the last three centuries. We could not find a direct relationship between Jewish haplogroups and the p. G47D mutation in our study sample. Nevertheless, this study is a first approach towards understanding the origin of the p. G47D mutation in the Colombian population. Key Words:Oculocutaneous albinism, ethnicity, mtDNA haplogroups, Y-STR haplogroups, Jewish.

2826T

Novel *FGF10* Mutation in Autosomal Dominant Aplasia of Lacrimal and Salivary Glands. *J. -W. Kim[1,3], F. Seymen[2], M. Koruyucu[2].* 1) Pediatric Dentistry, Seoul National University School of Dentistry, Seoul, South Korea; 2) Department of Pedodontics, Faculty of Dentistry Istanbul University, Istanbul, Turkey; 3) Department of Molecular Genetics & Dental Research Institute School of Dentistry, Seoul National University, Seoul, Korea.

Aplasia of lacrimal and salivary glands (ALSG) is a rare autosomal dominant inherited disease, characterized by aplasia, atresia or hypoplasia of the lacrimal and salivary systems with variable expressivity. The purpose of this study was to identify genetic etiology of an ALSG family. We recruited a Turkish family with ALSG and performed a mutational analysis, based on the candidate gene approach, to clarify the molecular genetic etiology. The candidate gene sequencing of the *FGF10* gene identified a novel heterozygous nonsense mutation (c. 237G>A, p. Trp79*) in the exon 1. The identified novel mutation would result in a haploinsufficiency of the FGF10, because of nonsense-mediated mRNA decay caused by a premature stop codon. This report further confirms that ALSG is caused by the haploinsufficiency of functional FGF10. This work was supported by grants by the National Research Foundation of Korea (NRF) grant funded by the Korea government (2014R1A2A1A11049931) and the Korea Science and Engineering Foundation (KOSEF) through the Biotechnology R&D program (2014020550).

2827F

Genome-wide linkage analysis of non-syndromic cleft lip with or without cleft palate in a large multigenerational Indian family. *U. Ratnamala[1], N. K. Jain[1], S. S. Chettiar[1], D. Jhala[1], S. Beiraghi[2], S. K. Nath[3], U. Radhakrishna[4].* 1) Department of Life sciences, University school of Sciences, Gujarat University, Ahmedabad, Gujarat, India; 2) Division of Pediatric Dentistry, University of Minnesota, Delaware Street SE, Minneapolis, MN 55455, USA; 3) Arthritis and Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, USA; 4) Department of Obstetrics and Gynaecology, Oakland University William Beaumont School of Medicine, Royal Oak, MI, USA.

Non-syndromic cleft lip with or without cleft palate (NSCL/P) is considered the most common severe craniofacial disorder having both an environmental and genetic origin. NSCL/P abnormalities are associated with increased morbidity and mortality in infancy, childhood, and adulthood. Nevertheless, the etiology of this disorder remains poorly defined. Earlier genetic studies by our group on NSCL/P has identified two loci on chromosome 13q33. 1-34 (Am J Hum Genet. 2006; 79(3):580-5) and a locus at 18q21. 1 (Am J Hum Genet. 2007; 81(1):180-8) in two large multiplex autosomal dominant NSCL/P families (UR17 and UR18). Also many susceptibility NSCL/P loci have been reported on different chromosomal regions. In an attempt to further characterize the causative factors of NSCL/P, a productive collaboration with clinicians and research scientists from India, Europe, Iran, other Middle Eastern countries and the USA has been established. This collaboration has resulted in the acquisition of samples from 50 multiplex NSCL/P families. We have recently conducted another genome-wide linkage analysis, usingOMNI-1S bead array Chips on a large multigenerational non-syndromic autosomal dominant, non-consanguineous Indian NSCL/P family (UR411) with 12 affected and 31 unaffected members. The study identified a NSCL/P susceptibility locus at 1q25-42 (non-parametric linkage [NPL] = 16. 61, p = 0. 0014). These results were also supported by multipoint parametric linkage analysis. A maximum multipoint LOD score of 3. 57 was detected assuming an autosomal dominant mode of inheritance with reduced penetrance. Haplotype analysis with informative crossovers enabled mapping of the NSCL/P locus to a region of χ22. 914 cM between single-nucleotide polymorphisms rs12064317and rs6668056. This genomic interval on 1q25-42 contains the known candidate gene IRF6. To our knowledge this is the first Indian NSCL/P family with a susceptibility locus mapping to chromosome 1q25-42 region. Mutation analysis of candidate genes is progressing.

2828W

Investigation into Clinical Significance of c-FLIP and Jun-B Expression in Psoriatic Patients. *N. Salleh[1], N. Shamsudin[2], H. Lai[1], R. Lai[1], I. Mustafa[1].* 1) Faculty of biomedical and Health Sciences, Universiti Selangor, Selangor Malaysia; 2) Department of Medicine, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Selangor Malaysia.

Psoriasis is a chronic T cell-mediated inflammatory skin disease. Studies have shown that angiogenesis plays an important role in the pathogenesis of psoriasis. In the present study, we investigated the involvement of c-FLIP and Jun-B protein expression in the possible mechanisms of psoriasis. The hyperproliferation of keratinocytes observed in psoriasis prompted us to evaluate c-FLIP and Jun-B expression on biopsies collected from involved and uninvolved skin of 42 patients with active plaque-type psoriasis with respect to healthy skin. We analyzed the expression of C-FLIP and Jun-B at transcript and protein levels by quantitative RT-PCR and tissue microarray based immunohistochemistr. We demonstrated the expression of both genes c-FLIP and Jun-B in a hyperproliferative skin condition not related to neoplastic transformation. Interestingly, we observed that c-FLIP and Jun-B is not expressed in healthy skin, but it becomes detectable in non-lesional areas and it is even more expressed in lesional psoriatic skin. In addition, we found that Jun-B expression is correlated to the rate of keratinocyte proliferation and activation compared to c-FLIP. Hence, our observations indicate Jun-B as a new possible player, involved in the development and/or maintenance of the hyperproliferative state of psoriatic keratinocytes. Our study confirms distinct prognostic relevance of c-FLIP and Jun-B protein expression levels in the psoriasis patients and identifies an association of high Jun-B levels with elevated expression of Jun-B target genes and markers for infiltrating immune cells.

2829T

Development of refractive errors – what can we learn from retinal dystrophies? *V. J. M. Verhoeven[1,2,3], M. Hendriks[1,2,4], G. H. S. Buitendijk[1,2], J. R. Polling[1,5], M. A. Meester-Smoor[1], A. Hofman[2,6], M. Kamermans[7,8], L. I. van den Born[4], C. C. W. Klaver[1,2], RD5000 Consortium.* 1) Dept. Ophthalmology, Erasmus Medical Center, Rotterdam, The Netherlands; 2) Dept. Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands; 3) Dept. Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands; 4) The Eye Hospital, Rotterdam, The Netherlands; 5) University of Applied Sciences Utrecht, Faculty Eyecare, Utrecht, The Netherlands; 6) Netherlands Consortium for Healthy Ageing, Netherlands Genomics Initiative, The Hague, The Netherlands; 7) Netherlands Institute for Neuroscience, Amsterdam, the Netherlands; 8) Department of Neurogenetics, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands.

Purpose: Myopia (nearsightedness) is widely recognized as a multifactorial, complex genetic disorder. The global incidence of myopia is rising, and reaches epidemic proportions in some parts of the world. Previously, multiple loci for refractive phenotypes were identified. Still, it is unknown which retinal cells are involved in the retina-to-sclera signaling cascade causing myopia. Inherited retinal dystrophies (RD) are characterized by dysfunction of a single retinal cell type. We studied the association between affected cell type, causal gene and refractive error in RDs. **Methods:** We included patients with RD (N=302; RPE related, cone dominated (CD), rod dominated (RP), bipolar cell related dystrophies) from two ophthalmogenetic centers in the Netherlands. Distributions and mean spherical equivalent (SE) of each diagnosis and causal gene were calculated, and risks of myopia and hyperopia were evaluated using logistic regression. The population-based Rotterdam Study-III and ERF Study (N=5,550) served as a reference. **Results:** Bipolar cell dysfunctions were associated with the highest risk of SE (OR high myopia 239. 7; OR mild hyperopia 263. 2, both P<0. 0001; SE -6. 86 D [SD 6. 38]); followed by CD (OR high myopia 19. 5, P<0. 0001; OR high hyperopia 10. 7, P=0. 033; SE -3. 10 D [SD 4. 49]); RP (OR high myopia 10. 1, P<0. 0001; OR high hyperopia 9. 7, P=0. 001; SE -2. 27 D [SD 4. 65]); and RPE related dystrophies (OR low myopia 2. 7; P=0. 001; OR high hyperopia 5. 8; P=0. 025; SE -0. 10 D[SD 3. 09]). Mutations in *RPGR* (SE -7. 63 D [SD 3. 31]) and *CACNA1F* genes (SE -5. 33 D [SD 3. 10]) coincided with the highest degree of myopia; in *CABP4* (SE 4. 81 D [SD 0. 35]) with the highest degree of hyperopia. **Conclusions:** Refractive errors, in particular myopia, are common in RD. The bipolar synapse, and the inner and outer segments of the photoreceptor may serve as predilection sites for myopia development.

2830F

A Splicing Mutation in *VPS4B* Causes Dentin Dysplasia I. *F. Xiong[1], Q. Yang[1], D. Chen[2], C. X. Liu[1], Q. X. Yu[1], Y. H. Liu[1,3], J. Z. Liu[2], K. Y. Li[2], L. F. Zhao[4], Y. H. Ye[1], H. Zhou[2], L. L. Hu[1], Z. H. T[5], X. Shang[1], L. T. Zhang[5], X. F. Wei[1], D. N. Chen[1], W. J. Zhou[1], D. R. Li[6], W. Q. Zhang[4], X. M. Xu[1].* 1) Department of Medical Genetics, Southern Medical University, Guangzhou, Guangdong, China; 2) School of Stomatology, Zhengzhou University, Zhengzhou, China; 3) Department of Prenatal Diagnosis Center, Maternal and Child Health Hospital, Dongguan, China; 4) Department of Cell Biology, School of Basic Medical Sciences, Southern Medical University, Guangzhou, China; 5) Department of Stomatology, Nanfang Hospital, Southern Medical University, Guangzhou, China; 6) Department of Forensic Science, School of Basic Medical Sciences, Southern Medical University, Guangzhou, China.

Dentin dysplasia I (DDI) is a genetically heterogeneous autosomal-dominant disorder characterized by rootless teeth with abnormal pulpal morphology. Using a cohort of a large Chinese family with 13 normal members and 10 DDI patients, we mapped to a 9. 63 Mb candidate region for DDI on chromosome 18q21. 2-q21. 33. We then identified a mutation IVS7+46C>G which derived a novel donor splice site in intron 7 of the *VPS4B* gene with co-segregation of all 10 affected individuals in this family. The aberrant transcripts encompassing a newly inserts of 45-bp in size were detected in gingival cells from affected individuals. Protein structure prediction showed that a 15-amino acid insertion altered the ATP-binding cassette of VPS4B. The mutation resulted in significantly reduced expression of mRNA and protein by semiquantitative molecular analysis, and altered subcellular localization of VPS4B by confocal image analysis, indicating a loss of function of VPS4B. Using human gingival fibroblasts, the *VPS4B* gene was found to be as an upstream transducer linked to Wnt/beta-catenin signaling regulating odontogenesis. Furthermore, knock-down of *vps4b* in zebrafish recapitulated reduction of tooth size and absence of teeth similar to the tooth phenotype exhibited in DDI index cases, and the zebrafish mutant phenotype could be partially rescued by wild-type human *VPS4B* mRNA. We also observed that *vps4b* depletion in the zebrafish negatively regulates the expression of some major genes involved in odontogenesis. The study thus identifies the *VPS4B* as a disease-causing gene for DDI, which is one of the important contributors to tooth formation through the Wnt/beta-catenin signaling pathway.

2831W

Ehlers-Danlos syndrome, hypermobility type is linked to chromosome 8p22-8p21. 1 in an extended Belgian family. *P. J. Coucke, D. Syx, S. Symoens, W. Steyaert, A. De Paepe, F. Malfait.* Center for Medical Genetics Ghent, University Hospital Ghent, Ghent, O-VL, Belgium.

Joint hypermobility is a common finding in the general population with epidemiologic studies showing its presence in over 10% of Caucasians. In a subset of individuals however, it causes a range of clinical manifestations, mainly affecting the musculoskeletal system. Joint hypermobility often presents as a familial trait and is shared by several heritable connective tissue disorders, including the hypermobility subtype of the Ehlers-Danlos syndrome (EDS-HT) or benign joint hypermobility syndrome (BJHS). These hereditary conditions provide unique models for the study of the genetic basis of joint hypermobility. Nevertheless, these studies are largely hampered by the great variability in clinical presentation and the often vague mode of inheritance in many families. Here, we performed a genome-wide linkage scan in a unique three-generation family with an autosomal dominant EDS-HT phenotype, and identified a linkage interval on chromosome 8p22-8p21. 1, with a maximum two-point LOD score of 4. 73. Sequencing a number of interesting positional candidate genes such as *BMP1*, *LOXL2*, *CSGALNACT1* and *SLC39A14* revealed no disease causing mutation. Therefore, we applied the whole exome sequencing strategy in an attempt to unravel the underlying genetic cause within this candidate locus. This latter approach revealed the presence of a unique missense variant p. (His211Gln) in the *LZTS1* gene, encoding leucine zipper, putative tumor suppressor 1, in the candidate region. Although *LZTS1* itself was not linked to joint hyperlaxity, disruption of the *ProSAP2* gene, encoding a protein interacting with LZTS2 (LAPSER1) resulted in a syndrome with mild intellectual disability and joint laxity. Subsequent analysis of 230 EDS-HT/BJHS patients resulted in the identification of three additional rare *LZTS1* variants. This is the first reported genome-wide linkage analysis in an EDS-HT family, thereby providing an opportunity to identify a new disease gene for this condition.

2832T

Molecular diagnosis of F8 gene in sever hemophilia A in the Iranian population. *A. Ebrahimi[1,2], S. Mansouri derakhshan[1,2], M. r. Ranjouri[3], M. Shekari Khaniani[1,2].* 1) Genetics and Molecular Medicine, Tabriz university of medical science, Tabriz, Iran; 2) Hematology & Oncology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran; 3) Medical Genetic Department, Faculty of Medicine, Zanjan university of medical science, Zanjan, Iran.

Background: Hemophilia A (HA) is an inherited X-linked disorder in blood coagulation cascade that caused by a variety of mutations in F8 gene. The most common mutation in sever HA is Intron 22&1 Inversion, exon deletion and point mutation in exon 14, respectively. **Objective:**This study aimed to detect these known DNA mutations in severe HA patients in Azeri Turkish population for the first time in North West of Iran. **Material and Methods:** Peripheral blood of 50 severe HA patients was collected and genotyped by Inverse-Shifting PCR (IS-PCR) method for intron 22 Inversion, long PCR for Intron 1 Inversion, multiplex–PCR for exon deletions and PCR- sequencing for exon 14 mutations. **Results:** Our study showed that 22 patients (44%) had Inversion 22, one individual (2%) had Intron 1 Inversion, no cases of exon deletions and one case presented mutation in exon 14 (2%). In this population, 52 % of patients don't carried mutations in the analyzed regions, but might have mutations in the other regions of gene. **Discussion:** We concluded that *F8* intron 22 inversion is causing mutation for nearly 50% in severe HA cases in Azeri Turkish population similar to other populations and IS-PCR a robust and rapid method for genetic analysis of sever HA patients in developing countries.

2833F

Exome sequencing identifies SPG3A and OTOF gene mutations in patients with Pure Hereditary Spastic Paraplegia. *S. Majid[1], D. Monies[1], S. Hagos[1], A. Qahtani[1], K. Ramzan[1], A. Abulaban[2], H. AlDossari[1], B. Meyer[1], S. Bohlega[2].* 1) Department of Genetics, KFSHRC, RIYADH; 2) Department of Neurosciences, KFSHRC, RIYADH.

Hereditary spastic paraplegias (HSP) comprises of clinically and genetically heterogeneous group of neurodegenerative diseases. SPG3A, a pure form of HSP which is characterized by progressive bilateral, mostly symmetric spasticity and weakness of legs, with diminished vibration sense caused by degeneration of the corticospinal tracts and dorsal columns. SPG3A encodes for protein called Atlastin (ATL1), a dynamin-related GTPase, which play a significant role in the assembly and functions of nuclear protein complexes. In this study, we describe a native Arabic consanguineous family comprising of four affected individuals where progressive spasticity started at age 2 to 8 years with no cognitive impairment or dysmorphic features. Two of the affected presented with severe dysarthria, disabling spasticity and they remained ambulatory until the third decade of life through multiple corrective surgeries. Using combined homozygosity and exome sequencing approaches, we identified two novel pathogenic mutations in ATL1 and OTOF gene that underlies HSP and hearing loss in this family respectively. The novel mutations were not detected in >600 ethnically matched normal chromosomes. By*in silico*analysis, OTOF mutation was predicted to damage protein structure or function. Most of the reported SPG3A cases are caused by heterozygous mutations in ATL1 gene. We have identified a homozygous non-sense mutation p. K330X in the ATL1 gene which is the causative gene for this rare autosomal recessive childhood HSP with complete penetrance for the mutation in the homozygous state. This study delineates the genetic spectrum spectrum of rare autosomal recessive form of Spastic Paraplegia 3A.

2834W

Informativeness of St14 VNTR marker for carrier detection and prenatal diagnosis of hemophilia A in Iranian families. *S. Mansouri derakhshan[1,2], A. Ebrahimi[1, 2], P. Aob[3], M. Shekari Khaniani[1,2].* 1) Department of Medical Genetics, School of Medicine, Tabriz university of medical science Tabriz, Iran; 2) Hematology & Oncology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran; 3) Medical Genetic Department, Faculty of Medicine, Zanjan university of medical science, Zanjan, Iran.

Background: Due to the large size and hetrogenous mutations of F8 gene, direct mutation analysis is not practical. Therefore, indirect allele characterization using DNA markers linked to the *F8* locus appears to be more available approach for hemophilia A carrier detection. **Objective:** The aim of this study was to investigate the usefulness of an extragenic variable number of tandem repeat (VNTR) St14 (DXS52) on the human X-chromosome in linkage analysis to develop an informative and accurate carrier detection and prenatal diagnosis. **Material and Methods:**Hemophilia A patients, 104 male and female non-hemophiliac family members, and 35 potential female carriers were studied. They were sampled from North West of Iran (Azeri Turkish population). Molecular analysis for St14 marker was performed by long-distance Polymerase chain reaction (PCR) method. **Results:** 12 different alleles ranging from 700 to 1800 bp in size were observed. The most common allele corresponded to 1100 bp amplified fragment (18%) whereas the alleles with amplified fragment of 1700 bp (1. 6%) and 1800bp (1. 2%) were less prevalent among Iranian Azari Turkish population. Compared to Caucasians, this population showed a markedly higher occurrence of medium molecular weight fragments and a relatively low occurrence of high molecular weight fragments. **Discussion:**The observed heterozygosity among female subjects was 61% Linkage analysis using St14 marker in 35 suspected hemophilia A carriers revealed (9) 26% to be carrier. In consequence, St14 VNTR analysis by PCR should prove to be a useful tool in the genetic diagnosis of hemophilia A carriers and prenatal diagnosis in Azeri Turkish population. However, the use of other polymorphic markers such as *Bcl1*, *XbaI* and *HindIII*/RFLP and dinucleotide repeats (STR) in combination with VNTR St14 should raise the information of indirect markers for clinical diagnosis.

2835T

Oculocutaneous albinism in a boy with a complex phenotype, and homozygosity for an ancestral block on chromosome 11 bearing a double set of *TYR* mutations. *S. L. Everhart[1], N. Butala[2], W. R. Heymann[2], S. M. Manders[2], J. Chiang[3], A. Levin[4], R. E. Schnur[1].* 1) Division of Genetics, Department of Pediatrics, Cooper University Health Care, Camden, NJ; 2) Division of Dermatology, Department of Medicine, Cooper University Health Care, Cooper Medical School of Rowan University, Camden, NJ; 3) Casey Eye Institute, Oregon Health Sciences University, Portland, OR; 4) Wills Eye Hospital, Philadelphia, PA.

Oculocutaneous Albinism Type 1 (OCA1) is an autosomal recessive disorder due to mutations of tyrosinase (*TYR*) on chromosome 11q14. 3. OCA1A patients show nearly complete lack of pigmentation due to absence of tyrosinase activity; OCA1B patients, in contrast, have some residual TYR activity with milder pigmentary effects. We report a 10 year-old Hispanic (non-Puerto Rican) boy with features of OCA, including a complete absence of pigment. Non-OCA-related features include macrosomia, visceromegaly, history of pulmonary infections, easy bruising and learning disability. A distant relative on each side of the family was reported to have albinism. Chromosome microarray analysis detected a single, contiguous, 16. 15 Mb region of homozygosity on chromosome 11q13. 5-q14. 3 including the *TYR* locus [arr 11q13. 5q14. 3(76,447,458-92,594,422)x2 hmz], allowing us to target *TYR* as the likely etiology of his OCA. Subsequent *TYR* sequencing revealed homozygosity for a double set of variants, each previously individually reported to be pathogenic: p. R299H:c. 896G>A and L216M:c. 646T>A, consistent with our patient's OCA1A phenotype. Biparental inheritance of the double mutation set was confirmed. Thus, the single block of homozygosity likely represents very distant ancestral relatedness of the parents rather than segmental isodisomy. Interestingly, although the patient is homozygous for these two discrete missense changes, more distal SNPs within *TYR* are heterozygous, suggesting that an intragenic crossover event occurred at some point. In summary, our patient has OCA1A due to biparental transmission of a double set of *TYR* missense mutations that are carried on a larger ancestral homozygous block of chromosome 11. Our patient's "albinism-plus" phenotype may be the result of homozygosity for deleterious changes of other regional genes, although correlation of his non-albinism features with specific genes could not be made.

2836F

Hereditary motor and sensory neuropathy with proximal dominancy in the lower extremities, urinary disturbance, and paroxysmal dry cough maps to chromosome 1p13. 3-1q23. *S. Miura[1], S. Sano[1], K. Kosaka[2], R. Fujioka[3], H. Saitsu[4], T. Taniwaki[1], K. Yamamoto[5], H. Shibata[2].* 1) Department of Medicine, Kurume University School of Medicine, Kurume City, Fukuoka Prefecture, Japan; 2) Medical Institute of Bioregulation, Kyushu University, Fukuoka City, Fukuoka Prefecture, Japan; 3) Department of Food and Nutrition, Beppu University Junior College, Beppu City, Oita Prefecture, Japan; 4) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama City, Kanagawa Prefecture, Japan; 5) Department of Medical Chemistry, Kurume University School of Medicine, Kurume City, Fukuoka Prefecture, Japan.

We previously reported a clinically new type of autosomal dominant disorders of motor and sensory neuropathy with proximal dominancy in the lower extremities, urinary disturbance, and paroxysmal dry cough (Miura et al. J Neurol Sci 2008). The cardinal clinical features of the disease were as follows: i) slow progression; ii) muscular weakness with dominancy in the proximal portion of the lower limbs; iii) sensory involvement with dominancy in the lower limbs; iv) areflexia; v) fine postural tremor; vi) painful muscle cramping; vii) elevated serum creatine kinase level; viii) urinary disfunction; and ix) recurrent paroxysmal dry cough. To identify the causative mutation for the disease, we studied a Japanese pedigree with the disease consisted with 19 family members including 9 patients in five generations. On multipoint linkage analysis of the pedigree by randomly selected 1,752 cleaned SNPs with an average interval of 2 cM using GeneHunter implemented in easyLINKAGE package v5. 08, maximum LOD scores of > 2. 0 was obtained on 1p13. 3-q23 (LOD = 2. 24). The exome sequencing upon five patients and one healthy relative from the pedigree revealed 2526 patient-specific single nucleotide variants (SNVs). By the filtering processes using the information from the public SNP database and our linkage analysis, we identified 19 candidate SNVs in 14 genes. Moreover, we selected 3 candidate SNVs in 2 genes through validation by Sanger sequencing, confirming cosegregation, and genotyping 520 healthy Japanese individuals. Through the functional predictions of the cosegregated SNVs, we currently focus on one SNV located in *IQGAP3* which is associated with neurite outgrowth. The SNV possibly alters the splicing of *IQGAP3* and also located within a non-coding RNA which may affect the expression of IQGAP3. This study was approved by the Ethics Committees of Kurume University School of Medicine and Kyushu University, Faculty of Medicine.

2837W

Adult onset motor neuron disease predominant Triple 'A' syndrome in South India- ? due to a novel mutation. *D. Selvarajan[1], P. Jayapal[2], S. Krishnan[3], NJ. Elakatt[4].* 1) Internal Medicine, Vinayaka Mission's Medical College, , Karaikal, Pondicherry, India; 2) Internal Medicine, Madras Medical College, Chennai, Tamilnadu, India; 3) Internal Medicine, Presence Saint Francis Hospital, Evanston, Illinois; 4) Strand Centre for Genomics and Personalised testing, Bengaluru, Karnataka, India.

Triple A syndrome (Allgrove syndrome)is a rare autosomal recessive disease involving Chr. 12q13producing defective AAAS gene, encodes a protein, Aladin, which is a member of the WD-repeat family of regulatory proteins involved in normal development of nervous system. The protein consists of 546 amino acids and the gene is made up of 16 exons. Triple A syndromehas myriad of features mainly Addison's disease, Achalasia cardia and Alacrimia followed by peripheral neuropathy, Autonomic dysfunction, Microcephaly, Mental retardation, Ataxia ,etc. This is usually a childhood disease. We present a 20 year old adult female from South India who was born of second degree consanguity presented with dysphagia and severe small muscle wasting for the past three years and elsewhere diagnosed as motor neuron disease. But on further evaluation she revealed presense of Addison's disease and Type 11 Achalasia cardia and alacrimia and other above mentioned features. Mutation analysis was planned and using the Patient's DNA, Nextera technology was utilized in the enrichment of target sequences, which includes coding exons (>95%) and splice junctions of 550 genes. The generated library was subjected to next generation sequencing on the Illumina NGS platform (MiSeq). Variations were identified using the STRAND® NGS software and interpreted using StrandOmics. Of the 550 genes, 460 had >85% coverage at 20X. The patient harbored a novel homozygous variant (c. 545+2T>G) in the AAAS gene. The variant(NM_015665 c. 545+2T>G, chr12:53708533A>C)appears in the essential splice site region of intron 6. Online splice site prediction tools such as Alternative Splice Site Prediction and Berkeley Drosophila Genome Project have predicted the variant to disrupt the wild type splice site and favour an alternative splice site 19 nucleotide bases into intron 6. This causes the insertion of additional 19 nucleotide bases into the coding sequence. The online tool ORF finder has predicted that this variant could cause a frameshift and an eventual termination (p. Ser182ArgfsTer30) leading to a final peptide product of only 210 amino acids instead of the original 546. Her parents were tested to confirm heterozygosity (carriers) and younger brother had same disease with same homozygous mutation. Early diagnosis will lead to effective management by treating Addison's disease, Artificial tear drops and surgical correction of Achalasia cardia. Genetic counseling prevents further occurrences.

2838T

Novel Gene Discovery Using Whole Exome Sequencing and Linkage Analysis in a Large Family with Autosomal Dominant Ataxia. *S. Cho[1], M. Seong[1], H. Kim[2], S. Seo[1], S. Lee[1], J. Kim[1], B. Jeon[2], S. Park[1].* 1) Dept Laboratory Medicine, Seoul National Univ Hosp, Seoul, South Korea; 2) Departments of Laboratory Neurology, Seoul National Univ Hosp, Seoul, South Korea.

The autosomal dominant spinocerebellar ataxias (SCAs) represents a group of neurodegenerative diseases, characterized by a slowly progressive incoordination of gait. The genetic background of SCAs is heterogeneous, having 40 or more genetic subtypes. In addition, some SCAs are caused by sequence variants, but others by nucleotide repeat expansions. Here, we investigated genetic etiology using genomic approach in a large family with autosomal dominant spinocerebellar ataxia patients. We collected peripheral blood from 14 family members including 5 patients. Whole exome sequencing (WES) was done for 4 patients and 2 normal members. Sequence variants were called and filtered to find out pathogenic mutations located within coding regions and their flanking intronic regions. Also, we performed exome-wide linkage analysis using variants information discovered in WES to find out coding or non-coding nucleotide repeat expansion. Sanger sequencing and repeat-primed PCR were done for each repeat regions to reveal abnormal expansion. No sequence variant consistent with segregation pattern was identified in this family. However, we identified 10 genomic regions with highest LOD score. In these regions, there were 20 genes and 10 tri- to hexa-nucleotide repeat regions. Interestingly, *TGM6* (genetic cause of SCA30) and *NOP56* (SCA36) were included also. WES and exome-wide linkage analysis may be a useful tool to discover novel gene in genetically heterogeneous disease caused by different genetic mechanisms like autosomal dominant spinocerebellar ataxia.

2839F

A Newly Recognized Intellectual Disability Disorder Caused by Variants in *TELO2,* a gene encoding a component of the TTT complex. *J. You[1,2], N. Sobreira[2], D. Grange[3], D. Gable[1,2,7], J. Jurgens[1,2], N. Belnap[8,9], A. Shiniard[8,9], M. Huentelman[8,9], I. Schrauwen[8,9,10], R. Richholt[8,9], S. Vallee[11], M. Palko[11,12], D. Valle[2,5,6], M. Armanios[2,7], J. Hoover-Fong[2,4,6].* 1) McKusick-Nathans Institute of Genetic Medicine, The Johns Hopkins School of Medicine, Baltimore, MD; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 3) Department of Pediatrics, Washington University, St. Louis, MO 63110, USA; 4) Greenberg Center for Skeletal Dysplasias, McKusick-Nathans Institute of Genetic Medicine the Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA; 5) Center for Inherited Disease Research, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21224, USA; 6) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 7) Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD, USA; 8) Dorrance Center for Rare Childhood Disorders, TGen, Phoenix, AZ 85012, USA; 9) Neurogenomics Division, TGen, Phoenix, AZ 85004, USA; 10) Department of Medical Genetics, University of Antwerp, 2650 Antwerp, Belgium; 11) Dartmouth-Hitchcock Medical Center, 1 Medical Center Dr, Lebanon, NH 03766; 12) Pediatrics and Pathology department, Geisel School of Medicine at Dartmouth, 1 Medical Center Dr, Lebanon, NH 03766.

The proteins encoded by *TELO2*, *TTI1*, and *TTI2* interact to form the TTT complex, a co-chaperone system for maturation of the phosphatidylinositol 3-kinase-related protein kinases (PIKKs). Here we report 6 individuals from 4 unrelated families with compound heterozygous mutations in *TELO2.* These individuals share overlapping features including global development delay, intellectual disability, dysmorphic facies, microcephaly, abnormal movements and abnormal auditory and visual function. WES of three affected individuals in Family 1 identified rare compound heterozygous missense variants in *TELO2,* C367F (c. 1100G>T) and D720V (c. 2159A>T). Subsequently, using GeneMatcher (www. genematcher. org), we identified affected individuals in 3 additional families all with compound heterozygous variants in *TELO2*. The proband in Family 2 is compound heterozygous for the C367F variant (found in Family 1) and the novel variant V766M (c. 2296G>A). The proband in Family 3 carries the P260L (c. 779C>T) and R609H (c. 1826G>A) variants and that in Family 4 is a compound heterozygote for the D720V (found in Family 1) and a complex allele, Q172X (c. 514C>T) with IVS16+1G>A, which encodes a markedly truncated protein from a transcript that is subject to nonsense mediated mRNA decay. All variants are rare and predicted to be damaging by Polyphen-2 and SIFT. TELO2, TTI1 and TTI2 protein levels were reduced to 34%, 27% and 15% of controls in lymphoblasts and fibroblasts from affected individuals in Family 1. Fibroblasts from the proband in Family 2 showed similar reductions. Despite this, PIKK functions were normal in ionizing irradiation, mitomycin C treatment, and telomere length assays in Families 1 and 2. Under stress conditions (17-AAG, treatment), TTT protein expression decreased more in Family 1 individuals' fibroblasts than in control fibroblasts (TELO2, TTI1, and TTI2 decreased by 11. 9%, 31. 6 %, and 20%, respectively). This further reduction in TTT proteins was associated with decreased ATM and DNA-PK levels in patients in Family 1 (34% and 56% of control fibroblasts, respectively). ATR and mTOR expression was not reduced under these conditions. This result indicates that with Hsp90 inhibition, the reduction of the TTT complex resulting from *TELO2* variants decreases abundance of some PIKK proteins. In summary, our results suggest that these *TELO2* missense variants result in loss-of-function, perturb TTT complex stability and cause a newly recognized autosomal recessive form of ID.

2840W

The Expanding Genetic and Phenotypic landscapes of Deafness Genes. *H. Azaiez[1], K. T. Booth[1], C. M. Sloan-Heggen[1], M. Beheshtian[2], H. Najmabadi[2], R. J. H Smith[1].* 1) Dept Otolaryngology, Univ Iowa, Iowa City, IA; 2) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran.

The genetic, allelic and clinical heterogeneity associated with hereditary deafness are well established, with over 90 genes and 2000 mutations currently identified to cause non-syndromic hearing loss (NSHL) and a wide spectrum of phenotypic diversity in onset, severity, stability and audioprofiles. Less established although rapidly expanding, are genotype-phenotype relationships. This expansion directly correlates with the advancement in sequencing technologies. Here we expand the phenotypic spectrum associated with three genes; *PTPRQ*, *CEACAM16* and *COL11A1* using a combination of linkage analysis, custom-targeted genomic enrichment of all known deafness genes as well as whole exome sequencing in three large families. In an extended three-generation family with autosomal dominant (AD) non-syndromic post-lingual and progressive deafness, we have identified a novel nonsense mutation (p. W2122X) in *PTPRQ* gene that segregates with the disease phenotype. To date, pathogenic variants in *PTPRQ* have been exclusively associated with autosomal recessive (AR) NSHL that is congenital, severe to profound and progressive in nature. The mutation is located in the phosphatase domain and produces a protein that is lacking the last 6 amino acids. Given that many of the recessive mutations are loss of function, haploinsufficiency as the mechanism of action of the p. W2122X can be excluded. In a second consanguineous Iranian family segregating moderate to severe postlingual progressive ARNSHL we have identified a novel homozygous canonical splice-site mutation in the *CEACAM16* gene. Currently mutations in *CEACAM16* are only described in two families with ADNSHL. Interestingly all three families exhibit a similar phenotype suggesting that loss of function is the most probable mechanism of action of this novel variant. Finally, in a large 4- generation family of European descent with early-onset high frequency progressive ADNSHL, we have identified a novel canonical splice-site mutation in *COL11A1* segregating with the disease. Mutations in *COL11A1* have been previously linked to Stickler and Marshall Syndromes. High throughput next generation sequencing has already advanced our understanding of the molecular genetics of hearing loss, here we use it to expand the phenotypic landscape of deafness genes and shed light on their extraordinary pleiotropy.

2841T

A Comprehensive Search for Deafness Genes Associated with Inner Ear Anomalies. *G. Bademci[1], F. B. Cengiz[1,2], O. Diaz-Horta[1], D. Duman[2], J. Foster II[1], T. Atik[3], T. Kirazli[4], H. Alper[5], I. Menendez[1], A. Pandya[6], I. Akalin[7], G. Sennaroglu[8], S. Guo[1], F. Huesca Hernandez[9], J. Dominguez-Aburto[9], G. D. González Rosado[10], T. M. Felix[11], A. Incesulu[12], M. T. Kalcioglu[13], E. Karatas[14], F. Ozkinay[3], L. Sennaroglu[15], M. Yildirim-Baylan[16], M. Tekin[1].* 1) John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL, USA; 2) Division of Genetics, Department of Pediatrics, Ankara University School of Medicine, Ankara, Turkey; 3) Division of Genetics, Department of Pediatrics, Ege University School of Medicine, Izmir, Turkey; 4) Department of Otolaryngology, Ege University School of Medicine, Izmir, Turkey; 5) Department of Radiology, Ege University School of Medicine, Izmir, Turkey; 6) Division of Genetics and Metabolism, Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC, USA; 7) Department of Medical Genetics, Istanbul Medeniyet University School of Medicine, Istanbul, Turkey; 8) Department of Audiology, Hacettepe University Health Sciences Faculty, Ankara, Turkey; 9) Neurociencias Service, National Institute of Rehabilitation, Mexico D. F, Mexico; 10) Computed Tomography Service, National Institute of Rehabilitation, Mexico D. F, Mexico; 11) Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre, Porto Alegre, RS, Brazil; 12) Department of Otolaryngology and Head and Neck Surgery, Eskisehir Osmangazi University School of Medicine, Eskisehir, Turkey; 13) Department of Otolaryngology, Istanbul Medeniyet University School of Medicine, Istanbul, Turkey; 14) Inonu University, Medical Faculty, Turgut Ozal Medicine Center , Otolaryngology, Head and Neck Surgery, Malatya, Turkey; 15) Department of Otolaryngology, Hacettepe University School of Medicine, Ankara, Turkey; 16) Department of Otorhinolaryngology, Dicle University School of Medicine, Diyarbakir, Turkey.

Inner ear anomalies (IEAs) demonstrated with an imaging study have been reported in approximately 30% of the children with congenital hearing loss. While recent discoveries make molecular diagnosis accessible for a number of deaf individuals, recognized genetic causes of non-syndromic IEAs remain limited. In this work we present the results of whole exome sequencing (WES) in unrelated 68 families with non-syndromic IEAs. Sixty-four probands were simplex and four were multiplex with autosomal recessive inheritance. Agilent SureSelect Human All Exon 50 Mb kits and an Illumina Hiseq2000 instrument were used. Candidate variants were confirmed with Sanger sequencing and evaluated for co-segregation with IEAs in all available family members. We identified causative variants in 9 (13. 2%) families in the recognized IEA genes including *SLC26A4* (MIM 605646) (10. 29%), *FGF3* (MIM 164950) (1. 47%), and *CHD7* (MIM 608892) (1. 47%). While *CHD7* and *FGF3* mutations are known to cause CHARGE (MIM 214800) and LAMM syndromes (MIM 610706), respectively, affected children in this study were either non-syndromic or showed very subtle findings of these syndromes. In an additional family with two affected children we identified a mutation in *ROR1* (MIM 602336) as a novel gene for non-syndromic IEAs. Our data show that WES can be used to identify causative DNA variants in known and novel genes for IEAs and provide important insights into the etiology of deafness.

2842F

***PDZD7* and hearing loss: more than just a modifier.** *KT. Booth[1], H. Azaiez[1], K. Kahrizi[2], AC. Simpson[1], WTA. Tollefson[3], CM. Sloan[1], NC. Meyer[1], M. Babanejad[2], F. Ardalani[2], S. Arzhangi[2], MJ. Schnieders[3], RJH. Smith[1], H. Najmabadi[2].* 1) Molecular Otolaryngology and Renal Research Laboratories, Department of Otolaryngology - Head & Neck Surgery, University of Iowa, Iowa City, Iowa, 52242, USA; 2) Genetics Research Center, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran; 3) Department of Biomedical Engineering, University of Iowa, Iowa City, IA 52242, USA.

Deafness, the most frequent sensory disorder, affects roughly 1 in 500 newborns and by the age of 80, half of all persons require amplification for meaningful communication. Over 90 genes are implicated in non-syndromic hearing loss, a testament to the extreme heterogeneity of this condition. By using targeted genomic enrichment with massively parallel sequencing (using OtoSCOPE®), it is possible to screen all known deafness-causing genes simultaneously. Herein we present four consanguineous families of Iranian origin segregating autosomal recessive non-syndromic hearing loss (ARNSHL). Audiological examination revealed bilateral moderate downsloping to severe hearing loss; ophthalmological examination was normal. Linkage analysis and homozygosity mapping identified segregation with the *DFNB57* locus in two families. OtoSCOPE® results identified five pathogenic variants in the *PDZD7* gene: two nonsense and three missense mutations. Variant segregation with the hearing loss was confirmed by Sanger sequencing in all families. The *PDZD7* gene encodes for the PDZ-domain- containing-7 protein (PDZD7). It is expressed in both the retina and cochlear hair cells, where it plays a vital role in the organization and anchoring of the USH2 protein complex. The two nonsense mutations are predicted to produce null alleles, while the missense mutations are predicted to alter conformational flexibility of specific protein domains. This report is the first to establish that pathogenic variants in *PDZD7* cause non-syndromic hearing loss and not Usher syndrome. These results in association with published data support an essential role for PDZD7 in normal auditory function.

2843W

Exome Sequencing for Deafness Genes: Reducing False Positives by Functional Analysis. *N. Danial-Farran[1], Z. Brownstein[2], A. Abu-Rayyan[3], M. Birkan[2], F. T. A. Martins[2], M. Khayat[1], M. Kanaan[3], S. Shalev[1], K. B. Avraham[2].* 1) Genetics Institute, Ha'Emek Medical Center, Afula, Israel and Rappaport Faculty of Medicine, Technion-Israel Institute of Technology, Haifa, Israel; 2) Department of Human Molecular Genetics and Biochemistry, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 3) Hereditary Research Lab, Department of Life Sciences, Bethlehem University, Bethlehem, Palestine.

Since Next Generation Sequencing has become available as a strategy in genetic research and disease gene discovery, the number of reported causative variants has increased dramatically. Recent studies have shown that more than 25% of reported variants may be false positives, precluding them being the main contributing factors to disease. One of the main pitfalls in characterizing a new variant by computational prediction tools is its absence in sequencing databases or low minor allele frequency (MAF) scores. Our work focuses on identifying pathogenic variants causing hearing loss in the Middle Eastern population. Determining the etiology of hearing loss is crucial for clinical management, determining the prognosis of hearing loss and other abnormalities and assessing the recurrence rate in family members. The use of isolated populations, such as those of the Middle East, has had a strong influence on finding deafness genes worldwide, due to consanguinity and founder effects. We applied targeted gene capture, along with massively parallel sequencing (MPS), of 407 human genes and human orthologues of mouse deafness genes on unrelated families of Israeli Jewish and Palestinian Arab origin. MPS was analyzed using a combination of open-source tools, including Burrows-Wheeler Alignment (BWA) and variant calling with the GATK Genome Analysis Toolkit. We discovered variants that are candidates for deafness in our patients, including in mouse genes not previously associated with human deafness. Some variants predicted to be pathogenic appeared in healthy family members, demonstrating that caution must be taken with prediction tools. Functional analysis was performed on a subset of variants, including immunofluorescence to determine cellular localization. The c. 349C>T, p. (Leu117Phe) variant in SLC26A4 is associated with congenital profound deafness [MIM 600791] in our population. Based on population-scale data, it is predicted to be benign, but pathogenic by Polyphen2, SIFT, Mutation Taster and ConSurf. Our localization experiments of the wild-type and mutant proteins in transfected COS7 cells demonstrated that the mutant is retained in the endoplasmic reticulum, whereas the wild-type targets the plasma membrane, suggesting this variant is indeed pathogenic. The auditory system, through the use of cell and cochlear cultures and mouse models for human deafness, offers suitable avenues for functional analysis to link variants to pathogenicity.

2844T

Genetic variation in genes associated with autosomal recessive non-syndromic hearing impairment. *H. Kremer[1,2], J. Oostrik[1], A. Henkes[2], C. Gilissen[2], I. Feenstra[2], C. Zazo Seco[1], H. G. Yntema[2].* 1) Dept Otorhinolaryngology, Radboud University medical center, Nijmegen, The Netherlands; 2) Dept Human Genetics, Radboud University medical center, Nijmegen, The Netherlands.

Early onset hereditary hearing impairment is assumed to be inherited in an autosomal recessive pattern in the majority of the cases. Currently, more than 60 genes are known to be associated with this type of hearing impairment. Whole exome sequencing (WES) or targeted massive parallel sequencing of all known genes for autosomal recessive non-syndromic hearing impairment (arNSHI) are state-of-the-art strategies for genetic testing, with varying yields. Interpretation of genetic variation is a major challenge in genetic testing and currently, revised genetic diagnoses are to be reported to families that received a genetic diagnosis in the past (Shearer et al. Am J Hum Genet. 2014 95:445-53). To obtain insight in the genetic variation in the population of the Netherlands in 51 genes known to be associated with arNSHI, we have selected rare variants (MAF <1%) identified in WES of 3620 individuals. To assess the predicted outcome *CDH23* (MIM *605516) of nonsynonymous Single Nucleotide Variants (SNVs), the Condel score was determined which combines various prediction programs (e. g. SIFT, Polyphen2, MutationAssessor). A potential effect on splicing was predicted for SNVs from nucleotide +1 to +8 of splice donor sites and -8 to -1 of splice acceptor sites with Genesplicer, NetGene and NNSplice. The total number of predicted pathogenic variants in the 51 arNSHI genes was 1223, the majority of which were private variants. Seven variants in six genes had an allele frequency of 0. 5% or higher including mutations in *GJB2* (MIM *121011) and *MYO15A* (MIM *602666) and part of these were repeatedly detected in WES of patients with arNSHI. These variants are likely to be founder mutations or represent mutational hotspots. Predicted carrier frequencies of the 51 known genes for arNSHI varied between 4. 97% (*CDH23*) and 0. 04% (*SYNE4*; MIM *615535). The carrier frequency in *CDH23* already predicts an incidence of arNSHI of 2 in 1000 which is are higher than the reported incidence of arNSHI which is about 1 in 1875 based on an incidence of 1 in 750 for early onset hearing impairment with 40% of these cases based on autosomal recessive mutations. These findings demonstrate the importance of segregation analysis in genetic testing and the pitfalls in diagnostics of isolated cases. On the other hand, the high carrier frequency of (predicted) pathogenic variants raises the question whether these variants contribute to complex inheritance of hearing impairment with onset later in life.

2845F

Application of a targeted next-generation sequencing panel (OTO-NGS-Panel) in a cohort of Brazilian patients with hereditary hearing loss. *PZ. Ramos[1], L. Borreguero[2], M. Morin[2], L. Serrao-de-Castro[2], I. del Castillo[2], SM. da Silva-Costa[1], GM. de Carvalho[3], AM. Castilho[3], EL. Sartorato[1], MA. Moreno-Pelayo[2].* 1) CBMEG, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil; 2) Genetics Department, Hospital Ramón y Cajal-IRYCIS, CIBERER-U728, Madrid, Spain; 3) ENT Department, Faculty of Medical Sciences, University of Campinas (UNICAMP), Campinas, São Paulo, Brazil.

Hearing loss is one of the most common sensory disorders in humans. Over 400 syndromes that include hearing loss have been described with at least 40 genes associated with the main types of syndromic hearing loss. Additionally, more than 80 genes and 140 loci have been identified for nonsyndromic hearing loss. Studies have shown that a single gene may be related to both syndromic and nonsyndromic hearing loss, or even associated with different patterns of inheritance, dominant or recessive. The identification of genetic variants can contribute to a better understanding of the molecular basis and the pathophysiological mechanisms involved in the different phenotypes of hereditary hearing loss, and can provide an accurate diagnosis, development of specific treatments and genetic counseling for patients and families. However, molecular diagnosis has been a challenge, mainly due to clinical and genetic heterogeneity of hearing loss. The application of new high-throughput technologies, such as next-generation sequencing, allows an optimization of genetic diagnosis and a comprehensive investigation of hereditary hearing loss. Therefore, the main goal of the present study was to elucidate the genetic etiology in a cohort of Brazilian patients with bilateral sensorineural hearing loss due to distinct inherited patterns using the OTO-NGS-panel, a targeted next-generation sequencing panel designed by the Genetic Department at the Hospital Ramón y Cajal, Madrid-Spain, and the Illumina MiSeq platform to simultaneously analyze 72 genes already known to be involved in hereditary hearing loss. In our pilot study we have screened 13 unrelated sporadic or familial cases and pathogenic mutations were identified in six out of 13 patients (approximately 50%). Nine mutations in five different genes were detected. The following genes were involved: *ACTG1* (MIM 604717) and *DFNA5* (MIM 608798), associated with autosomal dominant hearing loss and *CDH23* (MIM 601386), *MYO15A* (MIM 600316) and *OTOF* (MIM 601071), related to autosomal recessive hearing loss. We reported the identification of three novel mutations and the involvement of genes that had never been investigated before in Brazil. Our study confirms the usefulness of the OTO-NGS-panel for detecting the genetic etiology of hearing loss in a heterogeneous group of patients.

2846W

Whole Genome Sequencing Delineates Genetic Sub-Types of Neuromyelitis Optica. *K. Estrada[1,2,3], F. Zhao[2,3], C. Whelan[3], J. Flannick[3], C. Sun[1], T. Harris[1], B. Greenberg[4,5], J. Carulli[1], A. G. Day-Williams[1], D. MacArthur[2,3].* 1) Computational Biology and Genomics, Biogen, Cambridge, MA; 2) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston MA; 3) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA; 4) Department of Neurology and Neurotherapeutics; 5) Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX.

Neuromyelitis optica (NMO) is a rare autoimmune disorder of the central nervous system that affects the optic nerves and spinal cord with a prevalence of 1/100,000 in the USA. Auto-antibodies against Aquaporin 4 (AQP4) are found in about 70% of NMO patients (NMO-IgG+) yet no association to AQP4 has been identified in small genome-wide association studies (GWAS), but nominal associations to the HLA have been reported. We have performed the first whole-genome sequencing (WGS) analysis of NMO by deep sequencing (mean_DP>24X) 143 NMO patients and 658 population controls. We have performed joint variant calling yielding 42 million single-nucleotide variants (SNVs) and 6. 3 million insertions/deletions. After variant and sample quality control 29,092,408 variants were identified in 86 cases and 460 controls. In a single-variant association analysis of 9,357,936 variants with a MAF>0. 5%, we found at least two independent signals identified by the variants rs146828696 (OR=4. 8, P=4. 9x10-13) near the HLA-DRB5 gene and rs147131125 (OR=4. 5, P=6x10-9) near the HLA-DRB1 gene. In a stratified analysis by AQP4 auto-antibody status, both signals were significantly associated with NMO-IgG+, but only rs146828696 was associated in NMO-IgG- (OR=5. 6, P=3. 1x10-6). To further investigate the role of the HLA we imputed classical HLA alleles from the sequencing data. The HLA-DRB1:03*01 was the most significant imputed allele (OR=4. 5 P=6. 7x10-9) in the NMO-IgG+ and was not associated with NMO-IgG-. Both rs146828696 and rs147131125 remained associated after conditioning for HLA-DRB1:03*01. An imputed missense variant in HLA-DRB1 Arg74 carried by the HLA-DRB1:03*01 haplotype had the most significant association with NMO-IgG+ (OR=4. 2, P=6. 8x10-9). Gene-based analyses using only missense and/or loss-of-function variants with a MAF <1% did not show any gene with P<2. 5x10-6. Finally, we performed a CNV discovery and association analysis where we also identified a 3. 7kb deletion carried by HLA-DRB1:03*01 near the *C4* gene associated with NMO-IgG+ (OR=3. 7, P=7. 8x10-8). These genome-wide significant SNPs as well as 16 additional SNPs with P<1x10-5 are being replicated in 150 cases and 920 controls. Identification of the first genome-wide significant association in NMO is an important discovery in the field, and replicating differential genetic associations between NMO-IgG+ and NMO-IgG- would be a major advancement in the molecular understanding of the heterogeneity of the disease.

2847T

HA20: A novel autoinflammatory disease caused by haploinsufficiency of A20, encoded by *TNFAIP3* . *Q. Zhou[1], H. Y. Wang[1], M. Stoffels[1], Y. H. Park[1], M. Takeuchi[1], D. Yang[2], W. L. Tsai[3], Y. Zhang[4], E. Demirkaya[5], J. Lyons[4], X. M. Yu[4], H. Leavis[6], S. Ozen[7], M. Boehm[2], J. Milner[4], M. Gadina[3], J. J. Chae[1], D. L. Kastner[1], I. Aksentijevich[1].* 1) Inflammatory Disease Section, NHGRI, Bethesda, USA; 2) Laboratory of Cardiovascular Regenerative Medicine, NHLBI, Bethesda, USA; 3) Translational Immunology Section, NIAMS, Bethesda, USA; 4) Genetics and Pathogenesis of Allergy Section, NIAID, Bethesda, USA; 5) Institute of Health Sciences, R&D Center, Ankara, Turkey; 6) University Medical Center Utrecht, Utrecht, Netherlands; 7) Hacettepe University, Ankara, Turkey.

TNFAIP3/A20 protein functions as a potent negative regulator of the NFκB signaling pathway. Low-penetrance common variants of *TNFAIP3* have been associated with a number of autoimmune and inflammatory diseases. Biallelic somatic inactivating mutations in *TNFAIP3* have been described in B cell lymphomas. Here, we report 5 high penetrance dominantly inherited frameshift and nonsense *TNFAIP3* mutations in 11 patients from 5 unrelated families with early-onset systemic inflammation, arthralgia/arthritis, oral and genital ulcers, and ocular inflammation. The syndrome resembles Behcet's disease BDthat is typically considered a polygenic disease with onset in early adulthood. Targeted sequencing of *TNFAIP3* in Turkish and Japanese GWAS cohorts with BD identified one patient with a novel frameshift mutation. None of the mutations were reported in any public database. Expression of wild type (WT) A20 was reduced in patients' PBMCs and fibroblasts, while mutant proteins were not detectable. Over-expressed mutant proteins failed to suppress TNF-α-induced NFκB activity in transfected 293T cells in comparison to WT A20. Co-transfection with the WT A20 plasmid normalized the NFκB activity suggesting a haploinsufficiency mechanism of disease. Stimulated patients' cells showed increased phosphorylation of IKKα/β, increased degradation of IkBα, and increased nuclear translocation of p65. These data show strong evidence that mutant proteins cause enhanced signaling in the NF-κB pathway. Patients' cells displayed weak associations with the TNFR complex and a marked defect in the deubiquitination of target molecules RIP1, NEMO and TRAF6. This defect was partially rescued by co-transfection with a WT-A20 construct. Patients' cells had a sustained higher K63-ubiquitinaton level of NEMO and RIP1 and accumulated high-molecular weight Ub-aggregates. These results indicate that the reduced interaction of A20 with the TNFR signaling complex leads to inefficient deubiquitination of target proteins. Multiple experiments showed evidence for increased expression of NFκB target genes in patients' immune cells. Levels of IL-1β, TNF-α, IP-10, IL-17, and IL-9 were substantially increased in patients' serum and in the supernatants of stimulated PBMCs relative to healthy controls. Initial experience with agents targeting pro-inflammatory cytokines has been encouraging. This study provides the first example of the effects of high-penetrance loss-of-function mutations in *TNFAIP3* in humans.

2848F

Identification of uniparental isodisomy from sequencing data. *D. Bis[1], R. Schüle[2], P. Bauer[2], A. Söhn[2], T. Rattay[2], C. Saghira[1], L. Abreu[1], F. Tao[1], S. Züchner[1].* 1) Dr. John T. Macdonald Foundation Department of Human Genetics, University of Miami, Miami, FL; 2) Department of Neurodegenerative Disease, Hertie Institute for Clinical Brain Research and Center for Neurology, Tübingen, Germany.

Uniparental disomy (UPD) is the inheritance of two copies of a chromosomal region from one parent. The copied region can be an identical copy of one parental allele (isodisomy) or distinct alleles from the same parent (heterodisomy). Isodisomy can generate homozygosity for a deleterious recessive mutation, but without knowledge of the isodisomic event, the disease causing mutation is excluded from analysis since one parent would not carry the recessive allele. The extent of UPD as a cause of disease is still not fully understood. Only whole genome sequencing, and to some extent whole exome sequencing, allows for the routine detection of such events of nearly every size. We are systematically studying exomes and genomes of a sizable cohort (>100 families) of rare neurodegenerative diseases. Thus far, we have identified a case with Hereditary Spastic Paraplegia (SPG35) where a homozygous UPD spanned 80% of chromosome 16. Isodisomy was confirmed by excluding genomic deletions and genotyping polymorphic markers within the parents, and the homozygous disease-causing mutation was identified within this region. We will present the full evaluation of UPD in our cohort of rare Mendelian families, already excluded for candidate genes by whole exome sequencing. It is possible that the contribution of UPD has been significantly underestimated because of technical challenges to explore it in larger sample size.

2849W

Rare Variant Gene-Based Case-Control Burden Testing in a Mendelian Disease Framework Using Publically Available Control Data. *M. Lippincott[1], M. Guo[2], L. Plummer[1], J. Hirschhorn[2], W. Crowley[1].* 1) Reproductive Endocrine Unit, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114; 2) Boston Children's Hospital, 3 Blackfan Circle, Boston, MA 02115.

Background: Isolated GnRH Deficiency (IGD) is characterized as failure to enter or progress through puberty. Over 30 mendelian loci have been identified which together account for only 1/3 of the disease cases. However, the majority of IGD cases are sporadic given the reproductive consequences of this disorder. We hypothesized that IGD cases would be enriched for rare deleterious variants in genes causing IGD compared to controls. Aim: To develop a gene-based burden testing model of WES data using publically available control data and to use this method to discover new genes in a rare disease model. Methods: Whole-exome sequencing was done on 148 non-Finnish European cases, which were enriched for IGD-causing variants in *PROKR2* (a positive control). The publically available Exome Aggregation Consortium (ExAC) WES data on non-Finnish Europeans was used for controls. Each data set was restricted to rare (MAF less than 0. 0001), putatively deleterious (nonsense, frameshift, essential splice, missense with CADD score >10) variants in areas of shared coverage with an average read depth of 10. These multiple variants were then collapsed by gene and the gene-based burden of variants was compared between cases and controls using Fisher's Exact Test. Results: Burden testing revealed genes known to cause IGD: *PROKR2* (9 variants in cases, p<1. 3E-9), *FGFR1* (9 variants in cases, p<1. 2E-8), and *KAL1* (5 variants in cases, p<1. 8E-5). All variants were confirmed by Sanger sequencing. In addition, 2 other genes had exome-wide significance p values (p<2. 5E-6) and are novel candidates for genes causing IGD. Conclusions: We report the ability to use publically available WES data to design rare variant case-control gene-based burden testing in a rare disease model. This rare variant burden testing was validated by its ability to recapitulate genes known to cause IGD. As a result, novel genes passing the burden testing are being evaluated for their possible role in IGD.

2850T

Identifying candidate mutations for rare birth defects through family trio-based exome sequencing. *T. Nicholas[1], D. Koboldt[1], D. Wegner[2], R. Fulton[1], B. Ozenberger[1], J. Wambach[2], D. Grange[2], R. Wilson[1,3], F. Cole[2].* 1) McDonnell Genome Institute , Washington University, St. Louis, MO; 2) Department of Pediatrics, Washington University, St. Louis, MO; 3) Department of Genetics, Washington University, St. Louis, MO.

Genomic information is increasingly relevant for clinical care, particularly with the availability of rapid, inexpensive tools for genome and exome sequencing. The McDonnell Genome Institute and the Department of Pediatrics at the Washington University School of Medicine have initiated a program to identify candidate pathogenic variants associated with rare birth defects. It is estimated that as many as 80% of such cases can be attributed to underlying genetic variation. Families recruited into the program are selected by a board of medical geneticists, clinicians, and genome scientists, based on phenotypic information, relevant clinical diagnostic studies, family history, and availability of appropriate informed consented samples. Our strategy for identifying candidate causal variants involves exome sequencing of the proband and immediate family members (parents and available siblings). The sequence data are aligned to the human reference sequence and assessed for variants, including both single nucleotide variants and indels, which are further annotated with gene and functional information. Candidate causal variants (*de novo* mutations and rare recessive or compound heterozygous variants) are reported and prioritized based on relevant segregation patterns, predicted functional consequences and population allele frequencies. This program provides a streamlined approach for combining clinical expertise with genomic methods to better identify causative genetic variants that can inform and aid the diagnosis and treatment of patients with rare birth defects, and reveal the genetic risk of family members. Here, we present the initial findings from exome sequencing of 8 families (28 samples total) with rare birth defects including congenital heart disease, congenital diaphragmatic hernia, genitourinary defects, sternal malformation, and neuromuscular disease.

2851F

The UK 100,000 Genomes Project. *K. Smith[1], J. Davies[1], A. Devereau[1], T. Fowler[1], T. Hubbard[1,2], E. McDonagh[1], M. Parker[1], A. Rendon[1], L. Riley[1], A. Rueda Martin[1], M. Ryten[1], E. Thomas[1], C. Turnbull[1], M. Caulfield[1], Genomics England.* 1) Genomics England, London, United Kingdom; 2) King's College London, London, United Kingdom.

The UK's 100,000 genomes project has begun consenting participants to its main programme through a network of 11 NHS Genomic Medicine Centres (GMCs) spanning over 70 local delivery partner institutions. It has four main aims: (1) to bring benefit to patients; (2) to create an ethical and transparent programme based on consent; (3) to enable new scientific discovery and medical insights; and (4) to kick-start the development of a UK genomics industry. The project focuses on patients with a rare disease and their families (approximately 50,000 genomes), as well as patients with certain common cancers (about 25,000 tumour-normal pairs). Whole genome sequencing of DNA extracted from blood samples is performed to at least 30x depth for germline samples and 75x for tumour samples, using Illumina's HiSeq X Ten sequencing platform. PCR-free library preparation is employed when feasible. All germline samples are required to achieve at least 15x high-quality coverage over 95% of the autosomal genome. One of the innovations of the project is the collection of phenotype data from the GMCs in a comprehensive and standardised manner, either through direct data entry or by the population of data models directly from Electronic Health Record systems. Data models for each of the 120 currently eligible rare genetic conditions, as well as for cancers, have been developed in consultation with clinical experts to support clinical data capture and phenotyping. This approach was designed to enable clinical interpretation and large-scale genomic research. Clinical data models include questions about the presence or absence of human phenotype ontology (HPO) terms, additional clinical tests, and family history. To date, data models include 1370 different HPO terms, with a median 37 terms per condition and range 2 (hyperinsulinism) to 116 (mitochondrial disorders). Over 200 phenotypes have been proposed for addition to the HPO. One of the important benefits of the programme will be to obtain HPO term frequencies based on large numbers of patients, rather than occurrence in OMIM entries, informing more powerful variant prioritisation strategies.

2852W

Inherited disorders in apparently acquired severe aplastic anemia. *B. J. Ballew[1], S. Gadalla[1], S. Spellman[2], M. Haagenson[2], S. Lee[2], S. A. Savage[1], NCI DCEG Cancer Genomics Research Laboratory, NCI DCEG Cancer Sequencing Working Group.* 1) Clinical Genetics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Rockville, MD, USA; 2) Center for International Blood and Marrow Transplant Research, Minneapolis, MN, USA.

Severe aplastic anemia (SAA) is associated with increased risk of myelodysplastic syndrome, leukemias, and possibly solid tumors. A subset of patients with isolated SAA may have an underlying inherited bone marrow failure syndrome (IBMFS). The IBMFS require distinct medical management compared with acquired SAA. For example, the IBMFS do not respond to immunosuppressive therapy, the mainstay of treatment in acquired SAA. Moreover, hematopoietic cell transplant (HCT) regimens must be modified to reduce morbidity in IBMFS patients. In collaboration with the Center for International Blood and Marrow Transplant Research, we are exploring the genetic etiology of SAA by performing whole exome sequencing (WES) on ~600 individuals who underwent matched-unrelated HCT for SAA. We analyzed WES data on a test set of 293 unrelated SAA individuals, 91 of whom had been diagnosed with an IBMFS based on clinical criteria. Variants were called using GATK Unified Genotyper, GATK Haplotype Caller, and FreeBayes. We removed variants of low quality or with a MAF>1% in the NHLBI Exome Sequencing Project and 1000 Genomes data. We then evaluated variants present in 56 genes known to be associated with 20 IBMFS, and separated the variants into 3 categories: likely pathogenic, possibly pathogenic, and likely non-pathogenic. These distinctions were based on a variant's proximity to known deleterious mutations, impact on the protein, and *in silico* predictions. This resulted in an average of 4 potentially deleterious variants per individual. Based on known IBMFS inheritance patterns, we were able to identify potentially causal mutations in ~75% of the 91 individuals with a known IBMFS diagnosis; this is consistent with expected number based on the literature. Notably, ~20% of the 202 SAA individuals without an IBMFS diagnosis had a potentially causal mutation. 602 healthy controls, sequenced on the same platform and called concurrently, did not have potentially causal mutations in these 56 genes. We have sequenced a second validation set of SAA individuals and analysis is underway. Additionally, because individuals with IBMFS are more at risk of complications following HCT conditioning, we are assessing the association between a putative pre-existing IBMFS and outcomes following transplant. These findings may have significant clinical implications regarding genetic testing of individuals presenting with SAA to shape treatment decisions.

2853T

Genetic determinism of Primary Osteoarthritis: the EXORHUM project. *C. Bauge[1,2], J. AURY-LANDAS[1,2], S. LECLERCQ[3], C. MARCELLI[4], K. BOUMEDIENE[1,2].* 1) Normandie Université, France; 2) UNICAEN, EA4652 MILPAT, Caen, France; 3) Service de Chirurgie orthopédique, Hôpital privé Saint Martin, Caen, France; 4) Service de Rhumatologie, CHU, Caen, France.

Osteoarthritis (OA) is the most common joint disease observed worldwide and affects nearly 10 million adults in France and 27 million adults in USA. OA is mainly characterized by the gradual loss of the articular cartilage, particularly affecting the knee, hip, spine, hand and foot. Although the multifactorial nature of the vast majority of OA cases is well recognized, genetic risk factors have been found to be strong determinants of the disease. About 50% of hip and hand OA may be genetically determined. However, currently, only few genetic alterations responsible for OA determinism have been identified mostly in genes involved in the cartilage extracellular matrix. These alterations have been mainly detected in patients with syndromic early-onset OA like pseudoachondroplasia, multiple epiphyseal dysplasia or spondyloepiphyseal dysplasia tarda. The EXORHUM project is based on the recruitment of a well clinically characterized cohort of patients with non syndromic early-onset OA. The main aim is to identify new genetic alterations, exclusive to primary OA in patients with a familial presentation, strongly suggestive of a genetic determinism. Exome sequencing will be performed on two distant affected family members without known alterations. The sub aim is to assess the prevalence of known mutations in primary non syndromic OA. In conclusion, the EXORHUM project will help us to improve the understanding of molecular processes involved in pathogenesis of OA and to better understand the complexity of OA. This is essential to improve patient diagnosis and to identify new therapeutic targets.

2854F

Molecular etiology of arthrogryposis in a cohort of families of Turkish origin. *Y. Bayram[1], E. Karaca[1], Z. Coban Akdemir[1], H. Aydin[2], A. Gezdirici[3], D. Torun[4], S. Tug Bozdogan[5], S. Isikay[6], M. M. Atik[1], T. Gambin[1], A. Karaman[7], D. Pehlivan[1], H. Aslan[8], O. Ozalp Yuregir[9], S. N. Jhangiani[10], E. Boerwinkle[10,11], R. A. Gibbs[10], N. Elcioglu[12], B. Tuysuz[13], J. R. Lupski[1,10,14,15], Baylor-Hopkins Center for Mendelian Genomics.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 2) Department of Medical Biology, Namik Kemal University Medical Faculty, Tekirdag, TURKEY; 3) Department of Medical Genetics, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul, TURKEY; 4) Department of Medical Genetics, Gulhane Military Medical Academy, Ankara, TURKEY; 5) Department of Medical Genetics, Mersin University Faculty of Medicine, Mersin, TURKEY; 6) Department of Pediatric Neurology, Gaziantep Children's Hospital, Gaziantep, TURKEY; 7) Center of Genetics Diagnosis, Zeynep Kamil Women`s and Children's Diseases Training and Research Hospital, Istanbul, TURKEY; 8) Department of Medical Genetics, Eskisehir Osmangazi University Faculty of Medicine, Eskisehir, TURKEY; 9) Genetics Diagnosis Center, Seyhan Practice Center, Adana Numune Training and Research Hospital, Adana, TURKEY; 10) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 11) Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA; 12) Department of Pediatric Genetics, Marmara University School of Medicine, Istanbul, TURKEY; 13) Department of Pediatric Genetics, Istanbul University Cerrahpasa Medical Faculty, Istanbul, TURKEY; 14) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 15) Texas Children's Hospital, Houston, TX, USA.

The term "arthrogryposis" is clinically defined as congenital joint contractures in two or more body areas. Arthrogryposis is a symptom rather than a specific diagnosis. To date, more than 400 different disorders have been described as having arthrogryposis and around 150 of these have been associated with a responsible gene alteration. Here we report whole exome sequencing (WES) data of 50 patients from 46 families (22 of which are consanguineous) with clinical findings that include arthrogryposis. While variants in known genes were found in affected individuals from 24 families (52%), WES analysis revealed deleterious variants in candidate genes that were not previously associated with arthrogryposis in 5 families (11%). The subjects from 17 families remained unsolved; the initial overall solved rate was calculated as 63%. The most common genes that we detected as harboring homozygous variants were *CHRNG* (in 6 patients) and *ECEL1* (in 4 patients). The identified variants in novel genes express different mode of inheritance such as recessive-homozygous (*VPS8*), recessive-compound heterozygous (*MYPBC2*), X-linked hemizygous (*ITGB1BP2*), *de novo* heterozygous (*FBN3*), and autosomal dominant (*PSD3*) in different families. Using WES as a genomic sequencing method enables us to investigate potential oligogenic inheritance or blended phenotype models (i. e. deleterious homozygous or compound heterozygous variants in two different loci or known genes in one proband) in some patients. For example, in addition to homozygous variants in *CHRNG* in a more severely affected patient with multiple pterygium syndrome, we identified another homozygous deleterious variant in *ERCC2*, which is also a known gene for an arthrogryposis related disorder. We suggest that increasing use of genome-wide technologies will provide opportunities to better understand molecular mechanisms of genetically heterogeneous disorders such as arthrogryposis.

2855W

Homozygous mutation in the eukaryotic translation initiation factor 2alpha phosphatase gene, *PPP1R15B*, is associated with severe microcephaly, short stature, and intellectual disability. *K. D. Kernohan[1], M. Tetreault[2, 3], U. Liwak-Muir[1], M. T. Geraghty[1, 4], W. Qin[1], S. Venkateswaran[5], J. Davila[6], M. Holcik[1], J. Majewski[2, 3], J. Richer[7], K. M. Boycott[1, 7], Care4Rare Canada Consortium.* 1) CHEO Research Institute, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 2) Department of Human Genetics, McGill University, Montreal, Quebec, Canada; 3) McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, Canada; 4) Division of Metabolics and Newborn Screening, Department of Pediatrics, University of Ottawa, Ottawa, Ontario, Canada; 5) Division of Neurology, Department of Pediatrics, University of Ottawa, Ottawa, Ontario, Canada; 6) Department of Radiology, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; 7) Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada.

Protein translation is an essential cellular process initiated by the association of a methionyl–tRNA with the translation initiation factor eIF2. The Met-tRNA/eIF2 complex then associates with the small ribosomal subunit, other translation factors, and mRNA, which together comprise the translational initiation complex. This process is regulated by the phosphorylation status of the α subunit of eIF2 (eIF2α); phosphorylated eIF2α attenuates protein translation. Here we report a consanguineous family with severe microcephaly, short stature, hypoplastic brainstem and cord, delayed myelination, and intellectual disability in two siblings. Whole-exome sequencing identified a homozygous missense mutation, c. 1972G>A; p. Arg658Cys, in protein phosphatase 1, regulatory subunit 15b (PPP1R15B), a protein which functions with the PPP1C phosphatase to maintain dephosphorylated eIF2α in unstressed cells. The p. R658C PPP1R15B mutation is located within the PPP1C binding site. We show that patient cells have greatly diminished levels of PPP1R15B-PPP1C interaction, which results in increased eIF2α phosphorylation and resistance to cellular stress. Finally, we find that patient cells have elevated levels of PPP1R15B mRNA and protein, suggesting activation of a compensatory program aimed at restoring cellular homeostasis which is ineffective due to PPP1R15B alteration. PPP1R15B now joins the expanding list of translation associated proteins which when mutated cause rare genetic diseases.

2856T

Mutations in the glutaminyl-tRNA synthetase gene cause early-onset epileptic encephalopathy. *H. Kodera[1], H. Osaka[2,5], M. Iai[2], N. Aida[3], A. Yamashita[4], Y. Tsurusaki[1], M. Nakashima[1], N. Miyake[1], H. Saitsu[1], N. Matsumoto[1].* 1) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Kanagawa, Japan; 2) Division of Neurology, Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 3) Division of Radiology, Clinical Research Institute, Kanagawa Children's Medical Center, Yokohama, Japan; 4) Department of Molecular Biology, Yokohama City University School of Medicine and Graduate School of Medical Science, Yokohama, Japan; 5) Department of Pediatrics, Jichi Medical School, Tochigi, Japan.

Aminoacylation is the process of attaching amino acids to their cognate tRNA, and thus is essential for the translation of mRNA into protein. This direct interaction of tRNA with amino acids is catalyzed by aminoacyl-tRNA synthetases. Using whole exome sequencing, we identified compound heterozygous mutations [c. 169T>C (p. Tyr57His) and c. 1485dup (p. Lys496*)] in *QARS*, which encodes glutaminyl-tRNA synthetase, in two siblings with early-onset epileptic encephalopathy (EOEE). Recessive mutations in *QARS*, including the loss-of-function missense mutation p. Tyr57His, have been reported to cause intractable seizures with progressive microcephaly. The p. Lys496* mutation is novel and causes truncation of the QARS protein, leading to a deletion of part of the catalytic domain and the entire anticodon-binding domain. Transient expression of the p. Lys496* mutant in neuroblastoma 2A cells revealed diminished and aberrantly aggregated expression, indicating the loss-of-function nature of this mutant. Together with the previous report, our data suggest that abnormal aminoacylation is one of the underlying pathologies of EOEE.

2857F

Exome sequence analysis identifies novel candidate genes associated with neuronal migration disorders. *W. Wiszniewski[1], P. Gawlinski[2], T. Gambin[1], S. Jhangiani[3], D. Muzny[3], J. Bal[2], E. Boerwinkle[3, 4], R. Gibbs[3], J. Lupski[1, 3, 5, 6].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Institute of Mother and Child, Warsaw, Poland; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 4) Human Genetics Center and Institute of Molecular Medicine, University of Texas-Houston Health Science Center, Houston, TX; 5) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 6) Texas Children's Hospital, Houston, TX.

Neuronal migration disorders (NMDs) are a group of developmental defects caused by the abnormal migration of neurons in the developing brain with over 100 causative genes identified to date. Whole exome sequencing (WES) of 29 individuals with NMDs and their parents identified apparent causative mutations in ~62% (18/29) of families. Mutations were found in known NMDs genes and novel candidate genes in 31% (9/29) and 20% (6/29) subjects respectively. Three additional patients (3/29) were found with mutations in known genes, *CACNA1A, SYNE1* and *FRAS1*, that have been already associated with human pathology but not with isolated brain abnormalities, thus they likely represent an example of phenotypic expansion. Three patients (3/29) represented a familial form of NMDs with mutations inherited from an affected parent in AD (*RELN*) and X-linked (*FLNA* and *L1CAM*) fashions. Patients with no family history of NMDs were found with compound heterozygous mutations suggestive of AR inheritance in 28% (8/29) of families and *de novo* variants in 24% (7/29). These findings provide us with information on distribution of mutations in known NMDs genes as well as potential new candidate genes involved in a process of neuronal migration. Identification of novel genes may contribute to our understanding of brain development in health and disease and improve a clinical classification of neuronal migration disorders.

2858W

Identification of a rare variant in *POLD1* in a family with familial multiple lipomatosis. *C. Mirlene[1], S. Raskin[2], J. You[3,4], J. Jurgens[3,4], M. Migliavacca[1], D. Valle[4,5,6], N. Sobreira[4].* 1) Clinical Genetics, Department of Morphology and Genetics, Universidade Federal de São Paolo, Rua Coronel Lisboa, 966, 04020-041, São Paulo, SP, Brazil; 2) Genetika Laboratory, Curitiba, Brazil; 3) Predoctoral Training Program in Human Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 4) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 5) Center for Inherited Disease Research, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21224, USA; 6) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA.

Lipomatosis, also referred as familial multiple lipomatosis (FML, MIM151900) is a rare autosomal dominant phenotype characterized by multiple overgrowths of mature adipose tissue and multiple lipomas on the trunk and limbs. Cytogenetic studies have shown that rearrangements involving chromosomes 12 and 3 are often associated with FML and *HMGA2* has been suggested as a causative gene but no molecular explanation has been defined. Here we report a family submitted to the Baylor-Hopkins Center for Mendelian Genomics with FML. Four affected with (male proband, his father, paternal uncle and paternal aunt) and 1 unaffected in which the phenotype co-segregates with a rare variant in *POLD1* (V70F, c. 208G>T) detected by WES. The proband also had short stature and Arnold-Chiari malformation. We analyzed the WES data using the PhenoDB variant analysis tool (http://phenodb. org) to identify rare functional variants that fit the autosomal dominant inheritance mode. Of the 98 variants identified by this analysis, none mapped to chr12q14 or chr3q27-q28. A variant in exon 3 of *POLD1* (V70F, c. G208T) was present in all 4 affected individuals and not present in the proband's unaffected brother. This variant is predicted to be a damaging with a Sift score of 0. 06, and a Polyphen score of 0. *POLD1* encodes the 125-kDa catalytic subunit of DNA polymerase delta. DNA polymerase delta possesses both polymerase and 3' to 5' exonuclease activity and plays a critical role in DNA replication and repair. Previously, homozygous mice with inactivating point mutation in the proofreading domain of polλ(D400A) are prone to cancer including lymphomas at a young age and have a high incidence of epithelial tumors later in life. Based on this information, we selected the *POLD1* variant as our strongest candidate. In 2013, Weedon et al. described 4 unrelated individuals (2 previously reported by Shastry et al. , 2010) with mandibular hypoplasia, deafness, progeroid feature and lipodystrophy (MDPL) syndrome all with a *de novo* heterozygous in-frame deletion (S605del, c. 1812_1814delCTC) of a single codon in *POLD1* affecting the polymerase active site. We suggest that the variant in our family may confer a gain of function effect causing the overgrowth of the adipose tissue and FML phenotype in contrast to the lipodystrophy phenotype described in the patient with the heterozygous in-frame deletion.

2859T

Novel gene discovery across a large cohort of patients with syndromic craniofacial anomalies. *E. J. Bhoj, D. Li, M. H. Harr, L. Tian, T. Wang, Y. Zhao, H. Qiu, F. Mentch, C. Kim, E. H. Zackai, H. Hakonarson.* Children's Hospital of Philadelphia Philadelphia PA.

For many complex disorders it can be difficult to identify patients with single-gene disorders suitable for novel gene discovery. Many such conditions can also result from multigenic and environmental insults, which dilute discovery cohorts for research sequencing efforts for novel Mendelian genes. We have addressed this issue by creating an international highly-informative cohort of patients with craniofacial anomalies and at least one other medical condition to enrich for families affected by single-gene disorders. Additional findings in this cohort include congenital heart disease, brain malformations, seizures, cardiomyopathy, autism, hypotonia, and anomalies of the limbs, kidneys, liver, genitourinary system ordiaphragm. Over 200 families with an undiagnosed craniofacial syndrome have been carefully screened for known mutations and enrolled in our study, over 50 have had exome sequencing performed. Initial discoveries from this cohort included SPECC1L mutations in Teebi Hypertelorism Syndrome, MYRF andHox3B mutations in Hardikar syndrome, and the identification of over a dozen genes not previously implicated in human disease, including MACF1, SUV420H1, H3F3A, ZNF599, FAT3, CDH18, RRAS2, MYRF,HOX3B, ZNRF3, Radil, and DCHS2. As these genes appear to play a role beyond the craniofacial pathology in these families, we were able to discover novel candidate genes underlying over twenty separate medical conditions.

2860F

Accelerating the pace of Mendelian gene discovery via multi-modal data sharing: Geno$_2$MP, the UW-RMD Family Portal, and MyGene^2. *J. X. Chong[1], J-H. Yu[1], Q. Yi[2], S. Jamal[1], H. K. Tabor[1,3], J. Shendure[2], D. A. Nickerson[2], M. J. Bamshad[1,2,4], University of Washington Center for Mendelian Genomics.* 1) Department of Pediatrics, University of Washington, Seattle, WA; 2) Department of Genome Sciences, University of Washington, Seattle, WA; 3) Treuman Katz Center for Pediatric Bioethics, Seattle Children's Research Institute, Seattle, WA; 4) Division of Genetic Medicine, Seattle Children's Hospital, Seattle, WA.

Since the first application of exome and whole-genome sequencing to Mendelian gene discovery in 2010, the rate of reported gene discoveries has increased ~40%, from an average of ~166/year from 2005-2009 to ~231/year from 2010-2014. However this increase is perhaps more modest than anticipated for many reasons including limited adoption of open sharing of sequence data and candidate genes/variants. Furthermore, existing resources explicitly target professional users such as researchers, clinicians, and diagnostic laboratories. We introduce three new, related resources that will help tackle these issues: Geno2MP, the University of Washington Repository of Mendelian Disorders Family Portal (FP), and MyGene2. Geno2MP ("Genotype to Mendelian Phenotype") (http:// geno2mp. gs. washington. edu) is a browser that, in its initial form, releases de-identified, aggregated variant data linked to structured phenotypic data for ~3,000 individuals sequenced by the UW-CMG. Geno2MP is intended to facilitate candidate variant prioritization and identification of individuals with variants in a shared candidate gene not necessarily selected for follow-up analysis. It has already been used to identify two unrelated patients sharing a phenotype caused by nonsense mutations in the same novel gene. Family-driven social networking to identify "second families" sharing the same Mendelian gene/phenotype was pioneered by parents of children with *NGLY1* deficiency and led to the delineation of a novel disorder. However, replicating their success requires families to have some technical knowledge and resources. Following our involvement in a similar effort, we developed the FP to expedite this process. After consenting to broad data sharing, families (and/or investigators) without a diagnosis may submit their existing phenotypic and genomic data through the FP (http://uwcmg. org/#/family), giving them access to in-house clinical and analysis expertise. All candidate genes/variants resulting from a standard analysis will be deposited in MyGene2. org, a publicly-accessible and family-friendly site that can be searched by gene and inheritance model. Users who identify a variant of interest may then contact the submitting families, investigators, or clinicians to follow-up. We encourage researchers, clinicians, diagnostic labs, and families worldwide to contribute data to Geno2MP and MyGene2, which will increase the utility of both databases for the entire human genetics and patient communities.

2861W

***De novo* variants in *CHAMP1* associated with neurodevelopmental abnormalities and dysmorphic features.** *A. Telegrafi[1], A. J. Tanaka[2], M. T. Cho[1], J. R. Jones[3], C. Nowak[4], J. Douglas[4], Y. H. Jiang[5], A. McConkie-Rosell[5], B. Schaefer[6], J. Kaylor[6], K. Retterer[1], G. Douglas[1], K. G. Monaghan[1], W. K. Chung[2].* 1) GeneDx, Gaithersburg, MD; 2) Columbia University Medical Center, New York, NY; 3) Greenwood Genetic Center, Greenwood, SC; 4) Boston Children's Hospital, Boston, MA; 5) Duke University Medical Center, Durham, NC; 6) Arkansas Children's Hospital, Little Rock, AR.

Next generation sequencing has enabled the identification of many novel genes contributing to Mendelian disorders. Previously, *de novo* loss-of-function variants in the *chromosome alignment-maintaining phosphoprotein* (*CHAMP1*) gene (MIM 616327) have been reported to be associated with non-syndromic intellectual disability and developmental disorders. *CHAMP1* is located on chromosome 13q34, encodes a mammalian zinc finger protein involved in kinetochore-microtubule attachment and is required for the proper alignment and segregation of chromosomes during mitosis. Here we provide detailed molecular and clinical characterization of four unrelated individuals with an as yet undefined disorder consisting of neurodevelopmental and dysmorphic features associated with novel *de novo* predicted pathogenic variants in *CHAMP1*. All patients were identified through whole exome sequencing (WES) after previous comprehensive diagnostic genetic testing on all patients had been unrevealing, and connection between the cases was facilitated through GeneMatcher. All of the patients had *de novo* heterozygous variants which were either nonsense or frameshift and predicted to be deleterious. No other variants identified via WES were deemed as likely candidates to explain the patients' phenotype. All variants were clustered in exon 3 in the *CHAMP1* gene (NM_001164144. 1) and none of the variants were present in the Exome Variant Server, 1000 Genomes, Exome Aggregation Consortium (ExAC) or our internal GeneDx database. Our patient cohort included three children and one adult ranging in age from 3-23 years old. The clinical phenotype was dominated by severe developmental delay, intellectual disability and dysmorphic features. The common dysmorphic features included upslanting palpebral fissures, epicanthal folds, hypertelorism, flat and broad nasal bridge, and short philtrum. Additional clinical features enriched among these patients include microcephaly in three of the patients, hypotonia, seizures, behavioral issues, sleep disturbances, and gastrointestinal issues. Our findings provide further evidence that predicted loss of function variants in *CHAMP1* are responsible for a severe neurodevelopmental phenotype characterized by developmental delay, intellectual disability, microcephaly, and dysmorphic features.

2862T

Left ventricular hypertrabeculation and gastropathy associated with *MIB2* variants altering NOTCH signaling. *S. Attanasio[1], P. Piccolo[1], R. Sangermano[1], C. Strisciuglio[2], E. Miele[2], G. Limongelli[3], M. Mutarelli[1], S. Banfi[1], T. Pons[4], S. Campione[5], F. D'Armiento[5], G. Nardone[6], A. Staiano[2], T. C. Lynnes[7], P. B. Celestino-Soper[7], K. G. Spoonamore[7], M. Vatta[7], N. Brunetti-Pierri[1,2].* 1) Telethon Institute of Genetics and Medicine, Pozzuoli, Italy; 2) Department of Translational Medicine, Federico II University, Naples, Italy; 3) Monaldi Hospital, Second University of Naples, Naples, Italy; 4) Structural Biology and BioComputing Programme, Spanish National Cancer Research Centre (CNIO), Madrid, Spain; 5) Advanced Biomedical Sciences, Federico II University, Naples, Italy; 6) Department of Clinical Medicine and Surgery, Gastroenterology Unit, Federico II University, Naples, Italy; 7) Department of Medical and Molecular Genetics and the Krannert Institute of Cardiology, Division of Cardiology, Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, USA.

We performed exome sequencing in patients from a family of Italian descent with an autosomal dominant gastropathy resembling Menetrier disease, a premalignant disorder of the stomach with epithelial hyperplasia and enhanced EGFR signaling due to local overproduction of TGF-α. Menetrier disease is believed to be an acquired disorder but its etiology is unknown. In affected patients, we found a missense p. V742G variant in *MIB2*, a gene regulating NOTCH signaling that has not been previously linked to human diseases. The mutation segregated with the disease in all nine affected members available for analysis. It affects a highly conserved amino acid residue, it was predicted to be deleterious, and it was found with a frequency of 0. 0015-0. 003 in control populations. In addition, the purified mutant protein showed reduced ubiquitination activity in vitro and white blood cells from patients exhibited reductions of *HES1, NOTCH3*, and *NRARP* expression reflecting alteration of NOTCH signaling. Because mutations of MIB1, the homolog of MIB2, have been recently found in patients with left ventricle non-compaction (LVNC), we screened members of the family with Menetrier-like disease for this cardiac abnormality by cardiac ultrasound and MRI. Left ventricular hypertrabeculation was detected in 2 out 4 members carrying the *MIB2* variant who were both free of cardiac symptoms at the time of evaluations. Finally, in a screening of *MIB2* mutations in a series of patients with LVNC, we found one affected patient carrying the p. V984L variant which was found in 1 out 10,2118 alleles of control populations. This variant was predicted to affect protein stability and when transfected in cells it resulted in significantly reduced levels of MIB2 protein compared to wild-type. The low frequency of the p. V742G variant in the control population and the rare occurrence of Menetrier-like syndrome suggest that *MIB2* mutations might not be the only cause for the gastropathy. Additional genes or environmental factors, such as *H. pylori* or CMV that have been previously associated with Menetrier disease, might be necessary for clinical expression of the gastropathy in addition to *MIB2* mutations. In conclusion, we report the first example of left ventricular hypertrabeculation with germline *MIB2* mutations resulting in altered NOTCH signaling that might be associated with a gastropathy clinically overlapping with Menetrier disease.

2863F

Mutations in VEGFR3 signaling pathway explain a third of familial primary lymphedema. *E. Fastre[1], M. J Schlögel[1], A. Mendola[1], N. Revencu[2], I. Quere[3], L. M. Boon[1,4], P. Brouillard[1], M. Vikkula[1,4,5].* 1) Human Molecular Genetics, de Duve Institute, Université catholique de Louvain, Brussels, Belgium; 2) Center for Human Genetics, Cliniques universitaires St-Luc, Université catholique de Louvain, Brussels, Belgium; 3) Department of Vascular Medicine, Montpellier University Hospital, Montpellier, France; 4) Center for Vascular Anomalies, Plastic Surgery Division, Cliniques universitaires St-Luc, Université catholique de Louvain, Brussels, Belgium; 5) Walloon Excellence in Lifesciences and Biotechnology (WELBIO), Université catholique de Louvain, Brussels, Belgium.

Background: Lymphedema, caused by dysfunction of lymphatic vessels, leads to disabling swelling that occurs mostly on the extremities. Lymphedema can be either primary (congenital) or secondary (acquired). Familial primary lymphedema commonly segregates in an autosomal dominant or recessive manner. It can also occur in combination with other clinical features. Nineteen mutated genes have been identified in different isolated or syndromic forms (Brouillard et al. , 2014). However, the prevalence of primary lymphedema that can be explained by genetic alterations in these genes is unknown. In this study, we investigated systematically 17 of these genes by targeted next generation sequencing. **Methods:** We screened 416 index patients, a quarter of which were from families with inherited primary lymphedema. A targeted NGS panel for IonTorrent (Personal Genome Machine, PGM) was designed. Data analysis was performed with the in-house developed Highlander software (Helaers et al. , in prep.). The filtering criteria were: checked manually with IGV (true), allelic depth proportion of the alternative allele (≥ 0. 25), less than 30 patients with the same change in our series, 1000 genome (< 10), goNL (≤ 5), consensus of impact predicted by 6 programs (≥ 3) and variant evaluation probably disease-causing. Validations and cosegregations were done using Sanger sequencing. **Results:** The filters allowed us to keep 227 variants among the approximately 19,500 detected. We limited this number to 100 by validations and cosegregation analyses, thus explaining the etiology in 24 % of the cases. We currently perform more detailed clinical phenotyping for those patients. We also identified 109 variants of unknown significance (VUS), which we are in the process of characterization. **Discussion:** The genetic cause of primary lymphedema remains unexplained for about 60 % of patients with a family history and 85 % of patients with sporadic or unspecified occurrence. Identification of the genes is important for understanding etiopathogenesis, stratification of management and generation of disease models for future development of therapies. Interestingly, most of the proteins that are encoded by the genes mutated in primary lymphedema seem to act in a common functional pathway involving VEGFR3 signaling. This underscores the important role this pathway plays in lymphatic development and function, and suggests that the unknown genes may also have a role in this pathway.

2864W

Mutations in YY1AP1 cause fibromuscular dysplasia in patients with Grange syndrome. *D. Guo[1], E. S. Regalado[1], L. Gong[1], X. Duan[1], L. Mellor-Crummey[1], R. L. P. Santos-Cortez[2], M. J. Bamshad[3], J. Shendure[3], D. A. Nickerson[3], S. M. Leal[2], A. Schinzel[4], K. Lieberman[5], A. Braverman[6], D. K. Grange[7], D. M. Milewicz[1], University of Washington Center for Mendelian Genomics.* 1) Department of Internal Medicine, University of Texas Health Science Center, Houston, Texas, USA; 2) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas; 3) Department of Genome Sciences, University of Washington, Seattle, Washington, USA; 4) Institute for Medical Genetics, University of Zurich, Zurich, Switzerland; 5) Pediatric Nephrology, Joseph M. Sanzari Children's Hospital, Hackensack University Medical Center, Hackensack, New Jersey, USA; 6) Cardiovascular Division, Department of Medicine, Washington University School of Medicine, St. Louis, Missouri, USA; 7) Division of Genetics and Genomic Medicine, Department of Pediatrics, Washington University School of Medicine, St. Louis, Missouri, USA.

Fibromuscular dysplasia (FMD) is a non-atherosclerotic arterial disease for which the molecular basis is unknown. FMD ischaracterized by arterial stenoses due to neointimal lesions rich with cells with a smooth muscle phenotype. It commonly affects the renal arteries, presenting as hypertension, and the internal carotid arteries, presenting as ischemic strokes. Grange syndrome is an autosomal recessive disease with FMD and additional complications, including variable brachydactyly and syndactyly, bone fragility, and learning disabilities. Exome sequencing was performed using DNA from three affected siblings with Grange syndrome and their parents. Previously described variant filtering strategies were used and variants with heterozygous genotypes in parents and homozygous or compound heterozygous genotypes in affected children were identified. The only variants that met these criteria were two *YY1AP1* (NM_001198903. 1) nonsense variants, c. 724C>T (p. Gln242*) and c. 2390T>G (p. Leu797*). The mother with the p. GLn242* variant has late onset hypertension and unilateral renal artery stenosis, suggesting that heterozygous YY1AP1 mutations may predispose to a milder form of FMD. To further investigate if mutations in *YY1AP1* cause Grange syndrome, two unrelated probands with Grange syndrome underwent Sanger DNA sequencing and each proband had a unique homozygous *YY1AP1* variant: c. 1903_1906delTCTG (p. Ser635Serfs) and c. 2401G>T (p. Glu801*); these variants are not found in the ExAC database. Two-point linkage analysis using Superlink software resulted in a combined LOD score of 2. 40. *YY1AP1* encodes YY1 associated protein 1, which is an activator of the transcriptional activity of ying yang 1 (YY1). *YY1* is rapidly induced in smooth muscle cells (SMCs) after vascular injury and suppresses SMC proliferation; conversely, loss of YY1 leads to increased SMC proliferation. YY1AP1 has not been extensively studied and encodes multiple isoforms. We determined that human smooth muscle cells express an isoform of 2391 bp, leading to an 88 kDa protein. The mutations identified in patients with Grange syndrome are predicted to lead to little to no expression of *YY1AP1*, decreased YY1 transcriptional activity, and thus increased SMC proliferation. In summary, *YY1AP1* is the first gene identified for FMD, and indicates that this condition can be due to underlying genetic alterations, leading to inappropriate hyperplasia of SMCs and occlusion of arteries.

2865T

Mutation spectrum in a Pulmonary Arterial Hypertension (PAH) cohort and identification of associated truncating mutations in *TBX4*. *C. Gonzaga-Jauregui[1], L. Ma[2], A. King[1], E. Berman Rosenzweig[2,3], U. Krishnan[2], J. G. Reid[1], J. D. Overton[1], F. Dewey[1], W. K. Chung[2,3], Regeneron Genetics Center.* 1) Regeneron Genetics Center, Regeneron Pharmaceuticals Inc. , Tarrytown, NY; 2) Department of Pediatrics, Columbia University Medical Center, New York, New York, USA; 3) Department of Medicine, Columbia University Medical Center, New York, New York, USA.

Pulmonary arterial hypertension (PAH) is a rare disorder characterized by sustained elevated pulmonary arterial pressure that can eventually lead to right ventricular hypertrophy with right heart failure and is often fatal. Genetic studies in families with PAH identified mutations in *BMPR2* as the major causative gene for the disorder, accounting for approximately 80% of cases. Additional genes (*SMAD9*, *CAV1*, *KCNK3*) have been implicated in a smaller percentage of familial cases, while PAH associated with hereditary hemorrhagic telangiectasia has been found in patients with mutations in *ENG* and *ACVRL1*. We performed whole-exome sequencing in a cohort of 60 probands with PAH and family members when available (180 total individuals) without a molecular diagnosis after most of the series was screened for mutations in *BMPR2*. We identified mutations in the known PAH genes *BMPR2* and *ACVRL1*. In addition we identified novel truncating variants in *TBX4* occurring *de novo* or inherited from an asymptomatic parent in 6 patients and a *de novo* predicted deleterious nonsynonymous variant in one additional patient. *TBX4* is a transcription factor of the T-box gene family. It is expressed in a variety of tissues during early mouse development including the atrium of the heart, the limbs, and the mesenchyme of the lung and trachea. TBX4, jointly with TBX5, has been shown to interact with FGF10 during lung growth and branching. Mutations in *TBX4* have been previously reported to cause small patella syndrome (SPS; MIM# 147891), an autosomal-dominant skeletal dysplasia characterized by patellar aplasia or hypoplasia and anomalies of the pelvis and feet, including disrupted ossification of the ischia and inferior pubic rami. Subsequently, another study (Kerstjens et al. , 2012) identified an association of *TBX4* mutations with PAH in 6 patients. We identified 7 different predicted deleterious mutations in *TBX4* (2 inherited and 5 *de novo*) in our cohort of patients ascertained for primary pulmonary arterial hypertension, accounting for ~10% of the probands in this series. Our results confirm the role of *TBX4* as an important cause of hereditary PAH.

2866F

Whole genome sequencing of cytogenetically balanced germline chromosomal rearrangements identify gene disruptions in both phenotypically normal and clinically affected individuals. *A. Lindstrand[1,2,3], M. Pettersson[1,2], P. Gustavsson[1,2,3], A. Fürster[1,2], W. Hofmeister[1,2], J. Wincent[1,2], B-M. Anderlid[1,2,3], A. Nordgren[1,2,3], O. Mäkitie[1,2,3,5,6], V. Wirta[7], F. Vezzi[8], J. R. Lupski[9,10], M. Nordenskjöld[1,2,3], E. Syk Lundberg[1,2,3], C. M. B. Carvalho[9], D. Nilsson[1,2,3,4].* 1) Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; 2) Center for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden; 3) Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden; 4) Science for Life Laboratory, Karolinska Institutet Science Park, Solna, Sweden; 5) Children's Hospital, Helsinki University Central Hospital and University of Helsinki, Helsinki, Finland; 6) Folkhälsan Institute of Genetics, Helsinki, Finland; 7) SciLifeLab, School of Biotechnology, KTH Royal Institute of Technology, Stockholm, Sweden; 8) SciLifeLab, Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden; 9) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston TX, USA; 10) Texas Children's Hospital, Houston TX, USA.

Approximately 1/500 individuals carry a balanced chromosomal aberration (BCAs) and are generally healthy but may suffer from infertility, repeated miscarriages and a risk of having children with unbalanced rearrangements. However, in a small fraction of cases the BCA may associate with disease, usually neurodevelopmental disorders (NDDs) ranging from mild (ADHD, dyslexia) to severe (autism, intellectual disability, ID). We investigated low coverage (5X) mate pair whole genome sequencing (WGS) to characterize translocation breakpoints and to elucidate mechanisms of formation; 52 junctions from 24 carriers of balanced translocations and inversions including both phenotypically normal (n=9) and abnormal individuals (n=15) were characterized. In contrast to previous publications only one significant *de novo* copy number variant was detected in the abnormal cohort (6. 7%). Further, genes were disrupted at the breakpoints to the same extent in both cohorts (45% for normal and 47% for abnormal; P=1. 0). However, differences were observed: in the normal individuals 56% of the disrupted loci were known disease genes all with a recessive inheritance pattern (i. e. *LARGE, COG7, VDR, ALMS1, OCA2*) while only 20% of gene disruptions in the affected individuals were known disease genes, all linked to dominant NDDs (i. e. *EXO6B, GRIN2B,* and *CTNND2*). For 21 junctions (40%), WGS split read analysis delineated the breakpoint. In the remaining 31 junctions paired-end analysis mapped the breakpoint to within 2 Kb. Breakpoint junction features revealed a consistent pattern of interspersed short homology involving both chromosomes but restricted to only one of the junctions. This type of pattern resembles template-switching between homeologous sequences in yeast which can lead to translocations. It has been hypothesized that "islands of microhomology" flanking the breakpoint junctions contribute to determine strand-invasion. Furthermore, one of the junctions is usually blunt or presents microhomology whereas the other one shows SNVs confined to the junction; a mutational signature of human recombination coupled replicative repair. We propose that at least two mechanisms underlie the formation of BCAs: a replication-based mechanism such as break induced replication (BIR) followed by microhomology mediated BIR (MMBIR) or non-homologous end joining (NHEJ). Alternatively, MMBIR with iterative template switches could parsimoniously explain our observations.

2867W

Combined targeted sequencing and array CGH of ciliary genes reveals increased mutational load in Bardet-Biedl syndrome and identifies *CEP76* as a novel driver of BBS. *M. Kousi[1], S. Frangakis[1], A. Lindstrand[2,3], A. Sabo[4], J. Lu[5], R. Lewis[6], J. R. Lupski[4,6,7,8], R. A. Gibbs[4], E. E. Davis[1], N. Katsanis[1].* 1) Cell Biology, Duke University, Durham, NC; 2) Dept of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; 3) Dept of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 5) Dept of Electrical Engineering, Duke University, Durham, NC; 6) Dept of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 7) Dept of Pediatrics, Baylor College of Medicine, Houston, TX; 8) Texas Children's Hospital, Houston, TX.

Bardet-Biedl syndrome (BBS) has served as a useful model to investigate pleiotropy, mutational burden, and ciliary biology. BBS is a multisystemic pediatric disorder characterized by retinal degeneration, obesity, hypogonadism, renal anomalies and learning disabilities. Further, BBS is frequently manifested under oligogenic inheritance paradigms in which *trans* alleles can modify the penetrance and/or expressivity of primary causal mutations. To date, mutations in >20 genes involved in the function of the cilium or its anchoring structure, the basal body account for ~75% of families of European descent. Here, we sought to resolve further the mutational landscape of BBS by a) identifying novel BBS candidate genes and b) evaluating comprehensively the combined effects of copy number variants (CNV)s and point mutations towards causing or modifying BBS phenotypes. We performed targeted sequencing of 785 genes encoding the ciliary proteome in a cohort of 100 individuals with BBS. In parallel, we used high-density array comparative genomic hybridization (aCGH) of 17 BBS, 20 intraflagellar transport (IFT) complex and 37 other causal ciliopathy genes to detect CNVs in the same cohort. We found that 80% of families harbor a recessive locus in a known BBS gene, suggesting that the remaining causal genes are likely to be ultra-rare contributors to the disorder. Among the affecteds for whom a known BBS locus did not segregate with the disease, we identified *CEP76*, encoding a centrosomal protein critical to cell cycle, as a novel primary causal locus in one pedigree. *In vivo* zebrafish assays that recapitulate the human retinal and renal phenotypes provided further supportive evidence for the physiological relevance and likely causal role of this locus in disease. Moreover, by comparing mutational burden between BBS cases and controls, we show that CNVs in BBS genes are enriched significantly in cases (18. 5% of our cohort) demonstrating a previously underappreciated role of such variants. To complement these data, we are evaluating the total number and effect of single nucleotide variants in BBS to test whether burden and/or specific pairwise combinations can be predictive of clinical outcome. Together, we have expanded the genetic spectrum of BBS through the discovery of *CEP76* as a novel causative locus, and through our mutational load studies, aim to better understand the disease mechanisms not only in BBS but also in the broader spectrum of ciliopathies.

2868T

Whole Exome Sequencing (WES) to Analyze the Genetic Basis of Cleft Lip and Palate. *M. Basha[1], M. Quentric[1], R. Helaers[1], N. Revencu[2], B. Bayet[3], M. Vikkula[1,2].* 1) Human Molecular Genetics, de Duve Institute, Université catholique de Louvain, Brussels, Europe, Belgium; 2) Center for Human Genetics, Cliniques universitaires St-Luc, Université catholique de Louvain, Brussels, Europe, Belgium; 3) Centre Labiopalatin, Division of Plastic Surgery, Cliniques universitaires St-Luc, Université catholique de Louvain, Brussels, Europe, Belgium.

Background: Cleft lip with or without cleft palate (CL/P) is the most common craniofacial birth defect with an incidence of ~ 1/700 live births, varying with ethnicity and cleft type. It is a debilitating condition requiring an expensive and lifelong treatment. CL/P can occur isolated or as part of a syndrome. The etiology is still largely unknown. **Methods**: We performed whole exome sequencing (WES) on 85 CL/P subjects. The cohort predominantly consisted of non-syndromic (n=21/25) families. We selected at least two most distant relatives per family for WES (n=36). We also included few sporadic CL/P cases (n=12). The sub-phenotypes of these 73 patients ranged from a full- blown bilateral CL/P to a subtler velopharyngeal insufficiency (VPI). In addition, we WESed 8 unrelated individuals all affected with the rare Cerebro-oculo-nasal syndromethat has cleft lip and high arched palate as part of its spectrum. For two of the eight, parents were also WESed. **Results**: After bioinformatic processing of raw data with an in-house developed pipeline, we analyzed our samples on Highlander (a software developed in-house by R. Helaers). We retained variants for further analysis if they met the following criteria: (i) passed the GATK filter, (ii) reported allele frequency of ≤1% in the population from the 1000 genomes project, (iii) not reported in the population from the GO-NL (genome Netherlands) project, (iv) not reported in Exome Aggregation Consortium (ExAC), (v) and not occurring in our in-house WES controls (subjects affected with =different condition than CL/P). As "likely pathogenic" variants, we considered those with a high impact (premature termination codon=PTC) or a moderate impact (non synonymous-NS), as predicted by SnpEff software. On average, each family harbored 1 PTC and 30 NS variants. All NS variants were predicted as damaging by at least 3 softwares. Approximately 5 changes co-segregated with the phenotype. This led to identify the causative mutation in 3 cases in known CL/P genes (IRF6 and TBX1). For the remaining samples characterization of variants of interest is ongoing. **Discussion**: To be able to distill out the causal gene from the aforementioned data, we will WES additional affected subjects from multiplex families, for which continued collection is also ongoing.

2869F

Exome Sequencing in a Multiplex Family with Cleft Palate Only Identifies a Novel Damaging Mutation in *ARHGAP29*. *A. Butali[1], T Busch[1], I Gowans[1], H Liu[1], D Anand[2], S Lachke[2], N Nidey[1], B Amendt[1], R Cornell[1], J Murray.* 1) UNIVERSITY OF IOWA, IOWA CITY, IA; 2) Department of Biological Sciences.

Background: With advances in genomics research and availability of next generation sequencing approaches, scientists are now able to identify risk loci across the exome and the genome. In the current study, we present findings from an exome study conducted in a multiplex family with cleft palate only (CPO). Methods: Exome sequencing was conducted in 5 individuals with high quality DNA in a family of 8 affected individuals. The Ilumina HiSeq 2000 was used for sequencing and Agilent Sure Select for enrichment capture. Quality control (QC) steps included aligning raw sequence to the human genome build hg19 using BWA software, SNP and In/Dels analyses using PICARD, SAMTOOLS and GATK. WANNOVAR was used to functionally annotate the SNPs and In/Dels. Several filtration methods were applied in order to identify variants with minor allele frequencies ≤1% in 1000genomes, EVS and dbSNP138 including clinically associated variants that segregate in affected individuals. The bioinformatics tool *SysFACE* (Systems tool for craniofacial expression-based gene discovery) was used to prioritize cleft candidates from our list of variants. Sanger sequencing was used to validate identified variants in affected and unaffected relatives. Results: A yield of 3. 4Gbases per sample passed the quality control step. A missense variant predicted to be probably in exon 15 of *ARHGAP29* (pSer552Pro) was identified. This variant segregates in all 5 individuals. The pSer552Pro variant is novel and not in all of the known databases and has never been previously reported. This variant was validated using Sanger sequencing and none of the 4 unaffected individuals including a CEPH control carried the variant. We also identified 2 clinically associated variants in *FRGB3* gene (rs145018822 and rs201563622) previously reported for malignant melanoma in the ClinVar study. Functional experiments are ongoing to understand the biological effect of the pSer552Pro variant in Zebrafish and cell based assay.

2870W

Whole Exome Sequencing in a Family with Primary Open Angle Glaucoma. *HF Nunes[1], JPC Vasconcellos[2], I Lopes-Cendes[3], BS Carvalho[3], MG Borges[3], NIT Zanchin[4], MB Melo[1].* 1) Center of Molecular Biology and Genetic Engineering, University of Campinas; Campinas SP, Brazil; 2) Department of Ophthalmology, Faculty of Medical Sciences, University of Campinas - UNICAMP, Campinas SP, Brazil; 3) Department of Medical Genetics, Faculty of Medical Sciences, Unicamp, Campinas, SP, Brazil; 4) Fiocruz, Curitiba PR, Brazil.

Background: Primary open-angle glaucoma (POAG) is a chronic neurodegenerative disease that leads to progressive damage of retinal ganglion cells resulting in visual field loss. Glaucoma is recognized as the main cause of irreversible blindness worldwide. Although the pathophysiology of glaucoma is not well understood, positive family history is one of the most important risk factors for its development. **Objective:** To identify genes associated with POAG in one family from the state of São Paulo, Brazil, with no mutations in the MYOC gene. **Methods:** Four individuals were analyzed through Whole Exome Sequencing (WES), including two affected and two unaffected. Nextera Rapid Capture Expanded Exome kit (Illumina™) was used to obtain 50X average coverage per sample. Sequencing was performed in a high-performance Hiseq Illumina 2500 DNA Analyzer (Illumina™). Bioinformatics analysis was performed using the GATK software package. Sequences were aligned using BWA algorithm and variant calling and functional prediction was performed using Variant Annotator and SnpEff tools. Before applying any filtering method, the main genes related with glaucoma were evaluated for the presence of mutations (WDR36, OPTN, NTF4, ASB10). **Results:** Filtering process started considering heterozygous mutations in autosomes segregating in affected individuals, which totalized 7951 variants. Next step included only non-synonymous, splice site, small indels and stop codon variants, remaining 2346. Considering those with relative pathogenicity 367 remained and out of these 42 were present in less than 1% of the general population. The last step was based on ocular tissue expression and protein function, resulting in 6 genes that require confirmation through Sanger sequencing in other members of this family. **Conclusion:** Properties related to apoptotic processes, aqueous humor production and outflow make the six candidate genes identified relevant for POAG pathophysiology. Further resequencing and experimental verification are still needed to confirm the causative gene in this family. **Supported by:** FAPESP.

2871T

Functional evidence for the involvement of *WNT10A* in failure of tooth development in humans and zebrafish. *M. Zhao[1,2], Q. Yuan[3], X. Liu[5], A. Zhang[1,2], R. Silva[1,2,4], J. T. Hecht[2,3,4], E. C. Swindell[3,4], A. Letra[1,2,4].* 1) Department of Endodontics, University of Texas Health Science Center School of Dentistry, Houston, TX; 2) Center for Craniofacial Research, University of Texas Health Science Center School of Dentistry, Houston, TX; 3) Department of Pediatrics, University of Texas Health Science Center Medical School, Houston, TX; 4) Pediatric Research Center, University of Texas Health Science Center Medical School, Houston, TX; 5) Department of Human Genetics, University of Texas Health Science Center School of Public Health, Houston, TX.

Tooth agenesis is the most common developmental abnormality in humans. Severe tooth agenesis (oligodontia) is characterized by the absence of six or more permanent teeth, and can be syndromic or nonsyndromic, sporadic or segregating in families. Candidate gene sequencing and association studies have suggested *MSX1, PAX9, AXIN2,* and recently *WNT10A* as etiologic genes in oligodontia. We used whole-exome sequencing (WES) to identify the cause of nonsyndromic oligodontia in a 9-year old girl missing eleven permanent teeth. No defects in tooth shape, hair, skin or nails were noted in the child or unaffected parents. Exome capture was performed using Agilent Sure Select 50MB Target Enrichment System followed by massive parallel sequencing in an Illumina HiSeq 2000 sequencing platform. Paired-end reads were aligned against the GRCh37 human reference genome using the Burrows-Wheeler Aligner. WES data was annotated using the dbNSFP database. ANNOVAR was used to annotate indels and identify variants causing frameshifts. A novel compound heterozygous missense mutation (c. 237G>A; c. 1070C>T), in *WNT10A* was identified in the child and predicted to be highly deleterious and possibly leading to gene loss-of-function. Sanger sequencing validated these mutations in the child while revealing the carrier status for each mutation in the unaffected parents. Zebrafish was used for functional analysis of *wnt10A* during tooth development. Using real-time PCR, we detected significantly higher *wnt10a* mRNA expression level mRNA expression during craniofacial development stages in zebrafish embryos. Two antisense morpholinos (MO) were used to block wnt10a translation and splicing in one-to two cell stage zebrafish embryos, which showed similar survival rates to uninjected controls at 1dpf. Thereafter, survival rates decreased albeit not significantly different than UIC. While MO-injected embryos had no gross whole-body abnormalities, tooth agenesis was arrested in 91% of the fish by 5dpf. In contrast, 100% of UIC presented normal teeth. Our results show a novel compound heterozygous mutation in *WNT10A* as the cause of recessive oligodontia and further provide the first functional evidence for a role of *WNT10A* in tooth development.

2872F

Deletions of the 5' HOXC genes are associated with lower extremity abnormalities including clubfoot and vertical talus. *D. M. Alvarado[1], M. B. Dobbs[1,2], C. A. Gurnett[1,3,4].* 1) Orthopaedic Surgery, Washington Univ, St Louis, MO; 2) Shriners Hospital for Children, St. Louis, MO; 3) Neurology, Washington University, St. Louis, MO; 4) Pediatrics, Washington University, St. Louis, MO.

Deletions of the HOXC gene cluster result in variable phenotypes in mice, but have been rarely described in humans. Here, we report four families with 30-176 kb chromosome 12q13. 13 microdeletions involving the 5'HOXC gene cluster that segregate with a spectrum of severe congenital lower limb birth defects, including clubfoot, vertical talus, hip dysplasia, lower extremity weakness, nail hypoplasia, and early-onset arthritis. Defects in limb morphogenesis are supported by the presence of shortened and overlapping toes, as well as peroneus muscle hypoplasia seen in two affected relatives with magnetic resonance imaging . While these deletions share minimal overlap within a noncoding region downstream of HOXC13, phenotypes vary depending upon inclusion of HOXC13, HOXC12 or the lncRNA HOTAIR within the deletion. Furthermore, we demonstrate reduced 5' HOXD gene expression in a fibroblast cell line from a patient with 175 kb 5' HOXC deletion, consistent with prior studies demonstrating that dosage of lncRNA HOTAIR alters expression of HOXD10 and HOXD11 genes in trans. Because HOXD10 has previously been implicated in the etiology of congenital vertical talus, variation in its expression may contribute to the lower limb phenotypes occurring with 5' HOXC microdeletions.

2873W

The role of combined SNV and CNV burden in patients with distal symmetric polyneuropathy. *C. R. Beck[1], D. Pehlivan[1,2], Y. Okamoto[1], T. Harel[1], S. N. Jhangiani[3], M. A. Withers[1], M. T. Goksungur[4], C. M. B. Carvalho[1], D. Czesnik[5], C. Gonzaga-Jauregui[1], W. Wiszniewski[1], D. M. Muzny[3], R. A. Gibbs[3], B. Rautenstrauss[6,10], M. W Sereda[5, 7], J. R. Lupski[1,3,8,9].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Section of Pediatric Neurology, Texas Children's Hospital, Houston, TX, USA; 3) Human Genome Sequencing Center, BCM, Houston, Texas; 4) Department of Neurology, Istanbul University, Istanbul Medical Faculty, Istanbul, Turkey; 5) Department of Clinical Neurophysiology, University Medical Center Göttingen, Göttingen, Germany; 6) Medizinisch Genetisches Zentrum, Munich, Germany; 7) Max-Planck Institute of Experimental Medicine, Göttingen Germany; 8) Department of Pediatrics, BCM, Houston, Texas USA; 9) Texas Children's Hospital, Houston, Texas USA; 10) deceased.

Charcot-Marie-Tooth (CMT) represents a heterogeneous group of genetic disorders of the peripheral nervous system. About 70 genes and 10 linkage regions have been identified for CMT-like distal symmetric peripheral neuropathies. Copy number variants (CNVs) are an important cause of CMT, as duplication of *PMP22* underlies ~70-80% of CMT1 cases. Indeed, a recent study of >17,000 neuropathy patients (DiVincenzo, *et al*. 2014, PMID25614874) found that *PMP22* CNVs accounted for ~80% of molecular diagnoses (dup 56. 7% and del 21. 9%). We hypothesized that CNVs involving genes other than *PMP22* and/or single nucleotide variants (SNVs) might exist in CMT patients where the molecular genetic etiology is not elucidated. Two hundred CMT patients who were negative for both SNV mutations in several CMT genes and for CNVs involving *PMP22* were screened for CNVs by high-resolution oligonucleotide array comparative genomic hybridization. Putatively causative CNVs were identified in five subjects (~2. 5%); four of the five map to known neuropathy genes. Breakpoint sequencing for each of these CNVs revealed *Alu-Alu* mediated junctions as a predominant contributor (present at 3/5 junctions). Exome sequencing additionally identified *MFN2* SNV in two of the individuals that are likely to underlie the neuropathy phenotypes. In one individual, a *de novo* CNV in *PRICKLE1* combines with a *de novo* point mutation in *MFN2* to give rise to a composite neurological phenotype including tonic-clonic seizures, developmental delay, and neuropathy. Our findings suggest that neuropathy-associated CNV outside of the *PMP22* locus are rare in CMT cohorts, and that these non-recurrent rearrangements are largely driven by *Alu*-mediated mechanisms. Nevertheless, there is potential clinical utility in testing for CNVs in CMT cases negative for the CMT1A duplication particularly if exome sequencing has also been conducted. Moreover, our findings suggest that complex phenotypes including neuropathy can be caused by a combination of SNV and CNV affecting more than one disease-associated locus.

2874T

New genetic insights into the spectrum of disorders of glycosylation: a patient with multiple congenital anomalies. *H. H. Arts[1,2], M. M. Oud[1], N. van Vlies[3], A. de Bruin[4], R. J. A. Wanders[5], Z. Ren[4], M. Alders[5], D. A. Braun[6], J. Lawson[6], F. Hildebrandt[6], S. T. Pals[4], H. G. Brunner[1], J. Cobben[7], R. Roepman[1], T. W. Kuijpers[7].* 1) Department of Human Genetics, Radboudumc, Nijmegen, Netherlands; 2) Department of Biochemistry, Robarts Research Institute, University of Western Ontario, London, Ontario, Canada; 3) Department of Pediatric Metabolic Diseases, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; 4) Department of Pathology, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands; 5) Laboratory of Genetic Metabolic Diseases, Academic Medical Centre, Amsterdam, The Netherlands; 6) Department of Medicine, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts; 7) Department of Pediatrics, Academic Medical Center University Hospital, Amsterdam, The Netherlands.

Introduction: Glycosylation is essential for human development. More than 100 clinically diverse inherited disorders are known to result from glycosylation defects. One subtype of the congenital disorders of glycosylation (CDG) is caused by defects in glycosaminoglycan (GAG) synthesis. Here, we describe an infant from a consanguineous family with a GAG-CDG. Phenotypic features include respiratory insufficiency, discrete skeletal anomalies, renal insufficiency, and T-cell depletion. **Methods:**We performed whole-exome sequencing (WES) on patient leucocyte DNA to identify the causative mutation. Immunocytochemistry was conducted to characterize the identified mutations functionally in patient-derived and control fibroblasts. The antibody that was used was validated with CRISPR/Cas9 experimentation. Synthesis of different GAG chains was measured in fibroblasts and urine from the patient and several controls. Mutated genes were screened with next-generations sequencing in a cohort of 96 patients with nephronophthisis and skeletal involvement. **Results:** Through WES we detected two homozygous missense mutations in highly conserved regions of two glycosylation genes regulating GAG synthesis. It remains unclear whether mutations in one or both of the genes contribute to the clinical phenotype. We found that one of the mutated proteins that normally localizes to the Golgi apparatus was absent in this organelle in patient cells. Immunocytochemical analysis of the other mutated protein failed. Finally, heparan sulphate GAG chains were significantly reduced in patient-derived fibroblasts, though within the normal range in urine. Mutation screening of both mutated genes in a cohort of 96 patients with renal and skeletal defects did not identify additional mutations. **Conclusion:** We describe an infant with multiple congenital anomalies. WES analysis revealed mutations in two glycosylation genes. One of the mutated proteins mislocalizes in patient-derived fibroblasts. Our data indicate that this patient has a rare GAG-CDG.

2875F

***De novo* mutations in novel histone and epigenetic regulator genes cause multiple genetic syndromes.** *D. Li[1], E. J. Bhoj[1,2], L. Tian[1], M. H. Harr[2], H. A. Dubbs[2], C. Hou[1], C. E. Kim[1], F. D. Mentch[1], R. Chiavacci[1], E. McCormick[2,4], M. J. Falk[2,3,4], E. H. Zackai[2,3], H. Hakonarson[1,2,3].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA; 2) Department of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Division of Metabolic Disease, The Children's Hospital of Philadelphia, Philadelphia, PA.

The post-translational modification of histones is emerging as an important mechanism for regulation of organism development and associated cellular function. Epigenetic dysregulation is observed in many human diseases, including cancer and neuropsychiatric disorders. Here we report genetic mutations in genes that encode histone and epigenetic regulator cause monogenic syndroms with complex phenotypes. We obtained clinical data and performed whole exome sequencing (WES) for 2 kindreds. The affected boy in the first family had 1-2-3 toe syndactyly on the left foot, agenesis of the corpus callosum, atrial septal defect, micrognathia, broad thumbs, microcephaly, failure to thrive, developmental delay, very short stature, and craniosynostosis. *H3F3A*, H3 histone family 3A, was the leading disease-causing candidate because histones play a central role in transcription regulation, DNA repair, DNA replication and chromosomal stability. Recent studies have emphasized its role during early development processes in mouse models that involves both major genome reorganization, discrete changes to sustain cell fate specification, and pluripotency. A dominant zebrafish mutant (D123N in *h3f3a*) showed variable reductions of the jaw-support skeleton, where wild type h3f3a rescued head skeletal development in mutants. Functional assessment in the zebrafish model is underway to elucidate the role our detected *H3F3A* mutation (p. A16G) plays in multiple congenital anomalies. A second, unrelated proband had developmental delay with no speech, congenital hypotonia, congenital heart disease, pulmonary hypertension, chronic lymphocytopenia and multiple dysmorphic features (sagittal craniosynostosis, midline cleft, hypertelorism, large hands and feet, long fingers and toes). He had a nonsense mutation (R1019X) in *KAT6A*, a histone acetyltransferase (HAT) gene, which was recently implicated in a cohort with variable findings of many overlapping clinical features. Knockout mice in *kat6a* show H3K9 hypoacetylation and reduced transcription at Hox gene loci. Our identification of mutations in *H3F3A* and *KAT6A* in patients with multiple congenital anomalies emphasizes the importance of epigenetic control of gene expression programs in diverse human disorders, including syndromic disease.

2876W

Whole exome sequencing detects variants in *MBTPS2* and *ITGB4* in a Brazilian patient with severe ichthyosis congenita and limb malformations. *M. P. Migliavacca[1], S. Raskin[2], J. Jurgens[3], D. Valle[3], N. L. M. Sobreira[3].* 1) Clinical Genetics, Department of Morphology and Genetics, Universidade Federal De São Paulo, São Paulo, SP, Brazil; 2) Genetika Laboratory and Pontifícia Universidade Católica do Paraná, Curitiba PR, Brazil; 3) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 733 N. Broadway Street Suite 512, Baltimore, MD 21205, USA.

The IFAP syndrome with or without BRESHECK syndrome (MIM 308205) is a phenotypically variable X-linked multiple congenital anomaly characterized by ichthyosis follicularis, atrichia, and photophobia (IFAP) with or without additional features affecting the skeletal, genitourinary, and central nervous systems. This phenotype is caused by variants in *MBTPS2*, which encodes a developmentally integral zinc metalloprotease. Heterozygous or homozygous variants in the integrin beta 4 gene *ITGB4* are associated with epidermolysis bullosa, which is distinguished by severe cutaneous symptoms. Here we describe a Brazilian patient harboring a *de novo* variant in *MBTPS2* and compound heterozygous variants in *ITGB4* with developmental anomalies including intrauterine growth retardation, microcephaly, patent foramen ovale,ichthyosis congenita, atrichia, cryptorchidism, and limb malformations including bilateral radius aplasia. We performed whole exome sequencing on the proband and his unaffected parents through the Baylor-Hopkins Center for Mendelian Genomics. Selecting for rare (MAF<0. 01), functional single nucleotide variants and indels, we identified a novel hemizygous *de novo* variant in exon 6 of *MBTPS2* (c. 727C>T, Pro243Ser) and compound heterozygous variants in *ITGB4* (c. 1940C>T, Thr647Met; c. 4217G>C, Arg1406Pro). Using Sanger sequencing, we validated the presence of the *MBTPS2* variant in the proband and absence of the variant in his unaffected mother, father, and brother. Sanger sequencing also confirmed that the unaffected mother, father, and brother of the proband each had only one of the two *ITGB4* variants found in the proband. Pietrzak et al. (2012) described a Polish patient with an *MBTPS2* variant and similar cutaneous and skeletal features to those observed in our patient, although our patient has much more severe skin anomalies. Interestingly, our patient also has features overlapping those of epidermolysis bullosa. While the phenotype of our patient may be attributed to the *MBPTS2* variant alone, his skin phenotype may also be caused by the combination of the *MBPTS2* and *ITGB4* variants.

2877T

Targeted NGS reveals a high incidence of digenic inheritance in monogenic diabetes and congenital hyperinsulinism. *A. Campos-Barros[1,2], A. De la Peña-Abad[1], A. Del Pozo[1,2], K. Ibáñez[1], J. C. Silla[1], V. E. F. Montaño[1], A. Gómez-Núñez[1], P. D. Lapunzina[1,2], K. E. Heath[1,2], E. Vallespín[1,2], ENDOSCREEN-CM CONSORTIUM.* 1) INGEMM (Institute of Medical & Molecular Genetics), IdiPAZ, Hospital Universitario La Paz, UAM, Madrid, Madrid, Spain; 2) CIBER de Enfermedades Raras (CIBER-ER) U 753, ISCIII, Madrid, Spain.

Background: Monogenic diabetes (MoDiab) and congenital hyperinsulinism (CHI) constitute a group of Mendelian genetic diseases caused by gain/loss-of-function mutations resulting in an inappropriate secretion of insulin by the pancreatic β-cells, which may cause hypo- (CHI) or hyperglycaemia (MoDiab). *Aim:* The overlapping clinical features and phenotypic heterogeneity of MoDiab andCHI patients, together with the increasing number of implicated genes, makes a genetic diagnosis often difficult to obtain using a traditional gene cascade sequencing approach. Therefore, we aimed to design and validate a targeted NGS assay to screen 13 selected candidate genes involved in the etiology of MoDiab or CHI. *Methods:* An exon-capture assay was designed to enrich coding regions and splice sites of *ABCC8, GCK, HNF-1α, HNF-4α, HNF-1β, INSR, KCNJ11, AKT2, GLUD1, HADH, SLC16A1, UCP2,* and *DIS3L2* relevant transcripts. To evaluate its performance, a total of 24 patient samples with 24 known mutations, were retested. *Results:* 99% of captured bp had a coverage >20x; 96% of exons were well captured, and 100% of the previously identified 24 variants were detected, including 17 SNVs, 7 indels (<21 bp) and 2 CNVs (1x whole gene deletion and 1x intragenic duplication). Interestingly, this targeted NGS approach allowed the identification of a second genetic hit in 7/24 (29. 2%) of the retested samples, in which five different coding variants, potentially relevant to the pathological entities under study, were identified and confirmed by Sanger sequencing: NM_152383. 4(*DIS3L2*): c. 1448G>A (p. Arg483Gln) (n=3); NM_152383. 4(*DIS3L2*): c. 410A>G (p. Tyr137Cys); NM_005271. 3(*GLUD1*): c. 1568G>A (p. Arg523His); NM_000162. 3(*GCK*): c. 56T>A (p. Ile19Asn); and NM_000208. 2(*INSR*): c. 3034G>A (p. Val1012Met). *Discussion:* Though cases of digenic inheritance in MoDiab and CHI have been considered to be very rare, our preliminary results obtained during the validation of a targeted NGS assay indicate that their incidence may be under ascertained due to the traditional sequential single candidate gene testing strategy, which implies that once a mutation in one of the candidate genes is found, the analysis of additional genes is not undertaken. The implementation of NGS will reveal the true prevalence of digenic/polygenic inheritance in these disorders and the influence of co-inheritance of additional mutations/variants which may act as modifiers on the clinical phenotypic expression.

2878F

Molecular characterization of Epidermolysis Bullosa in Iran. *H. Vahidnezhad[1,2,3], L. Youssefian[1,2], S. Zeinali[3], MR. Barzegar[4], S. Sotoudeh[2], A. Ertel[1], N. Mozaffari[4], P. Fortina[1], J. Uitto[1].* 1) Thomas Jefferson University, Philadelphia, PA; 2) Tehran University of Medical Sciences, Tehran, Iran; 3) Pasteur Institute of Iran, Tehran, Iran; 4) Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Epidermolysis bullosa (EB) is a clinically and genetically heterogeneous group of heritable blistering diseases caused by mutations in at least 18 genes. There are 4 types of EB based on the ultrastructural level of tissue separation: EB simplex (EBS), junctional EB (JEB), dystrophic EB (DEB), and the Kindler syndrome (KS). There is currently no comprehensive study about the molecular epidemiology of EB in Iranian populations. We exploit the principles of whole-genome homozygosity mapping via 240K Human Exome Beadchip and immunofluorescence mapping as screening methods before formal mutational analysis in an attempt to facilitate the molecular diagnosis of EB. In the present study, we are assessing a cohort of 181 DEB patients for pathogenic sequence alterations in the *COL7A1* gene with 118 exons. Results from Sanger sequencing of 124 patients show 46 distinct mutations, 23 being previously unreported. Among the mutations 14 were missense, 14 small indels, 4 splice-junction mutations, 11 nonsense, and 3 large indel mutations. Targeted sequencing of COL7A1 was also performed on the IonTorrent PGM, using an AmpliSeq 3-pool primer design designed to capture all *COL7A1* exons plus 50-bp flanking regions in 57 patients disclosed 34 different mutations. A recurrent c. 6269_6270delC (p. Pro2090LeufsX116) was found in 14 lorish and azeri families. Haplotype analysis of the 20 Mb flanking region of COL7A locus on chromosomal region 3p21. 1 identified 2. 5Mb conserved haplotypes in all 14 families indicating a founder effect. Sanger sequencing of *KRT5* and *KRT14* in 10 EBS families revealed 9 recurrent and novel mutations in 8 families. We also examined a total of 21 Iranian families with KS, including 70 affected individuals. Sequencing of *FERMT1* revealed 14 distinct mutations in 17 families, 12 of the mutations being novel. Two novel mutations, p. Q226X and c. 1253delA, were found in more than one family (4 and 2, respectively). Haplotype analysis of the 9 Mb flanking region of the *FERMT1* locus in these families indicated the presence of a conserved haplotype, indicating founder effect. In two families, no mutation in *FERMT1* gene was found and autozygosity mapping excluded this locus. Careful examination of patients in these two families revealed unusual clinical features suggesting new entities with clinical overlapping features with KS. Our data emphasize the need for population-specific diagnostic and management approaches for EB.

2879W

***De novo* truncating mutations *MSL3* cause a new recognizable syndrome.** *A. L. Bruel[1,2], I. van der Burgt[3], J. Morton[4], J. St-Onge[1,2], S. El Chehadeh-Djebbar[5], A. Hoischen[3], C. Gilissen[3], L. Vissers[3], R. Pfundt[3], C. Thauvin-Robinet[1,2], H. G. Brunner[3], J. B. Rivière[1,2], L. Faivre[1,2], J. Thevenon[1,2].* 1) Team EA4271 GAD, Université de Bourgogne, Dijon, Bourgogne, France; 2) FHU-TRANSLAD, Université de Bourgogne/CHU Dijon, Dijon, Bourgogne, France; 3) Clinical Genetics Unit, Birmingham Women's Hospital, Birmingham, United Kingdom; 4) Radboud University Medical Center, Department of Human Genetics, Nijmegen, The Netherlands; 5) Centre de Génétique, CHRU de Strasbourg, Strasbourg, France.

Whole-exome sequencing has proven to be successful for the identification of the molecular basis of heterogeneous conditions such as intellectual disability with congenital anomalies. A large number of results remain non-conclusive because the diagnosis of ultra-rare conditions limits the genotype-phenotype correlations. Broad international data-sharing catalyzes the identification of additional cases and allows the delineation of new syndromes. A female patient presented with developmental delay and moderate intellectual disability, absent speech, facial dysmorphism, severe constipation. Paraclinic and genetic investigations did not orientate the diagnosis. A proband whole-exome sequencing was performed with no causative variant was identified with a diagnostic interpretation. Data were further processed in a research setting looking for compound heterozygous or *de novo* variants with parental segregation by means of Sanger sequencing. A *de novo* acceptor-splice variant of *MSL3* was identified. The *MSL3* gene, located on the X chromosome encodes a transcriptional regulator implicated in the KAT8-complex and directly interacting with *KANSL1*. To assess the relevance of this finding, international collaborators were asked for patients with *de novo MSL3* truncating variants. Two additional European cases were identified with a *de novo* frameshift variant affecting the same splice-site in a female patient and *de novo* non-sense variant in a male patient. Reverse phenotyping of the cases evidenced a recognizable facial dysmorphism, developmental delay, intellectual disability and severe constipation. Interestingly, de *novo* loss-of-function of *KANSL1* causes the Koolen-De Vries syndrome, with clinical features overlapping the reported condition. Based on molecular and clinical evidences, we report on the first implication of *MSL3* in a human disorder with *de novo* truncating variants causing a new recognizable ultra-rare condition.

2880T

Investigating the genetic architecture of Congenital Hypothyroidism using next-generation sequencing. *E. Goncalves-Serra[1], A. Nicholas[2], VK. Chatterjee[2], CA. Anderson[1], N. Schoenmakers[2], UK10K Consortium.* 1) Human Genetics, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, Hinxton, United Kingdom; 2) University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, United Kingdom.

Congenital hypothyroidism (CH) is the most common endocrine disorder among newborns but its genetic aetiology is still largely unknown, with mutations identified in less than 20% of the cases. The increased CH frequency in consanguineous populations and in relatives of CH cases suggests the genetic contribution to CH might be greater than previously expected. Known CH genes have not previously been screened comprehensively in a CH cohort with gland-in-situ (GIS). We evaluated the relative contribution of mutations in eight known causative genes (*TG*, *TPO*, *DUOX2*, *DUOXA2*, *SLC5A5*, *SLC26A4*, *IYD* and *TSHR*) in 50 CH cases with GIS from 35, ethnically diverse families, using exome and targeted next-generation sequencing. Patient genotypes were correlated with biochemical phenotype and pathogenicity of mutations was analysed in silico using deleteriousness and conservation scores (PolyPhen2, SIFT and GERP) and protein structural modelling. Twenty-eight cases harboured likely, disease-causing, mutations. Nineteen cases with single gene defects, involving *TG* (12 cases), *TPO* (4 cases), *DUOX2* (2 cases) and *TSHR* (1 case), were documented. Novel variants were identified in *TG* (5 variants), *TPO* (5 variants) and *DUOX2* (2 variants). Nine cases harboured pathogenic variants in two different genes: *TG* and *TPO* (1 case), *SLC26A4* and *TPO* (2 cases) and *DUOX2* and *TG* (6 cases), yet no clear correlation between genotype and biochemical phenotype was found in these cases. We demonstrated that our case set is enriched for mutations in known-CH genes when compared to ~2,000 exome-sequenced healthy individuals and that the probability of double-hit mutations in the above gene pairs was unlikely to have occurred by chance. Genetic aetiology was not ascertained in 22 patients, generally with milder biochemical CH and some familial cases. Exome-wide gene-based burden tests in these cases and controls did not reveal novel CH genes. Still, gene prioritization analyses are ongoing to highlight biologically relevant candidate genes to be included in future screens of larger CH-GIS cohorts. In conclusion, the aetiology of CH with GIS is complex, with only 56% being due to mutations in known CH-genes. Combinations of defects in two different genes are common, especially in consanguineous families (57%). Severe CH is most commonly mediated by biallelic *TG* or *TPO* mutations. A high proportion (~44%) of unsolved cases suggest contribution of hitherto unknown genes or environmental factors to GIS CH.

2881F

***De novo* mutations in the eukaryotic translation elongation factor, EEF1A2 cause epileptic encephalopathy.** *G. L. Carvill[1], J. M. McMahon[2], A. Schneider[2], C. Myers[1], R. S. Moller[3], I. E. Scheffer[2], H. C. Mefford[1].* 1) Department of Pediatrics, Division of Genetic Medicine, University of Washington, Seattle, WA 98195; 2) Epilepsy Research Center and Department of Medicine, University of Melbourne, Austin Health, Australia; 3) Danish Epilepsy Centre, Dianaland, Denmark.

Epilepsy is one of the most common neurological disorders, with a lifetime incidence of 3%. The epileptic encephalopathies are the most severe of all the epilepsies. Patients typically present with refractory epilepsy with multiple seizure types, cognitive arrest or regression, and have a poor prognosis. *De novo* mutations have been increasingly recognized as causative for these disorders, and mutations in over 30 genes have been described. Recently, *de novo* mutations in *EEF1A2* were described in patients with intellectual disability and epilepsy. To define the role of this gene in epilepsy more broadly, we performed targeted resequencing of *EEF1A2* in 683 patients with epileptic encephalopathy. Overall, on average we sequenced 82% of the gene at a coverage of >50X across all samples. We identified four patients with likely pathogenic mutations, three patients had a recurrent Asp91Asn mutation and a fourth had a *de novo* Ala125Glu mutation. We review the phenotypes of these patients and contrast with those previously described. *EEF1A2* encodes for the alpha subunit of the eukaryotic translation elongation factor that controls the delivery of tRNAs to the ribosome. The function of this protein reveals a novel biological mechanism for epileptogenesis and broadens our appreciation for the complex mechanisms that underpin epilepsy. Future functional studies will not only enhance our understanding of epileptogenesis, but also potentially illuminate new options for therapeutic developments.

2882W

A missense variant in KRT25 causes autosomal recessive woolly hair. *M. Ansar[1,2], S. I. Raza[1], K. Lee[2], S. Shahi[1], A. Acharya[2], H. Dai[2], J. D. Smith[3], J. Shendure[3], M. J. Bamshad[3], D. A. Nickerson[3], R. L. Santos-Cortez[2], W. Ahmad[1], S. M. Leal[2], University of Washington Center for Mendelian Genomics.* 1) Department of Biochemistry, Faculty of Biological Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan; 2) Center for Statistical Genetics, Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA; 3) Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA.

A novel gene *KRT25* that encodes Keratin, type I cytoskeletal 25 (K25) protein, was identified for woolly hair (WH), which is primarily characterized by tightly curled hair with abnormal growth. We analyzed two consanguineous Pakistani families with non-syndromic autosomal recessive (AR) WH. Seven individuals, from these families present with tightly curled brittle scalp hair since early childhood. In both families homozygosity mapping and linkage analysis identified overlapping regions within 17q21. 1-q22. Exome sequencing of a DNA sample from each family result in the identification of a homozygous missense variant c. 950T>C (p. (Leu317Pro)) within *KRT25*. The variant segregates with ARWH in both families and has a combined maximum two-point LOD score of 7. 9 at λ=0. The KRT25 p. (Leu317Pro) variant is predicted to be deleterious by multiple bioinformatics tools. Molecular modeling of K25 protein with the variant predicted disruption of the second α-helical rod domain, which probably perturbs the helix structure resulting in the loss of surface contact residues. These perturbations possibly interfere with heterodimerization of K25 with type II keratins within the three inner root sheath layers of the hair follicle and the medulla of the hair shaft. In the hair follicle, KRT25 is expressed in Henle layer, Huxley layer and in the inner root sheath cuticle. Additionally *Krt25-/- mice* have fragile and curled fur. Furthermore micewith*Krt25*mutationshave mislocalized K71 protein as well as irregularlayers within the inner root sheath. In humans variants in *KRT71* have previously been shown to underlie non-syndromic WH. Our findings implicate a novel gene involved in human hair abnormality and are consistent with the curled, fragile hair found in *Krt25-/-* mice, and further support the role of inner root sheath-specific type I keratins in hair follicle development and maintenance of hair texture.

2883T

De Novo Variants in *GABBR2* Associated with Developmental Delay, Hypotonia, and Cyanosis. *N. J. Boczek[1], M. A. Cousin[1], P. R. Blackburn[1], A. N. Sigafoos[1,2], R. A. Urrutia[2,3], R. H. Gavrilova[4], K. J. Clark[1,2], E. W. Klee[1,5].* 1) Center for Individualized Medicine, Mayo Clinic, Rochester, MN; 2) Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, MN; 3) Department of Biophysics, Mayo Clinic, Rochester, MN; 4) Departments of Medical Genetics and Neurology, Mayo Clinic, Rochester MN; 5) Department of Biomedical Informatics, Mayo Clinic, Rochester, MN.

Whole exome sequencing (WES) is a sequencing strategy utilized in diagnostic odyssey cases in attempt to identify the underlying genetic substrate responsible for the patient phenotype. Recent publications have reported that WES reveals the genetic cause in ~25% of these cases, and is most successful when applied to children with neurological disease. The residual ~75% of cases remain genetically elusive until more information becomes available in the literature or applicable functional studies are pursued. Our diagnostic odyssey case presented at 6 months of age with severe developmental delay and hypotonia, with a normal brain MRI. At five years of age the patient began having spells of altered breathing rate with cyanosis with no documented seizure activity. In addition, she has repetitive hand movements and intermittent posturing of her legs. The patient is currently 7 years of age and unable to stand, walk unassisted, communicate, or follow instructions. Clinical WES testing by Baylor College of Medicine did not identify a genetic variant with clear causal association with this patient's phenotype. Subsequent research WES was performed at Mayo Clinic on the proband and her unaffected parents. Utilizing an in house variant prioritization tool, a de novo heterozygous variant within *GABBR2* was identified as the top candidate variant. *GABBR2*, encoding the G-protein coupled receptor, gamma-aminobutyric acid beta receptor 2, is known for its regulation of neurotransmitter release in the brain. The de novo variant, p. M702V, falls is within the highly conserved 6th transmembrane spanning region of GABBR2 and is absent in the publically available exome/genome databases including NHLBI ESP, ExAC, and the 1000 Genomes Project. A recent manuscript examining epileptic encephalopathy identified de novo GABBR2 variants, p. S695I and p. I705N, in the 6th transmembrane spanning domain of GABBR2 in two cases with almost identical phenotypes to our proband. The identification of de novo mutations in a neurologically expressed gene in our patient as well as two other individuals with comparable phenotypes, suggests the *GABBR2* variants within the 6th transmembrane spanning domain may be responsible for phenotypes consistent with developmental delay, hypotonia, and cyanosis. We are currently pursuing functional studies within zebrafish to confirm these findings.

2884F

Compound heterozygous loss of function variants in the signaling transducer *GNB5* in two sisters with severe hypotonia and hyporeflexia, cognitive deficit and epilepsy. *P. De Nittis[1,2], L. Gueneau[2], D. Cocciadiferro[1], A. A. Alfaiz[2], V. Napolioni[3], B. Augello[1], M. T. Pellico[1], L. Zelante[1], A. Reymond[2], G. Merla[1].* 1) Medical Genetics Unit, IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy; 2) Center for Integrative Genomics, University of Lausanne CH-1015 Lausanne; 3) Department of Neurology and Neurological Sciences, Stanford University School of Medicine, Stanford, CA, USA.

We identified two Italian sisters affected by the same multisystem disorder, defined by a global developmental impairment, psychomotor disability, delayed speech development, learning deficiency, nystagmus, epileptic seizures, generalized muscular weakness, bradycardia, hyporeflexia and characteristic facial features. Exome sequencing of the two affected siblings and their non-consanguineous healthy parents uncovered bi-allelic loss-of-function variants in the *GNB5* gene in both siblings. *GNB5* encodes the fifth member of the heterotrimeric G-protein β-subunit family, which integrate signals between receptors and effector proteins. It is a 353-aminoacids protein, principally expressed in brain and tissues such as pancreas, heart and kidney. GNB5 differentially governs the proteinstability of G protein signaling regulators (RGS), and is thus involved in the control of complex neuronal G protein signaling pathways. Further, RGS6/GNB5 complex modulates parasympathetic regulation ofheartrate in mice. The paternal missense variant r. [249G>A;249-250ins25] alters splicing accuracy, as assessed by RT-PCR and subsequent sequencing of the amplified products. The retention of 25bp of the intervening intron results in a frame-shift that shortens the encoded amino acid sequence. The maternal allele variant c. 994C>T results in a stop gain. Both variants appear to target the corresponding mRNA transcripts to nonsense-mediated mRNA-decay. Consistent with the possible involvement of these variants in the phenotype of the two sisters, knock-out of the paralogous gene in mice results in marked neurobehavioral developmental delay, impaired gait, motor learning and coordination deficiencies. *Gnb5*-deficient mouse models also display defective visual adaptation and abnormal functioning of retinal bipolar cells. Further validation of these findings through identification of more patients with variant in *GNB5* is warranted.

2885W

Identification of New Genes and Pathways for Rare Infantile Forms of Myopathies and Neuromuscular Disorders. *J. M. Hunter[1], C. Balak[1], M. E. Ahearn[1], W. Liang[1], M. Russell[1], M. Huentelman[1], D. Craig[1], J. Carpten[1], S. M. Bernes[2], L. Baumbach-Reardon[1].* 1) Integrated Cancer Genomics, TGen, Phoenix, AZ; 2) Phoenix Children's Hospital, Phoenix, AZ.

Neuromuscular disorders and myopathies account for a significant proportion of infant and childhood mortality and chronic disease. The phenotypes and genetics of these disorders cover a broad spectrum but have many overlapping features. As we learn more about the etiology of human disease we have found that most severe disease causing variants are very rare, only occurring in a single or a few individuals. Developing treatments for these individual diseases begins with identification of the underlying genetic mutations. Our goal is to use next generation sequencing to identify mutations that cause neuromuscular disorders and muscular dystrophies. In many of the cases we have studied, prior genetic testing and histology, which was often very extensive and spanned many years of evaluation, did not reveal the genetic etiology of disease. Over the past 2 years we have completed whole exome sequencing on 172 individuals including 64 affected individuals and family members. From this data set, we have identified novel probable pathogenic mutations in genes known to cause disease as well as in genes not previously associated with disease in ~30% of cases. We highlight some of our results and will summarize all of our findings. First, we identified novel compound heterozygous *PLA2G6* pathogenic variants in a male infant with regression of acquired motor skills, progressive cerebellar atrophy, and generalized hypotonia. Second, we studied a male child with a form of limb girdle muscular dystrophy, generalized weakness and dystrophic muscle biopsy results. We identified novel micro deletion and frameshift compound heterozygous mutations in *COL6A2* for a genetic diagnosis of Bethlem myopathy. Third, we describe a severe case of *possible* Nemaline myopathy caused by compound heterozygous splice mutations in *NEB*. The child presented with congenital global hypotonia, arthrogryposis, deformed diaphragm, and was ventilator dependent. Finally we present a child with a severe congenital neuromuscular disease presenting with global severe hypotonia, contractures, severe myofiber disorganization, and fractures. We identified a novel frameshift and premature truncation in TTN which we suspect is the cause of disease. The diversity of our results demonstrates the necessity of using next-generation sequencing to find answers for families with these rare and devastating diseases.

2886T

A comprehensive genomic analysis reveals the genetic landscape of mitochondrial respiratory chain complex deficiencies. *M. Kohda[1], Y. Tokuzawa[1], Y. Kishita[1], Y. Moriyama[1,2], Y. Mizuno[1], T. Hirata[1], Y. Yatsuka[1], Y. Yamashita-Sugahara[1], Y. Nakachi[1], H. Kato[1], S. Tamaru[3], H. Nyuzuki[1], N. N. Borna[1], H. Harashima[4], T. Yamazaki[4], M. Mori[5], K. Murayama[6], A. Ohtake[4], Y. Okazaki[1].* 1) Research Center for Genomic Medicine, Saitama Medical University, Hidaka, Japan; 2) Department of Anatomy II and Cell Biology, Fujita Health University School of Medicine, Toyoake, Japan; 3) Department of Obstetrics and Gynecology, Saitama Medical University, Saitama, Moroyama, Japan; 4) Department of Pediatrics, Saitama Medical University, Moroyama, Japan; 5) Department of Pediatrics, Matsudo City Hospital, Matsudo, Japan; 6) Department of Metabolism, Chiba Children's Hospital, Chiba, Japan.

We present the first report of a comprehensive genomic analysis of a cohort of patients with mitochondrial respiratory chain complex deficiencies. Using high-throughput sequencing and chromosomal aberration analysis, we ultimately identified 31 novel mutations in genes previously reported to be linked to mitochondrial disorders, and novel mitochondria-related genes (*NDUFB11*, *COQ4*, *GTPBP3*, and etc) as novel causative genes. We also identified noncanonical mitochondria-related disease genes (*MECP2* and *TNNI3*) and 3 chromosomal aberrations (6q24. 3-q25. 1, 22q11. 21, and 17p12) in patients with mitochondrial respiratory chain complex deficiencies. Using comprehensive genomic analyses, we achieved firm genetic diagnoses in 49 of 142 patients (34. 5%). Comprehensive genomic analyses also enabled us to identify closely related disorders, which will be important for patient care and the application of an efficient drug repositioning. We demonstrated that extensive clinical heterogeneity exists in this disorder, and comprehensive genomic analyses could reveal its clinical and genetic heterogeneity for better understanding this complex disorder.

2887F

Genetic analysis of autosomal recessive primary microcephaly in Indian population. *A. Kumar[1], S. C. Girimaji[2].* 1) Molecular Reproduction, Development & Genetics, Indian Institute of Science, Bangalore 560012, Karnataka, India; 2) Department of Child and Adolescent Psychiatry, National Institute of Mental Health and Neuro Sciences, Bangalore 560029, Karnataka, India.

Human brain development is a complex process that involves generation of neural progenitors in periventricular zones, cell proliferation through symmetric and asymmetric cell divisions, and finally the migration of post-mitotic neurons to their final destinations. An understanding of the molecular mechanisms guiding the complex process is facilitated by discovery of causative genes for a cortical (brain) malformation disorder, autosomal recessive primary microcephaly (MCPH), which is characterized by a smaller than normal brain and intellectual disability. Interestingly, the brain volume (500 cc) of an MCPH patient is similar to early hominids, such as Australopithecus who lived 1. 8-4. 4 myr ago. Occasionally, MCPH patients also have other brain malformations, such as pachygyria, polymicrogyria, simplified gyral pattern etc. We have been working on the genetics of MCPH for some time, which led to discovery of the seventh MCPH gene *STIL* in 2009. So far, a total of 13 genes have been identified for MCPH: *MCPH1* on ch8p23. 1, *WDR62* on ch19q13. 12, *CDK5RAP2* on ch9q33. 2, *CASC5* on ch15q15. 1, *CEP152* on ch15q21. 1, *ASPM* on ch1q31. 3, *CENPJ* on ch13q12. 12*, STIL* on ch1p33*, CEP63* on ch3q22. 2, *CEP135* on ch4q12, *PHC1* on ch12p13. 31, *CDK6* on ch7q21. 2 and *SASS6* on ch1p21. 2**.** We have ascertained a total of 39 MCPH families. Using homozygosity mapping and Sanger sequencing, we have identified mutations in *WDR62* (3 families), *ASPM* (8 families) and *STIL* (6 families) genes only in our families, with a mutation detection rate of 43. 59% (17/39 families). Thus, the *ASPM* gene is found to be the most commonly mutated gene followed by *STIL* and *WDR62* in Indian population. There are still 22/39 families in our family dataset which do not harbor mutations in any of the known MCPH genes, suggesting the future discovery of additional novel gene(s). We are currently using homozygosity mapping and whole exome sequencing to identify the novel MCPH gene(s). A comprehensive analysis of our molecular genetic analysis of Indian MCPH families will be presented and discussed.

2888W

Description of a male with Microphthalmia Syndromic 2 and a rare *BCOR* missense variant. *A. L. Pilla[1], A. B. Alvarez Perez[1], M. P. Migliavacca[1], J. Jurgens[2], L. Brunoni[1], N. L. M. Sobreira[2].* 1) Clinical Genetics Dept. of Morphology and Genetics, Universidade Federal de São Paolo, São Paolo, SP, Brazil; 2) Institute of Genetic Medicine, Johns Hopkins University School of Medicine, 733 N. Broadway Street Suite 512, Baltimore, MD 21205, USA.

Ramos-Arroyo syndrome (MIM122430) was described as a new syndrome by Ramos-Arroyo in 1987 in two brothers with corneal anesthesia, bilateral sensorineural deafness, patent ductus arteriosus, moderate intellectual disability and dysmorphic facial features. In 2008, Spurrier and Weaver described the same family including two other affected individuals and additional findings such as Hirschsprung's disease. The presumed mode of inheritance is autosomal dominant, and to date the molecular basis is unknown. In 2004, a variant in the *BCOR* gene (BCL6 corepressor) was identified in a family with Microphthalmia Syndromic 2 (MIM300166), characterized by X-linked dominant inheritance and eye abnormalities including microcornea, congenital cataracts and microphthalmia. These affected individuals also had short stature, dysmorphic facial features, patent ductus arteriosus and intellectual disability. Here we describe an individual with a clinical diagnosis of Ramos-Arroyo syndrome referred to the Baylor-Hopkins Center for Mendelian Genomics for whole exome sequencing (WES). He is the 3rd son of non-consanguineous parents. His mother used misoprostol in the 1st trimester of pregnancy, followed by a minor hemorrhage episode. He was born by uneventful normal delivery at term. His birth weight was 3500 g and height was 47 cm. His mother reported hypotonia and little eye movement since the neonatal period. Later ophthalmologic examination diagnosed corneal anesthesia and alacrimia. An eye surgery was performed at 7 months for suture of the eyelids. He had delayed psychomotor development. His physical examination at 7 years 8 months showed weight: 17 kg (<p3), height: 1. 18 cm (p25), PC: 52 cm (p50), prominent front, hypertrichosis and flattening in the face, anteverted nostrils, and hypoplasia of the middle face. WES identified a rare missense variant (c. 2288G>T, p. Arg763Leu) in exon 4 of *BCOR*, located on chromosome Xp11. 4. To date, many affected females have been described with truncating variants in *BCOR* and Microphthalmia Syndromic 2, also known as oculocardiofaciodental syndrome. Affected males have been described in only 3 families, all with the p. Pro85Leu missense variant in *BCOR*. Based on our patient's phenotypic similarity to Microphthalmia Syndromic 2, we suggest that his p. Arg763Leu *BCOR* variant is pathogenic and that he has Microphthalmia Syndromic 2. We also suggest that *BCOR* should be tested in individuals with the clinical diagnosis of Ramos-Arroyo syndrome.

2889T

Whole exome sequencing analysis of intracranial aneurysm in multiplex families. *A. Powell[1], B. Fernandez[1], F. Maroun[2], K. Mahoney[1], M. Woods[1].* 1) Discipline of Genetics, Faculty of Medicine, Memorial University, St. John's, NL, Canada; 2) Department of Surgery, Faculty of Medicine, Memorial University, St. John's, NL, Canada.

An intracranial aneurysm (IA) is a dilatation of the wall of a cerebral artery. The weakened vessel walls of IAs have the potential to rupture and lead to subarachnoid hemorrhage, a devastating type of stroke. In addition to environmental factors, numerous studies have supported genetic predisposition as a risk factor for IA development and rupture. However, no causal genes have been identified. The purpose of this study is to identify genetic variants that cause susceptibility to familial IA in the Canadian province of Newfoundland and Labrador. This involved the recruitment of 53 families that have a history of IA, and an additional 33 sporadic IA cases. Currently, we are focusing on two multiplex families, each with greater than ten affected individuals, and pedigrees consistent with Mendelian inheritance. A pilot study of twelve individuals was selected from this cohort, for whole exome sequencing of genomic DNA. To identify disease-causing genes, we developed a filtering strategy for variant prioritization. We focused on variants with potentially high or moderate effect (non-synonymous SNVs, frameshifts, splice-site acceptors and donors) that were shared by multiple affected individuals within a single family. Common variants with a minor allele frequency greater than 1% in public databases and a population of control exomes were excluded. Bioinformatics tools such as SIFT and Polyphen2 were also used to predict pathogenicity, along with data concerning the functional relevance of affected proteins. Our comprehensive method of filtering resulted in an initial list of 67 variants in 58 genes from the 12 exomes. This was further reduced to 20 variants of interest in 16 genes, as greater emphasis was placed on novel variants shared by all exomes within a family. Sanger sequencing was used for the validation of variants and to test for segregation in additional family members. After filtering, there were no rare variants shared by the two families. As well, these families did not share any different variants in the same gene. The identification of rare and novel variants in familial cases of IA is a significant step toward furthering our understanding of the genetic etiology of this disease. Variants that were identified through this study may have significance in other populations. Greater knowledge at the molecular level could assist in the future diagnosis and treatment of asymptomatic unruptured IA, and aid in genetic counseling of families.

2890F

Molecular genetic studies of the oculocutaneous albinism phenotype in the Pakistani population. *M. Shahzad[1], S. YOUSAF[1,2], T. KAUSAR[2], Y. M. WARYAH[3], H. GUL[4], U. MEHMOOD[1], N. TARIQ[2], M. A. KHAN[4], M. ALI[5], M. A. KHAN[4,6], A. M. WARYAH[3], S. RIAZUDDIN[1], R. S. SHAIKH[2], Z. M. AHMED[1,2], University of Washington Center for Mendelian Genomics.* 1) Department of Otorhinolaryngology Head & Neck Surg, School of Medicine, University of Maryland, BALTIMORE, MD; 2) Institute of Molecular Biology and Biotechnology, Bahauddin Zakariya University, Multan, Pakistan; 3) Molecular Biology & Genetics Department, Medical Research Center, Liaquat University of Medical & Health Sciences, Jamshoro, Pakistan; 4) Gomal Centre of Biochemistry and Biotechnology, Gomal University, D. I. Khan, Pakistan; 5) Government College University, Faisalabad, Pakistan; 6) Genomic Core facility, Interim Translational Research Institute, Academic Health System, Hamad Medical Corporation, Doha, Qatar.

Oculocutaneous albinism (OCA) is a rare genetic disorder characterized by variable hypopigmentation in hair, skin, and eyes with or without other clinical manifestation. Currently, mutations in at least 18 genes have been causally linked to OCA phenotype in human. Our objective here is comprehensive genetic screening of known OCA genes in Pakistani families segregating various types of OCA phenotype. We ascertained 132 families with multiple affected individuals from different provinces of Pakistan, after obtaining IRB approvals and informed written consents. Sanger sequencing of common *OCA* genes followed by whole exome sequencing was performed for the identification of pathogenic variants in these families. We identified 42 new and 18 reported variants, in nine known *OCA* genes responsible for OCA phenotype in 121 families. These mutations include, thirty-six missense, seven splice site, six nonsense, four frame shift, one six base in-frame and six gross deletions. In our cohort, mutations in *TYR* (OCA1) are more prevalent (40. 91%) followed by *OCA2* alleles (34. 09%), *TYRP1* (OCA3; 3. 79%), *SLC45A2* (OCA4; 3. 79%), *SLC24A5* (OCA6; 2. 27%) and *HPS1* (2. 27%). We also identified single variants in *HPS3*, *HPS4*, *HPS6* and *HPS9* and a compound heterozygous mutation in *LYST*. Our findings indicate that p. (Arg278*) and p. (Gly419Arg) alleles of *TYR* and c. 1045-15T>G and p. (Asp486Tyr) of *OCA2* are recurrent mutations in Pakistani population. Therefore, we developed tetra primer screening assays for the detection of these highly prevalent recurrent mutations. In remaining 11 families, we did not identify mutations in known OCA genes, indicating the existence of yet unidentified *OCA* gene(s). To the best of our knowledge, this study represents the most comprehensive documentation of *OCA* alleles in the Pakistani population. Overall, our study contributes to the development of genetic testing protocols, future diagnosis, genetic counseling and molecular epidemiology of OCA in Pakistan.

2891W

Systematic evaluation of patients with idiopathic short stature using whole exome sequencing. *C. T. Thiel[1], N. N. Hauer[1], S. Schuhmann[1], E. Schoeller[1], M. T. Wittmann[1], S. Uebe[1], A. B. Ekici[1], H. Sticht[2], H. -G. Doerr[3], A. Reis[1].* 1) Institute of Human Genetics, Friedrich-Alexander-Universitét Erlangen-Nürnberg, Erlangen, Germany; 2) Institute of Biochemistry, Friedrich-Alexander-Universitét Erlangen-Nürnberg, Erlangen, Germany; 3) Department of Pediatrics and Adolescent Medicine, Friedrich-Alexander-Universitét Erlangen-Nürnberg, Erlangen, Germany.

Shortness of stature is a common medical concern in childhood and has an incidence of 3% in the general population. After excluding defects of the growth hormone pathway and recognizable syndromes the underlying cause remains unknown in approximately 70-80% of patients. In some of these patients the underlying diagnosis is omitted by the lack of clinical features characteristic for known syndromic forms of short stature. To address this in patients with idiopathic short stature we thoroughly built a study group of more than 500 families with idiopathic short stature. We systematically selected 100 individuals where growth hormone defects, common genetic causes of short stature or copy number variations were excluded and performed whole exome sequencing. Variants were selected unbiased based on all modes of inheritance in agreement with the segregation in the families and their potential effect on protein function using our NGS Variant Analyzer software. We confirmed mutations in known short stature genes in 11 patients. All these syndromes have been reported to be associated with further clinical issues providing mandatory medical guidance for these patients. In addition, we recognized novel candidate genes involved in epigenetic modification, cell cycle regulation, ubiquitination and protein synthesis in up to 54 % of the patients with idiopathic short stature. Even though, recessive, dominant and x-linked inherited variants were observed in these candidate genes, we confirmed autosomal dominant de novo as the predominant inheritance model in idiopathic short stature. Thus, our data underlines the rare variant – frequent disease hypothesis for the extreme end of the growth spectrum. In conclusion, whole exome sequencing identified the underlying genetic defect in up to 65% of the patients with idiopathic short stature. As the clinical spectrum of most genetic defects is yet to be explored, an unbiased genetic analysis of patients with idiopathic short stature can establish a diagnosis in these cases.

2892T

Expansion of the Troyer Syndrome phenotype by discovery of mutations in Turkish families. *H. Unal Gulsuner[1], O. Ceyhan-Birsoy[2,3], S. Gulsuner[1], O. E. Onat[4], T. Walsh[1], H. Shahin[5], M. K. Lee[1], A. B. Tekinay[6], M. -C. King[1], A. H. Beggs[7], T. Ozcelik[4,6], H. Topaloglu[8].* 1) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA; 2) Laboratory for Molecular Medicine, Partners HealthCare, Cambridge, MA, USA; 3) Department of Pathology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA; 4) Department of Molecular Biology and Genetics, Bilkent University, Ankara, Turkey; 5) Department of Life Sciences, Bethlehem University, Bethlehem, Palestine; 6) Institute of Materials Science and Nanotechnology, National Nanotechnology Research Center (UNAM), Bilkent University, Ankara, Turkey; 7) Manton Center for Orphan Disease Research, Department of Genetics and Genomics, Boston Children's Hospital and Harvard Medical School, Boston, MA, USA; 8) Department of Pediatric Neurology, Faculty of Medicine, Hacettepe University, Ankara, Turkey.

We report two unrelated Turkish families, both consanguineous, with two and one affected children, respectively, with progressive spasticity and weakness of the legs, spastic gait, developmental delay, and intellectual disability. Parents and all other family members were unaffected. Clinical evaluation suggested that all three children suffered from the same disorder, but did not establish a differential diagnosis. We exome sequenced genomic DNA from the three affected individuals. In the first family, both affected children were homozygous for SPG20 p. G580R (c. 1780G>A); in the second family, the affected child was homozygous for SPG20 p. Q229X (c. 685C>T). No other gene harbored homozygous damaging alleles in all three affected children. Neither mutation appears in any public database. SPG20 p. G580R lies at a residue completely conserved in all sequenced species, in a highly conserved domain. *SPG20* (spastic paraplegia 20) encodes spartin, a 666-residue protein that localizes to the mitochondria and is involved in endosomal trafficking. *SPG20* is widely expressed in both developing and adult brain. *Spg20-/-* mice have been reported to show significant defects in gait, in axon branching of cerebral cortical neurons, and in BMP signaling. Homozygosity for loss-of-function mutations in *SPG20* has previously been reported to cause Troyer syndrome, an autosomal recessive complicated spastic paraplegia, in the Old Order Amish and in an extended Omani kindred. Additional clinical evaluation and magnetic resonance imaging (MRI) of our patients revealed neurological symptoms similar to those reported for the Amish and Omani Troyer syndrome patients, and also several important differences. Both affected children in the first family in our study have early onset tremor, which is not reported in any other Troyer syndrome cases; and all three children in our study lack any skeletal abnormalities, which are reported to be present in all other Troyer syndrome cases. In conclusion, SPG20 mutations in two Turkish families are likely responsible for a Troyer syndrome phenotype that shares neurological and developmental features of previously reported cases, but including, in one family, early onset tremor, and without skeletal abnormalities.

2893F

Genetic characterization of childhood-onset cardiomyopathies in the Finnish population. *C. Vasilescu[1], C. Carroll[1], V. Brilhante[1], A. Hiippala[2], J. Lohi[3], E. Jokinen[2], T. Tyni[4], T. Ojala[2], A. Suomalainen[1,5].* 1) Research Programs Unit, Molecular Neurology, Biomedicum Helsinki, University of Helsinki, Helsinki, Finland; 2) Department of Pediatric Cardiology, Children's Hospital, Helsinki University Central Hospital, Helsinki, Finland; 3) Department of Pathology, Helsinki University Central Hospital, Helsinki, Finland; 4) Department of Pediatric Neurology, Helsinki University Central Hospital, Helsinki, Finland; 5) Department of Neurology, Helsinki University Central Hospital, Helsinki, Finland.

Childhood-onset cardiomyopathies (CMPs) are severe heart disorders, mainly of genetic origin, with autosomal dominant, *de novo*,recessiveormaternal inheritance. In this study we aim to characterize the collection of disease genes and mutations causing childhood-onset cardiomyopathies in the Finnish population. We used Next Generation Sequencing (NGS) to screen a group of 89 pediatric patients from the sole clinic in Finland treating severe childhood-onset CMPs. These are patients in whom the genetic diagnosis was not achieved with standard candidate gene approaches. Nine of the patients were screened by whole-exome sequencing (WES), 37 by a HaloPlex custom-designed panel of 117 CMP-related genes, and 43 with a commercial panel of 101 genes of nonsyndromic and syndromic forms of CMP (BluePrint Genetics). Variants were prioritized with respect to frequency and pathogenicity prediction. Candidate mutations were further verified to match the disease segregation in the families and to be absent in Finnish controls under the implicated mode of inheritance. Dominant mutations were confirmed in nine patients, affecting proteins involved in sarcomere structure and function (*MYH7* - 3 distinct mutations, *TNNC1*, *TNNI3*), calcium handling (*CALM1*, *CACNA1C*), autophagy/apoptosis (*BAG3*), and cellular energy sensing (*PRKAG2*). Among these, screening of family members revealed *de novo* mutations in six cases. Furthermore, a recessive mutation was found in JPH2, which encodes a protein with both structural and calcium regulation roles. Additional candidates in *TAZ*, *PTPN11*, *MAP2K2*, *JPH2*, *MYH7*, *ACTC1*, *TBX20*, and *NEXN* are under investigation. Overall, we estimate a diagnostic success rate of around 25%. The mutations in our cohort are private for families and most of them novel. *De novo* mutations were found to be a common cause of early-onset cardiomyopathy and recessive mutations explain only a small fraction of cases, even in a genetic isolate such as Finland.

2894W

Genetic and Functional studies of *MC1R* alleles associated with hypopigmentation phenotype in human. *S. Yousaf[1,2], M. Shahzad[1], J. Campos[3], N. Tariq[2], C. Serrano[3], R. Yousaf[1], C. Cervantes[3], Y. Waryah[4], T. Kausar[2], M. Ali[5], A. Waryah[4], S. Riazuddin[1], S. Shaikh[2], J. Garcia-Borron[3], Z. Ahmed[1,2,5].* 1) Department of Otorhinolaryngology Head and Neck Surgery, School of Medicine, University of Maryland, Baltimore, MD, USA; 2) Institute of Molecular Biology & Biotechnology, Bahauddin Zakariya University, Multan, Pakistan; 3) 3Department of Biochemistry and Molecular Biology, School of Medicine, University of Murcia and IMIB-Arrixaca, Murcia, Spain; 4) Molecular Biology & Genetics Department, Medical Research Center, Liaquat University of Medical & Health Sciences, Jamshoro, Pakistan; 5) Government College University, Faisalabad, Pakistan.

Background: Skin pigmentation is one of the most visible phenotypes in humans. Loss of pigmentation is a group of disorders caused by either complete absence or deficiency of melanin in skin, hair and/or iris. Our goal here is to identify and characterize the *MC1R* alleles associated with hypopigmentation disorders in the Pakistani population. We enrolled consanguineous Pakistani families segregating hypopigmentation phenotype with normal fundus. **Methods:** Sanger sequencing was performed for the known genes followed by linkage analysis and Whole-Exome enrichment and sequencing (NGS). *Ex-vivo* studies were carried out in transiently transfected cells with cDNAs encoding either the wild type or mutant proteins. **Results:** Bioinformatic analysis of NGS data revealed four missense, one in-frame deletion (p. (Val174del) and a nonsense mutation (p. (Tyr298*)) in *MC1R* co-segregating with the hypopigmentation phenotype in seven families. *MC1R* encodes a seven-transmembrane G-protein coupled receptor mainly present on the membrane of melanocytes. *Ex-vivo* studies revealed destabilization of *MC1R* due to p. Val174del and p. Tyr298* mutations. The p. Tyr298* mutant *MC1R* showed no agonist-induced signaling to the cAMP or ERK pathways, nor detectable agonist binding. Flow cytometry and confocal imaging studies revealed reduced plasma membrane expression of p. Val174del and p. Tyr298*. **Conclusion:** Our data provide further insight on the relationship of MC1R functional status with hypopigmentation. Our data support the current view that functional characterization of new variants might allow for a prediction of phenotypic effects for low frequency alleles that cannot be reliably analyzed by genetic epidemiology association studies, or for mutations not easy to analyze *in silico*. Our results also reveal the complexity of determinants of normal MC1R trafficking to the plasma membrane.

2895T

Compound heterozygous mutations leading to loss of JAK2 function cause an immune dysregulatory syndrome. *X. Yu[1], M. Chi[1], S. Pitaluga[2], S. Siegel[3], J. Niemela[4], J. Milner[1].* 1) Genetics and Pathogenesis of Allergy Section, NIAID, Bethesda, USA; 2) Laboratory of Pathology, NCI, Bethesda, USA; 3) New York Medical College, Westchester Medical Center, Valhalla, USA; 4) Department of Laboratory Medicine, Clinical Center, Bethesda, USA.

We have followed an infant who developed a shoulder abscess at one month of age followed by severe ichthyosis, hepatosplenomegaly, lymphadenopathy, histiocytic inflammation, and consumptive anemia. Laboratory analysis showed decreased NK and CD8 killing, panhypogammaglobulinemia, extraordinary polyclonal T-cell lymphocytosis, normal B-cell numbers without any class-switched B-cells, and a lack of B-cell follicles. Almost no PD-1 CXCR5+ circulating T-follicular helper (TFH)-like cells were found in blood, and staining for PD-1+ or BCL-6+ TFH cells in the lymph node was completely negative. Without a diagnosis at the time, the patient received systemic immune suppression and a matched unrelated bone marrow transplantation and is healthy now. Whole-exome sequencing was performed in the affected patient and his unaffected parents. Common variants (MAF > 0. 01) were filtered from exome data. We identified compound heterozygous mutations Ser46Tyr (MAF < . 001) and Leu393Val (MAF . 01) in *JAK2*, the gene encoding Janus kinase 2 (JAK2). Computer modeling based on the crystal structure of human tyrosine kinase 2 predicts that two mutations disrupt FERM-SH2 and receptor-binding domains of JAK2. While gain of function JAK2 mutations are associated with a number of proliferative disorders and blood dyscrasias, western blots for JAK2 showed a decrease in protein expression in PBMCs from the patient, and Stat3 and Stat1 activation after stimulation with JAK2 dependent cytokines IL-6 and IL-27 was low, despite normal pSTAT3 responses to non-JAK2 dependent cytokines. The patient had profoundly decreased peripheral T-cell IFN-γ production, consistent with a focal defect in T-helper 1 T-cell differentiation, which is highly dependent upon IL-12 and IFNg signaling, both of which require JAK2 for signaling. IL-2, Th2, and Th17 cytokine expression, was normal, consistent with their ability to be generated by non-JAK2 dependent cytokines. These data indicate that loss of function *JAK2* mutations can lead to a profound immune deficient and dysregulatory state. While lethality of JAK2 loss of function in mice has precluded specific examination of Thelper cell differentiation, JAK2 appears to be required for normal Th1 and TFH differentiation, but not IL-2 prodcution, Th2 or Th17 differentiation in humans. Study of the extrahematopoetic symptoms in this transplanted patient over time will further add to our understanding of the consequences of impaired JAK2 signaling in humans.

2896F

Novel and previously reported mutations in MSX1 gene in three Mexican families with non-syndromic dental agenesis. *E. A. Ramirez-Ramirez[1,2], KL. Valdes-Morales[2], F. Hernandez-Teran[1,3], LD. Mayoral-Amador[3], HM. Diaz-Martinez[3], FF. Gonzalez-Galarza[2], HM. Arredondo-Carrera[2], JE. Gaytan-Arocha[2], RD. Arellano-PerezVertti[2], D. Delgadillo-Gozman[2], AY. Garcia-Marin[1], FO. Gonzalez-Galarza[1], JR. Arguello[1,2], RI. De la Cruz-Granados[2].* 1) Bioinformatics, Instituto de Ciencia y Medicina Genómica, Torreon, Coahuila, Mexico; 2) Facultad de Medicina. Universidad Autónoma de Coahuila. Torreón, Coahuila, México; 3) Facultad de Odontología. Universidad Autónoma de Coahuila. Torreón, Coahuila, México.

IntroductionCongenital dental agenesis consists in the lack of one or multiple permanent teeth in the oral cavity, as a result of molecular anomalies during the first stages of tooth development. Non-syndromic familial hypodontia is usually an autosomal dominant disease, which features incomplete penetrance and variable expression. Mutations related with this disease may present a hypodontia or a saw-like incisors phenotype. Most common mutations occur in MSX1 and PAX9 genes. The aim of this study was to asses mutations on MSX1 gene in patients with non-syndromic hypodontia. **Methods**Non-syndromic dental agenesis of three Mexican families was confirmed with clinical and radiographic findings. For all patients, DNA was obtained through extraction of blood samples. End-point PCR was performed to amplify exons 1 and 2 of MSX1 gene and the 5'à3' byproducts were processed for automatic Sanger sequencing. All sequences were compared with the reference sequence for MSX1 gene published by the NCBI with BioEdit 7. 0 software. **Summary of results**A total of three individuals and both of their parents were included in the study. Family 1, both the subject and father presented hypodontia, while the mother had a saw-like incisors phenotype. The three of them shared the same two mutations: A SNV in position 1423 of 3'UTR region of MSX1 (rs12532) with an AG genotype, and a previously unreported SNV in position 1301 of 3'UTR of MSX1 with an AG genotype. Family2, subject and father had hypodontia. Both of them shared a CC genotype in rs36059701 on exon 1 and a CT genotype on rs8670 in the 3'UTR region. After analysis of both mutations on Polyphen-2, the mutation in rs36059701 is probably not the causal variant of the phenotype. In Family 3, the subject solely presented hypodontia , both parents were unaffected. However, both the subject and the mother presented a heterozygous mutation on rs1095 on 3'UTR region of MSX1. **Conclusion**We have demonstrated five different mutations in MSX1 gene. Four of them have been previously reported on the literature, and we have found a novel variant of unknown clinical significance, due to the fact that it is present along with a known variant, further studies will be required to prove its causative effect. On the other hand, we found an incomplete penetrance and a variable expression in two mutations, phenomena that have been previously described in the literature.

2897W

Achieving genetic diagnosis in families with enlarged vestibular aqueduct using massively parallel sequencing. *C. C. Wu[1,2], Y. H. Lin[1,3], Y. H. Lin[1], C. J. Hsu[1,4], P. L. Chen[3,5].* 1) Department of Otolaryngology, National Taiwan University Hospital, Taipei, Taiwan; 2) Department of Medical Genetics, National Taiwan University Hospital, Taipei, Taiwan; 3) Graduate Institute of Medical Genomics and Proteomics, National Taiwan University College of Medicine, Taipei, Taiwan; 4) Department of Otolaryngology, Taichung Tzu-Chi Hospital, Taichung, Taiwan; 5) Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan.

Enlarged vestibular aqueduct (EVA) is a common inner ear malformation associated with sensorineural hearing impairment in children. Although not all, the majority of EVA patients are linked to Pendred syndrome (PS) and non-syndromic DFNB4, two autosomal recessive disorders caused by mutations in *SLC26A4*. However, a significant percentage of EVA patients carry only 1 or 0 mutant *SLC26A4* allele, constituting a dilemma in genetic diagnosis and counseling. In the present study, we designed a massively parallel sequencing (MPS) panel targeting the whole length of three PS/DFNB4 genes (*SLC26A4*, *FOXI1* and *KCNJ10*) and all exons of ten other genes which have been related to EVA. We then compared the diagnostic yields between this one-step MPS panel and the convention genetic examination, i. e. combination of Sanger sequencing and MLPA, in a total of 157 EVA families with non-diagnostic *SLC26A4* genotypes. Using the MPS panel, we identified the second mutant *SLC26A4* allele in 19 (59%) of 32 families with 1 common *SLC26A4* mutant allele detected, including 18 with point mutations and 1 with a novel large 7666-bp deletion (g. -1066C_6602Adel). Meanwhile, the MPS panel confirmed genetic diagnosis in 6 (55%) of 11 families with 0 common *SLC26A4* mutant allele detected, including 3 families with bi-allelic *SLC26A4* mutations, 2 families with dominant *EYA1* mutations (p. R264Q and p. R328X), and 1 family with a dominant *SIX1* mutation (p. T120I). On the other hand, combination of Sanger sequencing and MLPA achieved genetic diagnosis in 52 (58%) of 90 families with 1 common *SLC26A4* mutant allele and 7 (17%) of 24 families with 0 common *SLC26A4* mutant allele, respectively. Our results demonstrated the utility of MPS in addressing EVA families by identifying various types of mutations with satisfactory diagnostic yields in one step.

2898T

Introduction of an effective gene-panel for DNA diagnosis of epidermolysis bullosa and skin fragility. *H. H. Lemmink[1], P. van den Akker[1], K. Abbott[1], M. Meems-Veldhuis[1], A. Scheper[1], M. Viel[1], M. Pasmooij[2], I. Turcan[2], K. Gostynska[2], M. Jonkman[2], R. Sinke[1].* 1) Genetics, University Medical Centre Groningen, Groningen, Netherlands; 2) Dermatology, University Medical Centre Groningen, Groningen, Netherlands.

AimThe number of candidate genes in patients suspected of having epidermolysis bullosa (EB) is large and testing them time consuming. Our aim was to develop a targeted next generation sequencing (NGS) gene panel that enables to analyze the large number of EB-genes in patients with EB and skin fragility at relatively low cost and high speed. MethodsAn EB-panel of 33 genes involved in EB and related skin fragility disorders was developed using the Agilent Sure Select Target Enrichment® for simultaneous mutation detection and validated in 15 patients with 21 gene variants previously identified through Sanger sequencing. The samples were sequenced using 151 base pair paired-end reads on an Illumina MiSeq® sequencer and analyzed using Softgenetics' NextGENe® and Cartagenia's Benchlab NGS® software. Subsequently, for diagnostic purposes, the targeted EB-panel was used to analyze 27 patients, in whom a pathogenic mutation had not been identified. Results and ConclusionIn the validation phase, 20/21 pathogenic mutations, unclassified variants and neutral polymorphisms detected previously were identified by the EB-panel except for one large duplication of 77bp. In 14 out of 27 patients, novel mutations were identified which were all confirmed by Sanger sequencing. The use of this gene panel facilitates a much faster analysis of the large number of currently known EB-genes for mutation analysis than the single gene approach. In three patients, mutations were identified in genes not considered the prime candidates at initial clinical evaluation. Gene panel analysis thus adds to fine tuning of the clinical phenotype and expands the genotype-phenotype correlations for EB. The EB-panel contains genes having known pseudogenes (i. e. KRT5, KRT14, etc) and exons with high sequence homology. Nonetheless, mutations could be accurately mapped to the correct genes based on the accurate detection of single nucleotide variations between genes-pseudogenes and homologous exons by analyzing the sequence of single reads without the need of additional and elaborate techniques.

2899F

Clinical and genetic analysis of Wiedemann–Steiner syndrome caused by *KMT2A* mutations. *N. Miyake[1], Y. Tsurusaki[1], E. Koshimizu[1], N. Niikawa[2], N. Matsumoto[1].* 1) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 2) Health Science University of Hokkaido, Hokkaido, Japan.

Wiedemann–Steiner syndrome (WSS, MIM#605130) is an autosomal dominant congenital anomaly syndrome characterized by hairy elbows, dysmorphic facial appearances (hypertelorism, thick eyebrows, downslanted and vertically narrow palpebral fissures), pre- and post-natal growth deficiency, and psychomotor delay. WSS is caused by heterozygous mutations in *KMT2A* (lysine [K]-specific methyltransferase 2A; NM_001197104. 1, previously called *MLL*), a gene encoding a chromatin methyltransferase. To date, only eight *KMT2A* mutations have been reported with limited clinical features. Here, we identified six novel *KMT2A* mutations in six WSS patients using whole exome sequencing. We confirmed four mutations occurring *de novo*. Interestingly, some of the patients were initially diagnosed with atypical Kabuki syndrome probably owing to overlapping clinical features including similar facial expressions and/or hypertrichosis. Kabuki syndrome is another autosomal or X-linked congenital anomaly syndrome which is caused by mutations in *KMT2D* or *KDM6A*, which are also involved in histone methylation as *KMT2A*. Then, we clinically re-evaluated them and retrospectively diagnosed five of six patients were WSS. Only one patient was initially diagnosed with atypical Kabuki syndrome but could not be clinically re-evaluated because she is now untraceable. Furthermore, we compared 56 clinical features between KS and WSS whose causative mutations have been identified. Long palpebral fissures, ectropion of the lateral third lower eyelids, and short fifth fingers were more commonly observed in KS patients with *KMT2D* mutations. As expected, generalized hypertrichosis was more commonly observed in WSS patients with *KMT2A* mutations. Interestingly, postnatal growth retardation was observed in all WSS patients while it was observed in less than half of KS patients with *KMT2D* mutation. These features may be hallmarks for differential diagnosis, with the other features observed in both WSS and KS. Clinical similarity between WSS and KS may be explained by the common protein function of KMT2A, KMT2D, and KDM6A, as they all play important roles in histone modification, and activate gene expression via H3K4 methylation (KMT2A and KMT2D) or H3K27 demethylation (KDM6A).

2900W

A Novel *OFD1* Missense Mutation Causes an Autosomal Recessive Dyskeratosis Congenita-Like Disorder Further Complicating the Clinical Heterogeneity of *OFD1* Mutations. *H. Shaath[1], A. Fadda[2], Y. Al-Sharraj[1], S. Tomei[3], E. Wang[3], H. El-Shanti[1,4], M. Kambouris[5,6].* 1) Medical Genetics Center, Qatar Biomedical Research Institute, Doha, Qatar; 2) Biomedical Informatics, Sidra Medical & Research Center, Doha, Qatar; 3) Research Branch, Sidra Medical & Research Center, Doha, Qatar; 4) Pediatrics, University of Iowa, Iowa City, IA, USA; 5) Pathology-Genetics, Sidra Medical & Research Center, Doha, Qatar; 6) Genetics, Yale University School of Medicine, New Haven, CT, USA.

A consanguineous [first cousin marriage] family of Arabic ethnic origin with four unaffected siblings and two male siblings affected by a Dyskeratosis Congenita-like disorder, was studied by Genome-Wide SNP homozygosity mapping, functional and positional candidate gene screening by Sanger sequencing,and Whole Genome Sequencing [WGS] to identify the offending gene and mutation. The disorder is marked by short stature, sparse hair including eyelashes, leukoplakia, dental carries and early tooth loss, osteoporosis, skin atrophy and hyper pigmentation, nail dystrophy with longitudinal ridging and splitting without bone marrow involvement. The offending gene was mapped to two possible homozygous genomic regions on chromosomes 3p and 6q cumulatively spanning 24Mb containing 86 protein-coding genes. Functional and positional candidate gene screening by Sanger sequencing for *ARL6IP, BCKDHB, DKC1, DNAH17, EEF1A1, LRIG1, ORC3, POLA1, SLC17A5* and *SYNCRIP*did not identify causative pathogenic mutations. Analyses and mining of WGS data for homozygous variants within the homozygous regions and X-linked variants [as only males are affected] in X-shared regions among affected siblings, identified a homozygous missense c. C2353T/p. P785S mutation in the *OFD1* gene, at Xp22. 2. The mutation co-segregates with the disease phenotype within the family, is absent in all public databases and in 500 ethnically matched control chromosomes. *OFD1* is a 1012 aa protein component of centrioles, controlling centriole length and involved in cilium biogenesis. *OFD1* mutations cause Oral-facial-digital syndrome 1, Simpson-Golabi-Behmel Syndrome Type 2, Joubert Syndrome 10 and Retinitis Pigmentosa 23 displaying significant phenotypic heterogeneity and clinical variability. The c. C2353T/p. P785S mutation ads a Dyskeratosis Congenita-like phenotype to the wide spectrum of *OFD1* disease phenotypes. In addition, it demonstrates that when applying WGS data to clinical diagnoses, all significant variants should be considered and scrutinized for their clinical impact, irrespective of the observed clinical presentation.

2901T

The Contribution of Mutations in Syndrome-Causing Genes to Non-Syndromic Retinitis Pigmentosa. *M. Xu[1,2], F. Wang[1,2], L. Zhao[1,2,3], Q. Zhang[1,2], E. Jones[1,2], X. Song[1,2], Z. Ge[1,2], K. Wang[1,2], R. Sui[4], R. Koenekoop[5], G. Silvestri[6], F. Porto[7,8], K. Zhang[9,10], B. Lam[11], D. Birch[12,13], S. Daiger[14,15], H. Wang[1,2], Y. Li[1,2], R. Chen[1,2,3,16,17].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Structural and Computational Biology and Molecular Biophysics Graduate Program, Baylor College of Medicine, Houston, TX; 4) Department of Ophthalmology, Peking Union Medical College, Beijing, China; 5) Ocular Genetics Laboratory, McGill University, Montréal, Canada; 6) Centre for Experimental Medicine, Queen's University Belfast, Belfast, United Kingdom; 7) INRET - Clínica e Centro de Pesquisa, Belo Horizonte, Brazil; 8) Departamento de Retina e Vítreo, Centro Oftalmológico de Minas Gerais, Belo Horizonte, Brazil; 9) Shiley Eye Center and Institute for Genomic Medicine, University of California-San Diego, La Jolla, CA; 10) United States Veterans Administration Healthcare System, San Diego, CA; 11) Bascom Palmer Eye Institute, University of Miami, Miami, FL; 12) Retina Foundation of the Southwest, Dallas, TX; 13) Department of Ophthalmology, The University of Texas Southwestern Medical Center, Dallas, TX; 14) Department of Ophthalmology and Visual Sciences, University of Texas Health Science Center, Houston, TX; 15) Human Genetics Center, University of Texas Health Science Center, Houston, TX; 16) The Verna and Marrs Mclean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX; 17) Program in Developmental Biology, Baylor College of Medicine, Houston, TX.

Retinitis pigmentosa (RP) is a rare inherited retinal disease affecting 1 in 3,000 to 7,000 of the population. It is featured by juvenile or adult-onset progressive retinal degeneration. The genetic etiology of RP is remarkably heterogeneous, with over 70 disease-causing genes identified. RP can be isolated with no extra-ocular symptoms manifested, or can be syndromic, which is frequently seen in ciliopathies and other syndromes. Recently, occasional reports showed that mutations in syndrome-causing (SDC) genes can also lead to non-syndromic RP. These reports helped to reveal the tissue-specific roles played by those genes and differential damaging effect of their disease-causing variants. However, due to the rarity of these studies, the proportion of non-syndromic RP cases involving SDC genes has never been evaluated. Here, by capture sequencing screening mutations in all genes involved in retinal degeneration phenotype, we systematically evaluated the genetic basis of over 1,200 non-syndromic RP cases. Specifically, we have identified about 20 cases which can be solved by mutations in SDC genes with good confidence. These genes include *BBS1*, *BBS5*, *BBS9*, *IFT140*, *WDR19*, *CLRN1*, *VPS13B*, etc. . The results provide the first estimate that mutations in SDC genes account for approximately 2% of non-syndromic RP cases. Our findings highlight the benefit of including all genes involving retinal degeneration phenotype during molecular diagnosis and provide hints for the retinal biology of these genes and their variants.

2902F

A truncating mutation in *CLMN* is associated with intellectual disability. *A. Alkhateeb[1,2], W. Habbab[1], R. Thompson[1], S. Aburahma[2].* 1) Qatar Biomedical Research Institute, Doha, Qatar; 2) Jordan University of Science and Technology, Irbid, Jordan.

Objectives: Rare autosomal recessive disorders have a higher occurrence in Arab countries due to higher rate of consanguinity. Here, we identified a consanguineous Arab family co-segregating a phenotype of intellectual disability in two out of a seven sibs, all children were females. We aimed to characterize the genetic etiology causing the phenotype in this family. **Methods:** One patient and the parents were genotyped by Illumina chip microarray to identify homozygous intervals. This was followed by whole exome sequencing. Mutations were filtered using Ingenuity variant analysis and validated by Sanger DNA sequencing. **Results:** Homozygosity intervals were identified in chromosomes 5, 8, 11, 13, 14, 18, 19, 20. After filtration, only 3 mutations were identified in the homozygosity intervals. Two missense mutations were reported previously and one frameshift mutation has not been reported in any database; c. 2585_2586insT (K863*) in *CLMN* gene (NM_024734. 3) results in truncation of the protein and a loss of 140 amino acids from the c-terminal domain which includes the transmembrane domain. The mutation was validated by Sanger sequencing and it was neither found in 200 controls from the same community nor found in any database. *Clmn* is needed both for all-trans retinoic acid to induce mouse neuroblastoma cells exit of cell cycle (requisite step in neuronal differentiation) and for the promotion of elongation and outgrowth of neurites. In addition, *Clmn* downregulates the expression of cyclin D1, which plays a key step in the decision to move through the G1 to S transition and cell cycle progression. In developing mouse embryos, *Clmn* is expressed in actively proliferating cells and in differentiating neurons. **Conclusion:** We have identified a novel truncating mutation in *CLMN* gene in a consanguineous family segregating intellectual disability phenotype. *CLMN* gene product plays an important role in neurogenesis and is found in regions where newly differentiated neurons are located and it is expressed in adult hippocampus and cerebellum, thus considered a plausible candidate gene. This is the first study to implicate *CLMN* gene in the pathogenesis of intellectual disability. Functional analysis should confirm the role of this mutation.

2903W

An Integrative Approach to Characterizing Human Diseases and Identifying Human Disease Genes. *A. Chhibber, J. Li, N. B. Asadi, H. Y. K. Lam.* Bina Technologies, Roche Sequencing, Redwood City, CA.

As the accessibility of sequencing data grows, exploration of potential disease-causing genetic events across the exome have become more and more commonplace. In such studies, mutations or other genetic perturbations that impact disease causality must be culled from the complex genetic background in individuals. For example, identification of disease-causing mutations from sequencing data typically involves narrowing thousands of variants down to a set of potentially pathogenic events. However, given that even healthy individuals carry on the order of 100 loss-of-function events, determining precisely which genes play a role in disease causality remains a significant challenge without extensive experimental follow-up. Computational approaches to gene-disease association can aid in identification and prioritization of candidates. A number of characteristics have been identified that each independently show modest utility in separating disease genes from non-disease genes, including the number and type of interactions with other genes or proteins, evolutionary conservation and polymorphism across human populations, and the size of the gene, among others. However, given that human diseases vary greatly in their pathogenesis and clinical severity, it is likely that different features will be valuable in identifying gene-disease relationships for different disease types.

In this study, we use the Human Phenotype Ontology to study the characteristics of human disease causing genes by disease type, and identify unique signatures of disease genes for different human disease classes. Further, we create a comprehensive database of genic features (including network centrality measures from protein-protein and coexpression networks, presence and role in biological pathways, gene expression signatures, known or predicted function, direct and indirect interacting genes, conservation across species, variation across human populations, as well as additional sequence and structural features) for human genes that can be used to facilitate interpretation of human disease-gene studies and identify potential novel disease genes.

2904T

Progressive encephalopathies in children: Clinical and molecular characterization. *E. Frengen[1], C. R. J. Pedurupillay[1], T. Barøy[1], A. Holmgren[1], M. D. Vigeland[1], T. Hughes[1], J. R. Helle[1], N. Skauli[1], I. Akkouh[1], E. Carlsen[1], S. S. Amundsen[1], M. Fannemel[1], D. Misceo[1], P. Strømme[2].* 1) Department of Medical Genetics, Oslo University Hospital and University of Oslo, Oslo, Norway; 2) Women and Children's Division, Department of Clinical Neurosciences for Children, Oslo University Hospital and University of Oslo, Oslo, Norway.

Progressive encephalopathies (PE) in children are clinically and genetically heterogeneous. Affected individuals usually exhibit developmental arrest before regression, and cerebral MRI show loss of neural tissue or abnormal signaling intensity. Our aim was to identify disease-causing genetic variants in undiagnosed PE and to perform functional analyses in novel disease genes. CNS infection, trauma, vascular accidents and sequelae after asphyxia and/or prematurity were ruled out. We have so far collected 80 PE patients from 50 families presumed to have a genetic etiology. Neurological signs included: ataxia, epilepsy, dystonia, autism, spasticity, retinopathy (stationary or progressive), dementia, intellectual disability, peripheral neuropathies and encephalopathies. Cerebral MRI examinations revealed cortical atrophy, subcortical white matter changes, cerebellar degeneration, or basal ganglia abnormalities. Congenital anomalies and dysmorphic traits, also those outside the nervous system, were evaluated using the London Medical Database for syndrome identification. After karyotyping, aCGH and analysis of candidate genes by MLPA and/or Sanger sequencing, we performed Whole Exome Sequencing (WES) in family trios (one child and parents) or inverted trios (two children and one parent) and continued with data filtering (population frequency, estimated severity of variants, inheritance pattern). Genetic variants were verified with Sanger sequencing and studied *in silico*. Hypotheses about their molecular consequences were explored in *in vitro* experiments in cell lines mostly established from the patients. We have so far detected the putative disease causing gene variant in 24 out of 50 families. In six families the genetic findings immediately led to a change in treatment: e. g. diet adjustment, hematopoietic stem cell transplantation (HSCT), or administration of NMDA receptor antagonists or a dopamine agonist. Six of the disease-causing variants were identified in genes not previously linked to human disease. These are studied functionally. We perform Whole Genome Sequencing on selected families where we have not detected clinically relevant WES results.

2905F

Canine disorders as models for corresponding human conditions - new mutations and aspects of rare diseases. *M. K. Hytönen[1,2,3], M. Arumilli[1,2,3], S. Hundi[1,2,3], E. Salmela[1,2,3], A. K. Lappalainen[4], T. Jokinen[4], P. L. Lukinmaa[5], E. Sarkiala-Kessel[4], A. Iivanainen[2], J. Kere[3], P. Nieminen[5], H. Lohi[1,2,3].* 1) Research Programs Unit, Molecular Neurology, Faculty of Medicine, University of Helsinki, Helsinki, Finland; 2) Department of Veterinary Biosciences, University of Helsinki, Helsinki, Finland; 3) Folkhälsan Research Center, Helsinki, Finland; 4) Equine and Small Animal Medicine, Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland; 5) Department of Oral and Maxillofacial Diseases, Faculty of Medicine, University of Helsinki, Helsinki, Finland.

Dog has established its position as a comparative model for human medicine due to shared genes, disorders and physiology with man. Our canine disease genetics program aims not only to improve canine health but also to provide information about the gene function and disease pathology that can be applied in human medicine. In this study, we have utilized various clinical, pathological and genetic approaches e. g. genome wide association study (GWAS), whole exome sequencing (WES) and whole genome sequencing (WGS) to clinically and genetically characterize several canine Mendelian diseases. First, we have encountered a dental disease with hypomineralized dentine resulting in severe attrition of teeth. Through WGS of affected and control dogs and filtering according to recessive model, we identified a mutation in FAM20C gene that has previously been associated with human hypophosphatemia conditions (Raine syndrome). Second, we describe a novel developmental disorder with prominent mandibular prognathia and joint hyperlaxity. We performed GWAS and mapped a locus overlapping human 22q11. 2 deletion syndrome region. Subsequently, we performed targeted resequencing for the associated region using two cases and three controls. We identified a homozygous variant likely causing the disease, and are currently validating the variant to confirm its causality. In addition, we have ongoing WES and WGS analyses on several other disorders that are expected to reveal a number of new disease mutations. This study establishes large animal models for human rare disorders for improved understanding of gene functions, disease mechanisms and comparisons of phenotypic overlaps across species. Dogs share many rare disorders with human and could serve as important preclinical models for new therapeutic approaches.

2906W

A point mutation in *PDGFRB* causes autosomal dominant Penttinen syndrome. *J. J. Johnston[1], M. Y. Sanchez-Contreras[2], K. Keppler-Noreuil[1], J. Sapp[1], M. Crenshaw[1], V. Cormier-Daire[3], R. Rademakers[2], V. P. Sybert[4], L. G. Biesecker[1].* 1) Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD; 2) Department of Neuroscience, Mayo Clinic, Jacksonville, FL; 3) INSERM, Département de Génétique, Université Paris Descartes Sorbonne Paris Cité, Institut Imagine, Hôpital Necker-Enfants Malades, Assistance Publique-Hôpitaux de Paris, Paris; 4) Division of Medical Genetics, University of Washington School of Medicine and Group Health Cooperative, Seattle, WA.

Penttinen syndrome is a distinctive disorder characterized by a prematurely aged appearance with lipoatrophy, epidermal and dermal atrophy along with scar-like hypertrophic lesions, thin hair, proptosis, underdeveloped cheekbones, and marked acro-osteolysis. All patients have been sporadic. Exome sequencing of an affected individual identified a *de novo* c. 1994T>C p. Val665Ala variant in *PDGFRB*, which encodes the platelet-derived growth factor receptor b. Three additional unrelated individuals with this condition were shown to have the identical variant in *PDGFRB*. Surprisingly, mutations in this gene have previously been associated with three other disorders. Distinct mutations in *PDGFRB* have been shown to cause infantile myofibromatosis, idiopathic basal ganglia calcification type 4, and an overgrowth disorder with dysmorphic facies and psychiatric symptoms, none of which overlaps with the clinical findings in Penttinen syndrome. Based on these substantially heterogeneous clinical phenotypes, we hypothesized that at least three of the four disorders were caused by gain of function mutations that activated distinct downstream signaling effectors. We evaluated the functional consequence of the Penttinen variant on the PDGFRB signaling pathway by transfecting mutant and wild-type cDNA into HeLa cells, showing that this mutation leads to ligand-independent constitutive signaling through STAT3 and PLCγ. We are currently evaluating the signaling consequences of other putative gain of function mutations in this gene. Based on these results, we conclude that Penttinen syndrome is a clinically distinct genetic condition caused by a gain of function mutation in *PDGFRB* that is associated with a specific and unusual perturbation of receptor function with aberrant STAT3 and PLCγ signaling.

2907T

The mutation spectrum of a large Brazilian inherited retinal disease cohort. *E. M. Jones[1, 2], Z. Soens[1, 2], S. A. Sampaio[3], I. F. G. Sena[4], K. C. S. Magalhaes[4], Y. Li[1, 2], S. Xu[1, 2], R. T. Simoes[4], F. B. O. Porto[3, 5], R. Chen[1, 2].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) INRET - Clínica e Centro de Pesquisa, Belo Horizonte, Minas Gerais, Brazil; 4) Instituto de Ensino e Pesquisa da Santa Casa de Belo Horizonte, Belo Horizonte, Minas Gerais, Brazil; 5) Departamento de Retina e Vítreo, Centro Oftalmológico de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil.

Inherited retinal dystrophies (IRDs) affect approximately 1 in 2,000 individuals. These disorders range from most severe cases of congenital blindness to more delayed cases of progressive visual and retinal degeneration in adulthood. A significant amount of clinical and genetic heterogeneity among these disorders makes diagnosis difficult without comprehensive sequencing of potentially causative genes. IRDs, moreover, can manifest in autosomal dominant, autosomal recessive, and X-linked inheritance patterns, adding to the complexity. Cohort studies in multiple populations have highlighted distinct genetic effects based on population. Here we characterize the molecular cause of IRDs in a large Brazilian cohort, a prominent world population that has largely remained unstudied. We report on the molecular diagnosis of 231 individuals in an available cohort of 1,492 Brazilians affected by IRDs. Through targeted capture and next-generation sequencing of 224 known retinal disease genes, we identify the leading molecular causes of retinitis pigmentosa (RP), Leber's congenital amaurosis (LCA), Stargardt's disease, and Usher syndrome among numerous other IRDs in our cohort. The solving rates for IRDs found within our cohort are slightly higher than generally reported rates in the current literature. This is likely in part a result of improved access to reported pathogenic variants in numerous publicly accessible databases. Roughly 35% of pathogenic variants in our cohort have been previously reported to cause retinal disease. The other 65% are novel variants likely to be pathogenic as either putative loss-of-function variants or predicted damaging variants in a gene associated with the expected diagnosis. Notably, 11% of RP individuals were found to have pathogenic or likely pathogenic variants in *GPR98*. This is a much higher percentage than what is seen in other populations and may highlight the presence of a founder mutation. Furthermore, in LCA individuals the most common genes were *CRB1* (12%) and *ABCA4* (12%) with less than 1% of individuals affected by *CEP290* variants, differing significantly from that observed in Caucasian populations. These findings highlight the need for a precise and population-based molecular diagnosis of IRDs. With the development of current gene therapies for diseases such as LCA, an accurate molecular diagnosis is essential for appropriate genetic counseling and informed treatment recommendations.

2908F

A *PPP1R21* Frame-Shift Insertion Mutation Causes a Novel Autosomal Recessive Neuro-Developmental Disorder. *M. Kambouris[1,2], A. Rajab[3], A. Fadda[4], H. Shaath[5], Y. Al-Sarraj[5], S. Tomei[6], E. Wang[6], H. El-Shanti[5,7].* 1) Pathology-Genetics, Sidra Medical & Research Center, Doha, Qatar; 2) Genetics, Yale University School of Medicine, New Haven, CT, USA; 3) National Genetic Center, Ministry of Health, Muscat Oman; 4) Biomedical Informatics, Sidra Medical & Research Center, Doha, Qatar; 5) Medical Genetics Center, Qatar Biomedical Research Institute, Doha, Qatar; 6) Research Branch, Sidra Medical & Research Center, Doha, Qatar; 7) University of Iowa, Pediatrics, Iowa City, IA, USA.

A consanguineous [second cousin marriage] family of Arabic ethnic origin with two unaffected male siblings and two female siblings affected by a novel neuro-developmental disorder, was studied by Genome-Wide SNPhomozygosity mapping and Whole Genome Sequencing [WGS] to identify the offending gene and mutation. The disorder is marked by cognitive impairment, severe global developmental delay, and congenital malformations that include macrocephaly, coarse facial features, prominent forehead with bitemporal narrowing, telecanthus,blue sclerae, with convergent squint, prominent nasal bridge, long philtrum, low set ears, pectus carinatum, scoliosis, in-toeing and delayed bone age. Neurologic findings include attention deficit, dysarthria, hypotonia, clumsy ataxic gait and urinary and fecal incontinence. Brain MRI showed mild vermis hypoplasia, deep sulcation and thinning of the white matter at presylvian region bilaterally, in post portions of the brain, persistent cavum septi pellucidum. Additional findings include alternate esotropia, extensive caries and cardiac ASD in the older sibling, and myoclonus epilepsy in the younger. The offending gene was mapped to three possible homozygous genomic regions on chromosomes 2p, 10q and 20q, cumulatively spanning 20Mb containing 77 protein-coding genes. Analyses and mining of WGS data for homozygous variants within the homozygous regions from one affected sibling identified a homozygous protein-truncating c. 2169_2170insAGGT/p. K724fs, 4-bp insertion frame shift mutation in the *PPP1R21* gene, at 2p16. 3. *PPP1R21* [Protein Phosphatase 1, regulatory subunit 21] is a 780 aa protein with phosphatase binding properties and scant functional information. The c. 2169_2170insAGGT/p. K724fs insertion eliminates the last 45 C-terminal amino acids and causes a p. K724R substitution, immediately encountering a stop codon at position 725. The mutation co-segregates with the disease phenotype within the family, is absent in all public databases and in 500 ethnically matched control chromosomes. To date, this is the first report associating *PPP1R21* to an autosomal recessive neuro-developmental disorder.

2909W

A New Retinits Pigmentosa Gene Identified by Homozygosity Mapping and Whole Genome Sequencing. *N. H. Wang[1,2,3], S. J. Chen[4,5], C. F. Yang[2], H. W. Chen[2], H. P. Chuang[2], Y. H. Lu[6,7], Y. H. Huang[6,8], M. Y. Lo[6], C. H. Chen[2], Y. T. Chen[2], D. M. Niu[6,7], J. Y. Wu[2].* 1) Inst Biomedical Sci, Academia Sinica, Nankang, Taipei, Taiwan; 2) Molecular Medicine Program, Taiwan International Graduate Program; 3) Institute of Biochemistry and Molecular Biology, School of Life Sciences, National Yang-Ming University, Taipei, Taiwan; 4) Department of Ophthalmology, Taipei Veterans General Hospital, Taipei, Taiwan; 5) School of Medicine, National Yang-Ming University, Taipei, Taiwan; 6) Department of Pediatrics, Taipei Veterans General Hospital, Taipei, Taiwan; 7) Institute of Clinical Medicine, School of Medicine, National Yang-Ming University, Taipei, Taiwan; 8) Institute of Clinical and Community Nursing, School of Nursing, National Yang-Ming University, Taipei, Taiwan.

Retinitis Pigmentosa (RP) is an inherited retinal degeneration that leads to blindness and is characterized with broad heterogeneity both clinically and genetically. While more than 50 genes have been identified for the disease, these represent the cause in only 60% of the patients. To identify the gene mutation responsible for five families diagnosed with autosomal recessive RP (arRP) from a close community in Taiwan, we analysed the whole genome SNP genotypes of these family members with homozygosity mapping and located a 1. 8Mb region shared by all patients. Next generation sequencing was carried out and identified a missense mutation from this homozygous region. We have confirmed this mutation in all patients by Sanger sequencing and showed co-segregation in all five families. The identified mutation was cloned and expressed, and the protein was purified to compare the enzyme activity between wildtype and mutant proteins. We showed that the identified mutation resulted in decreased enzyme activity. The identified mutation locates in the coding region of an enzyme involved in glycosylation and was previously found to be responsible for several types of muscular dystrophies, including Walker-Warburg syndrome (WWS) or muscle-eye-brain disease (MEB), in addition to the less severe congenital form with mental retardation and a milder limb-girdle form. This is the first finding on the involvement of this gene in arRP.

2910T

***De novo* heterozygous *GMNN* mutations cause autosomal dominant primordial dwarfism associated with Meier-Gorlin syndrome.** *Y. Yang[1,6], L. C. Burrage[1], M. K. Eldomery[1], J. R. Willer[2], Z. Akdemir[1], M. Azamian[1], G. Zapata[1], W. Zhu[6], P. P. Hernandez[1], E. Davis[2], D. Lugtenberg[3], J. Schoots[3], S. Munnik[3], R. Roepman[3], S. Jhangiani[4], K. Katsanis[2], L. E. Vissers[3], H. G. Brunner[3], D. M. Muzny[4], R. A. Gibbs[1,4,6], C. M. Eng[1,6], J. R. Lupski[1,5], F. Xia[1,6], S. R. Lalani[1], E. M. H. F Bongers[3].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Center for Human Disease Modeling, Duke University Medical Center, Durham, NC; 3) Department of Human Genetics, Radboud university medical center, Nijmegen, the Netherlands; 4) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 5) Department of Pediatrics, Texas Children's Hospital, Houston, TX; 6) Exome Laboratory, Baylor Miraca Genetics Laboratories, Houston, TX.

Meier-Gorlin Syndrome (MGS) is a genetically heterogeneous primordial dwarfism syndrome caused by biallelic loss of function mutations in one of the five genes encoding pre-replication complex proteins: *ORC1*, *ORC4*, *ORC6*, *CDT1*, and *CDC6*. The molecular etiology was established in about three-fourths of MGS patients with disruption of the origin of replication initiation step as the primary cause and autosomal recessive as inheritance manner in all individuals studied. In about one-fourth of the clinically ascertained MGS patients, a molecular diagnosis is not obtained, suggesting the existence of additional causal genes for MGS. Here, we described three patients with a clinical diagnosis of MGS but who tested negative for pathogenic variants in the aforementioned known MGS disease genes. The patients shared clinical features including dwarfism, absent or hypoplastic patellae and microtia as well as the hallmark facial features of full lips with distinct ear abnormalities, narrow nose, high nasal bridge, microstomia, and micro/retrognathia. Heterozygous mutations located in the 5' end of the *GMNN* gene including p. K6X in patient 1 (nonsense), p. I12fs in patient 2 (frameshift) and p. K17R in patient 3 (missense) were identified by exome sequencing (patient 1 and 2) or Sanger sequencing (patient 3). Family studies did not detect the changes in the biological parents, indicating the three variants arose *de novo* in the patients. *GMNN* encodes geminin, an inhibitor of DNA re-replication that is expressed during the S and G2 phases of the cell cycle and degraded during the metaphase-anaphase transition by the anaphase-promoting complex (APC) recognizing the destruction box sequence in the geminin protein. Geminin acts by inhibiting DNA replication factor Cdt1, which has been known to be associated with MGS, and prevents loading of minichromosome maintenance proteins Mcm2-7 to the origins of replication. All the three *GMNN* mutations in our patients occurred 5' to the destruction box that is located from amino acid p. R23 to p. P31. We hypothesize that these MGS associated variants result in gain of function geminin proteins with diminished destruction signal, leading to increased protein stability, sequestered Cdt1 and decreased replication. In summary, gain of function de novo mutations in a candidate new disease gene *GMNN* causing inhibition of replication via sequestration of Cdt1 is the likely mechanism for the dominantly inherited MGS in the three individuals.

2911F

Mutation spectrum of autosomal recessive retinitis pegmentosa patients of India identified by whole exome sequencing. *X. J. Zhu[1], Y. Zhou[1], Y. Di[1], Z. Jiang[1], X. Zhu[1], P. Sundaresan[2], Z. Yang[1].* 1) Center for Molecular Genetics, Sichuan Provincial People's Hospital, Chengdu, Sichuan, China; 2) Department of Genetics, Aravind Medical Research Foundation, Aravind Eye Hospital, Madurai, Tamilnadu, India.

Retinitis pigmentosa (RP) is a group of inherited retinal diseases affecting the light-sensing photoreceptor cells. It is caused by mutations in at least 50 genes. However, known mutations can only explain around 55% of RP patients. To identify genetic mutations underlying autosomal recessive retinitis pigmentosa (arRP), we performed whole exome sequencing study on 14 consanguineous marriage Indian families. Here we reported identified mutations in *FAM161A, EYS, PDE6B*, *TRIM1*, *BBS10* and *NR2E3 in 10* families. Sequencing analysis revealed the presence of a novel homozygous frame shift mutation p. R592FsX2 in both patients of family RP-252 and family RP-182. We also identified two novel homozygous mutations *EYS* in families with arRP. Mutation types identified include missense mutation, frameshift deletion, and stopgain. Meanwhile, these mutations were absent in 1000 ethnicity-matched control samples screened by direct Sanger sequencing. In conclusion, we identified novel mutations of FAM161A, EYS, PDE6B, TRIM1, BBS10 and NR2E3 in arRP patients in India population. Our study expands our knowledge of mutation spectrum of arRP in India population.

2912W

The INVESTICATE project: Identification of New Variation, Establishment of Stem cells, and Tissue Collection Aimed at Treatment Efforts. *L. Crapper[1,2], R. Pujol[2], W. Al-Hertani[4], C. Ernst[3,4].* 1) Integrated Program in Neuroscience. McGill University, Montreal, QC; 2) McGill Group for Suicide Studies, Montreal, QC; 3) Department of Psychiatry, McGill University, Montreal QC; 4) Department of Medical Genetics, Alberta Children's Hospital, Calgary, AB.

Neurodevelopmental disorders (NDDs) are a large and complex group of disorders with varying etiologies. Recent advances in genomics and sequencing technologies have accelerated the pace of discovery of genes associated with NDDs, and suggested a large portion of non-syndromic NDDs are caused by rare, highly penetrant genetic variations. Exome sequencing studies and analysis of copy number variations in families are very effective in characterizing the genetic variability inherent in NDDs and suggesting that causal mutations cluster in several biological pathways, but so far they have not been successful in leading to the development of treatments for NDDs. The INVESTICATE project recruits patients for whom traditional genetic analysis has not identified the cause of their NDD, and uses exome or whole genome sequencing to provide state-of-the-art genetic information to families. After variant identification we generate patient derived induced pluripotent stem cells, which are used to determine functional information about the genetic variants we have identified. Stem cell data is complemented by the collection of whole brains from those cases that die during the course of the study. Finally we use high throughput screening to assess potential therapeutic molecules. INVESTICATE aims to provide genetic information to families not avialble through the pubic health system, understand how genetic variants cause neurological disease, and develop therapeutic strategies to provide treatment to affected individuals.

2913T

***HERC1*: a possible new candidate gene for Intellectual Disability.** *A. Das Bhowmik[1], S. Aggarwal[1,2], A. Dalal[1].* 1) Diagnostics Division, Centre for DNA Fingerprinting and Diagnostics, Hyderabad, India; 2) Department of Medical Genetics, Nizam's Institute of Medical Sciences, Hyderabad, India.

We studied a consanguineous family with unique phenotype of cognitive delay and dysmorphism in two affected children of both sexes. We did exome sequencing for both the affected children to assess coding sequence variants. Exome capture was performed on genomic DNA sample using Agilent SureSelect V5 exome capture kit. The libraries were sequenced to mean >80-100X coverage on Illumina HiSeq2000 platform. The reads obtained were mapped against human reference genome (GRCh37/hg19) followed by detection of SNPs and small indels. Variant annotation was performed using Annovar for location and predicted function. Exome sequencing analysis identified a novel c. 4906-2A>C 3' splice-site mutation variant in *HERC1* gene. The variant was shared by both the siblings. Array comparative genomic hybridization (CGH) was also performed in one affected child using Affymetrix CytoScan 2. 7M Array which identified a 12. 9 Mb LOH region on chromosome 15. *HERC1* gene was present in this region. Sanger sequencing in both the siblings and mother confirmed presence of the mutation in children and heterozygous status in mother. Further cDNA analysis using SuperScriptTM III First-Strand cDNA Synthesis assay system (Invitrogen) in both the patients revealed that this mutation results in a 10 base pair deletion in exon 27 leading to a frameshift that is likely to result in premature truncation of the protein at 1663 amino acids from start codon (original protein size 4861 amino acid). This is likely to affect the protein function which ultimately can lead to the disease. Herc1 is a giant protein with multiple structural domains, expressed ubiquitously in all human and mouse tissues, with the highest level of expression in testis and brain. This protein is believed to be involved in intra cellular membrane trafficking and ubiquitinization of specific targets which are essential for eukaryotic cell function. Recently another member of this group of protein family (Herc2) has been implicated in nonsyndromic intellectual disability and autism. Till date no other disease causing mutation in humans has ever been reported in *HERC1* gene. Hence we suspect *HERC1* could be a new candidate gene for intellectual disability. Further functional characterization of this mutation is in progress.

2914F

***De novo* mutations in *KAT6A* identified by whole exome sequencing cause a neurodevelopmental disorder with syndromic features.** *F. Millan[1], M. T. Cho[1], K. Retterer[1], K. G. Monaghan[1], R. Bai[1], P. Vitazka[1], D. B. Everman[2], B. Smith[2], N. L. Champaigne[2], L. M. Baggett[2], B. Angle[3], V. Roberts[3], L. Immken[4], H. Nagakura[4], M. DiFazio[5], E. Sherr[6], E. Haverfield[1], B. Friedman[1], A. Telegrafi[1], J. Juusola[1], W. K. Chung[7], S. Bale[1].* 1) GeneDx, Gaithersburg, MD, USA; 2) Greenwood Genetic Center, Greenwood, SC, USA; 3) Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA; 4) Specially For Children Genetics, Austin, TX, USA; 5) Department of Neurology, Children's National Health Care System, Washington, DC, USA; 6) Institute of Human Genetics, University of California, San Francisco, San Francisco, CA, USA; 7) Departments of Pediatrics & Medicine, Columbia University Medical Center, New York, NY, USA.

Neurodevelopmental disorders (NDDs) affect 1-3% of the general population. Their underlying genetic cause remained unexplained in many patients until next generation sequencing became available in clinical practice. Clinical whole-exome sequencing (WES) has become a powerful tool for the identification of pathogenic variants leading to Mendelian disorders, among which NDDs represent a large proportion. Performing WES with a trio approach has proven to be extremely effective in identifying *de novo* variants as a common cause of NDDs. Recently, *de novo* pathogenic variants in *KAT6A* have been reported to cause syndromic intellectual disability, including hypotonia, microcephaly, facial dysmorphism, craniosynostosis and/or cardiac defects. The mutation spectrum included sequence variants in exons 17 and 18 resulting in premature stop of translation, and one gene deletion encompassing the entire reading frame of *KAT6A*. We report seven additional unrelated individuals with a strikingly similar NDD phenotype including hypotonia, severe speech delay, and facial dysmorphism. Two of them had microcephaly and one congenital heart disease. WES trio testing revealed that each proband was heterozygous for a novel, *de novo* variant in the *KAT6A* gene. These variants included 2 frameshift and 3 nonsense variants in exons 7, 17 and 18, 1 canonical splice site variant predicted to damage the acceptor site for exon 18 and 1 missense variant located at the catalytic MYST-type histone acetyltransferase domain of the protein. The *KAT6A* gene encodes a histone acetyltransferase that has long been known for its structural involvement in acute myeloid leukemia. Loss of protein function in a published mouse model demonstrated craniofacial and cardiac defects due to homeotic transformations of both the axial skeleton and neural tube, suggesting that *KAT6A* plays a crucial role in development of the brain, face, and heart. In summary, our data expand the mutation spectrum of this newly described condition and identify exons 17 and 18 of *KAT6A* as mutation hotspots. Furthermore, we demonstrate that loss of *KAT6A* function is associated with a recognizable and relatively consistent phenotype.

2915W

A founder mutation in *C12orf4* gene - A candidate gene for autosomal recessive non-syndromic intellectual disability (NSID). *A. K. Philips[1], M. Pinelli[2,3,4], Y. C. Chen[5], A. Mustonen[6], T. Määttä[7], H. H. Arts[2,8], K. Wu[2], R. Roepman[2], J. S. Moilanen[6], S. Raza[1], T. Varilo[1], G. Scala[3], S. Cocozza[3], C. Gilissen[2], P. Panula[5], I. Järvelä[1].* 1) Department of Medical Genetics,University of Helsinki, Finland; 2) Department of Human Genetics, Nijmegen Centre for Molecular Life Science and Institute for Genetic and Metabolic disorders, Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands; 3) Dipartimento di Medicina Molecolare e Biotechnologie Mediche Université, Degle Studi di Napoli "Federico II", Italy; 4) Telethon Institute of Genetics and Medicine, Pozzuoli, Italy; 5) Neuroscience Center, Institute of Biomedicine, Department of Anatomy, University of Helsinki, Finland; 6) Department of Clinical Genetics, Medical Research Centre Oulu, Oulu University Hospital and University of Oulu, Finland; 7) Joint Authority for Kainuu, Disability services, Kainuu, Finland; 8) Department of Biochemistry, University of Western Ontario, Medical Sciences Building Rm. 342, Ontario, N6A 5C1, London, Canada.

Background: Intellectual disability is a major health problem in our society. Non-syndromic intellectual disability (NSID) represents a heterogeneous group of diseases that is often a diagnostic challenge in clinical medicine. **Methods and Results**: To facilitate gene identification, we applied whole-exome sequencing (WES) in a family with non-syndromic ID (NSID). This family originates from an isolated population in North-Eastern Finland. We found a missense variant in*C12orf4*gene on chromosome 12 in a consanguineous family with four affected males. The variant is inherited as an autosomal recessive trait. This variant has a high CADD score (23. 4), a high phyloP score (5. 02) and was predicted to be deleterious by the Condel server. Screening of our cohort of 200 NSID patients let to the discovery of another consanguineous family with two affected females and one affected male with the same missense mutation. In addition, two heterozygous carriers with different phenotypes were also found. The variant is enriched in the North Eastern subisolate of Finland with a carrier frequency of 1:53, characteristic to a founder effect. Localization studies in SH-SY5Y neuronal cells indicate that both the normal and mutated constructs are localized to the cell cytoplasm. Clinical features of the affected individuals are characterized by mild to severe NSID and delayed speech development. A frameshift mutation in exon 6 of the*C12orf4*gene has been previously reported in a Saudi family with three affected individuals with intellectual disability, ADHD and non-specific MRI changes. Our preliminary findings suggest that mutation in*C12orf4*gene causes autosomal recessive non-syndromic intellectual disability. **Ongoing work**: The function of this gene is not yet known. We are currently in the process of functionally characterizing this gene in zebrafish.

2916T

Analysis of a cohort of Polish patients with X linked intellectual disability by the targeted X chromosome exome sequencing (NGS). *S. O. Rzońca[1], M. Gos[1], A. Landowska[1], A. Charzewska[1], A. Kutkowska-Ka%8;mierczak[1], J. Castaneda[1], M. Janeczko[2], R. Posmyk[3], P. Stawiński[4], M. Rydzanicz[4], J. Kosińska[4], E. Obersztyn[1], R. Ploski[4], J. Bal[1].* 1) Department of Medical Genetics, Institute of Mother and Child, Warsaw, Poland; 2) Department of Medical Genetics, University Children's Hospital of Cracow, Cracow, Poland; 3) Centre of Clinical Genetics GENETICS, Bialystok, Poland; 4) Department of Medical Genetics, Medical University of Warsaw, Warsaw, Poland.

X-linked intellectual disability (XLID) is a term describing clinical entities caused by mutations in genes located on the X chromosome. Until recently, around 114 genes have been shown to be involved in NS-XLID, making a systematic screen useless in the diagnostic routine. The new molecular approach, next generation sequencing (NGS), provides a unique opportunity to analyze almost entire coding sequence of the chromosome X and is a promising method for the discovery of rare, disease causing sequence variants. In our study, we have performed the X chromosome exome sequencing to identify potentially pathogenic mutations in novel and/or known genes in patients with unexplained ID with a family history suggesting the X-linked inheritance. Twenty patients with excluded chromosomal aberrations (aCGH) and mutation in *FMR1* gene, were examined by targeted X chromosome sequencing. The X chromosome exome libraries were prepared with SureSelectXT X-Chromosome panel (Agilent, CA, USA) and sequenced on HiSeq1500 Sequencer (Illumina, CA, USA). Eleven potentially pathogenic variants were selected based on the results of *in silico* prediction tests (e. g. MutationTaster, SIFT, CADD) and data from population frequency databases (1000Genomes, ExAC, NHLBI/EVS, in-house variant database). The cosegregation analysis of the selected variants with phenotype confirmed the possible pathogenicity for 5 variants. Together, we have identified novel, not previously described, variants in 3 known genes associated with XLID (missense variants: *UBE2A*, *MID1,* nonsense: *DCX*). Moreover, variants in 2 new candidate genes have been found (nonsense variant: *FRMPD4*, missense: *MID2*). The functional analysis of the gene product function indicated that the proteins encoded by these genes are involved in proliferation, differentiation, adhesion and cell growth as well as ubiquitination and neuronal migration process. In conclusion, we have identified likely pathogenic mutations in 25% of the investigated patients (5/20). To further confirm the relation of novel variants with the etiology of intellectual disability, additional functional studies are needed. However, the presented data confirms that the NGS technique is highly useful tool to search for new variants in genes related to X – linked intellectual disability.

2917F

DCHS2, a novel autosomal recessive cause of Van Maldergem Syndrome. *N. Voisin[1], L. Ambrozaityte[2], A. Morkuniene[2], L. Gueneau[1], K. Mannik[1], Z. Ciuladaite[2], E. Preiksaitiene[2], E. Pranckeviciene[2], T. Rancelis[2], L. Cimbalistiene[2], N. Guex[3], V. Kucinskas[2], A. Reymond[1].* 1) Center for Integrative Genomics, University of Lausanne, Genopode Building, CH-1015, Lausanne, Switzerland; 2) Department of Human and Medical Genetics, Faculty of Medicine, Vilnius University, 08661, Vilnius, Lithuania; 3) Swiss Institute of Bioinformatics, Genopode building, CH-1015, Lausanne, Switzerland.

Within the frame of a Swiss-Lithuanian collaboration, we enrolled 132 families with one or two members affected with intellectual disability (ID) or developmental delay. 17 patients/families (12%) were found to carry causative CNVs such as del10q22. 1-q22. 3, del6q16. 1-q22. 31, del17p13. 2-p13. 3 and mos dup8p11. 22-q11. 23. The remaining patients and their unaffected parents and siblings were subjected to exome sequencing. In families with multiple affected members, variants were filtered based on quality scores, population prevalence, predicted deleterious and adherence to either an autosomal recessive inheritance or being *de novo* in both patients. For instance, we identified a family where 2 brothers shared mild ID with developmental delay (delayed speech and walk), a long face with synophrys, dental malocclusion, optic nerve atrophy, clinodactyly and skeletal abnormalities (arthralgia, joint dislocations, disc hernia and scoliosis). The older brother showed additional features such as strabismus, ptosis, high arched palate, sleep disturbance and hippocampal hypoplasia, whereas the younger one displayed febrile seizures, microdontia and optic nerves chiasm. Consistent with an autosomal recessive pathology they both harbored 4 heterozygous variants in *DCHS2*, 2 maternally- and 2 paternally-inherited. This cadherin gene is a paralog, of *DCHS1* and *FAT4*, mutations of which were previously associated with Van Maldergem syndrome 1 and 2, respectively (VMLDS; OMIM 601390 and 615546). Characteristic features of these cerebro-facio-articular syndromes overlap with several traits displayed by our patients, such as ID, delayed speech, flat face, ptosis, abnormal teeth, high-arched palate, hand and feet abnormalities with clinodactyly, scoliosis and hypoplastic optic nerve chiasm. Even though VMLDS are also characterized by microtia resulting in hearing loss and renal hypoplasia, clinical features not observed in our cases, the considerable overlap of traits suggest that the 2 brothers are affected by an autosomal recessive form of VMLDS caused by compound heterozygous variants in DCHS2, a gene not previously associated with these rare but very distinctive syndromes. UNIGENE project has received funding from Lithuanian-Swiss cooperation program to reduce economic and social disparities within the enlarged European Union under project agreement No CH-3-_MM-0.

2918W

Disease gene discovery in familial amyotrophic lateral sclerosis. *J. A. Fifita[1], K. L. Williams[1], K. Zhang[1], G. A. Nicholson[1,2], D. Rowe[1], I. P. Blair[1].* 1) Faculty of Medicine and Health Sciences, Macquarie University, Sydney, NSW, Australia; 2) University of Sydney, ANZAC Research Institute, Concord Hospital, Sydney, NSW, Australia.

Amyotrophic lateral sclerosis (ALS), also known as Lou Gehrig's disease, is a fatal neurodegenerative disease that is caused by the progressive death of motor neurons. Ten percent of cases are familial (FALS), with the remaining 90% considered sporadic (SALS). To date, the only known causes of ALS are gene mutations. Mutations in ALS genes, including *SOD1*, *TARDBP* and *C9ORF72,* only account for approximately 60% FALS and 10% of SALS. We aim to identify remaining gene mutations causing ALS. We apply both unbiased and hypothesis driven gene discovery strategies to analyze exome sequence data from Australian familial ALS cases. We have completed whole exome capture and sequencing of probands from 74 ALS families, as well as ten individuals from three multigenerational ALS families. Custom bioinformatics analysis of these families has produced small lists (5, 10, and 18) of novel candidate gene mutations. All variants were prioritized for functional studies based on factors such as protein predictions, conservation across species and gene function. The top-ranking candidates from each family will undergo functional studies to examine potential pathogenicity. Our ALS gene discovery strategies also include the identification and analysis of functional candidate genes, as well as mutation analysis of recently published ALS genes. Our standardized exome sequence data from probands allows rapid interrogation of candidate genes. This is illustrated by our recent analysis of *MATR3* a recently reported ALS gene, which demonstrates that mutations in this gene are a rare cause of Australian FALS. The penetrance of mutation-linked disease in familial ALS varies substantially, ranging from classic Mendelian inheritance to apparently sporadic disease. As such, familial ALS genes were also examined in apparently sporadic ALS cases. Rapid sequencing of 624 sporadic ALS cases using Fluidigm Access Array and Illumina MySeq sequencing is underway to identify variation in known ALS genes, including at least one novel gene identified by our laboratory. This has led to the identification of numerous novel non-synonymous variants in SALS cases, which may be a cause of, or contribute to ALS. The pathogenic basis of ALS remains poorly understood. The identification of novel ALS genes increases our knowledge of disease biology, adds to diagnostic regimes, and provides tools for the development of cell and animal models and long-term therapeutic discovery.

2919T

***CFTR* mutations in Chinese:Recurrent and novel mutations in cystic fibrosis.** *Y. Liu[1], J. Yang[1], X. Tian[2], K-F. Xu[2], X. Zhang[1]*. 1) Dept. of Medical Genetics, Peking Union Medical College, Beijing, Beijing, China; 2) Dept. of Respiratory Medicine, Peking Union Medical College Hospital, Beijing, China.

Background: Cystic fibrosis (CF) is a relatively common autosomal recessive disorder in Caucasians, caused by mutations in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene. The most common *CFTR* mutations have been well established among Caucasian CF patients. In Chinese, however, fewer than 40ChineseCFpatients are reported in the literaturesthus far. Previously, we reported on 7 Chinese CF cases and concluded that the mutations identified inChineseCFpatients seemed different from that in Caucasians. **Methods:**Here we further collected 8 more ChineseCFpatients, presented sequencing results in their*CFTR*gene, and summarized the recurrent and novel *CFTR* mutations from current (8) and our recent reported (7) ChineseCFpatients. The mutations were identified by sequencing the complete coding regions and flanking intronic sequences of the*CFTR*gene, followed by multiplex ligation-dependent probe amplification (MLPA) analysis in order to detect gene deletions or duplications. **Results:**In total, we identified 21 different mutations in 15 Chinese CF patients. We observed delF508, the most common *CFTR* mutation in Caucasians, in one patient as a heterozygote for the first time in Chinese CF patients. Four mutations are novel: ΔE7-E11; c. 3635delT, p. V1212AfsX15; c. 1997T>G, p. L666* and c. 2907A>C, p. A969A. Two mutations were recurrent: p. G970D, carried by 7 patients, was the most common one with a frequency of 23. 3% (7/30), and only seen in Chinese CFs previously; p. Q98R, carried by 2 patients, was the second common mutation with a frequency of 10% (3/30) and almost always seen in Asians. Furthermore, we noticed that another mutation together with p. G970D was always nonsense mutation, with only one exception (p. A969A). However minigene assay suggested that p. A969A led to skipping of exon 17 and generation of a premature stop codon after 3 codons. Other than delF508, we found only one more mutation (R553*) in one compound heterozygous patient that was included among the 32 common mutations in the screening panel for Caucasians. **Conclusion:**We further confirmed that the *CFTR* mutation spectrum ofChineseCFpatients was indeed quite different from that in Caucasians with p. G970D the most common one. Sequencing of the entire*CFTR*gene followed by MLPA analysis, rather than using the targeted sequencing-based screening panel for mutations commonly found in Caucasian populations, is recommended for genetic analysis in ChineseCFpatients.

2920F

Non-coding variants segregate with disease in South African families with keratolytic winter erythema (KWE). *T. Ngcungcu[1], B. Linghu[2], F. Yang[2], E. Oakeley[2], F. Staedtler[2], R. Bruccoleri[2], N. Nirmala[2], S. Buechmann-Moller[2], M. Marc Sultan[2], J. Szustakowski[2], M. Ramsay[1]*. 1) Division of Human Genetics, School of Pathology and the Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa; 2) Biomarker Development, Translational Medicine, Novartis Institutes for BioMedical Research, Basel, Switzerland and Cambridge, USA.

Keratolytic Winter Erythema (KWE) is a rare skin disorder of unknown aetiology that is characterized by erythema, keratin layer thickening and centrifugal skin peeling of the palms and soles, especially in cold and dry conditions. KWE shows an autosomal dominant mode of inheritance and the causal mutation was localised to chromosome 8p22-p23 (KWE critical region). The aim of this study was to identify and characterize the causal mutation for KWE in South African families. Targeted resequencing (8p22-23) and exome sequencing were performed in 21 affected and 30 unaffected individuals using Illumina HiSeq and MiSeq sequencing platforms. Reads were aligned to the reference genome using the Burrows-Wheeler Aligner (BWA) and the Genome Analysis Toolkit (GATK) and Pindel were used to call small and large structural variants, respectively. We identified a 7. 67 kb tandem duplication within the KWE critical region, upstream of the *CTSB* gene and encompassing two enhancers. This tandem duplication was validated in all affected individuals, highlighting the potential role of large structure variants for manifesting Mendelian diseases. Additionally, an *FDFT1* intronic variant also segregates with the disease. The two variants are in strong linkage disequilibrium, but the mechanism of action leading to the disease phenotypes remains unclear. The fact that these variants are located in non-coding regions, suggest that they may result in altered gene expression regulation. Further functional analyses are required, but these findings enforce the need to look beyond the exome when trying to identify causal mutations for a subset of Mendelian diseases.

2921W

Mutations in *STAC3*, the gene responsible for Native American Myopathy, are not responsible for the phenotypically similar Carey-Fineman-Ziter syndrome. *S. M. Robbins[1,2], A. Telegrafi[8], N. L. M. Sobreira[1], L. Fleming[9], C. E. Speck-Martins[6], J. C. Carey[5], D. FitzPatrick[7], D. Valle[1,4], J. E. Hoover-Fong[1,3,4].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore , MD 21205, USA; 2) Predoctoral Training Program in Human Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 3) McKusick-Nathans Institute of Genetic Medicine, Greenberg Center for Skeletal Dysplasias, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA; 4) Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA; 5) Department of Pediatrics, University of Utah, Salt Lake City, UT 84113, USA; 6) Genetic Unit, Sarah Rehabilitation Hospital, Belo Horizonte-MG, São Luiz-MA, Brasilia-DF, Brazil; 7) Human Genetics Unit, Medical and Developmental Genetics, University of Edinburgh Western General Hospital, Edinburgh, Scotland; 8) GeneDx, Gaithersburg MD 20877; 9) National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892.

Carey-Fineman-Ziter syndrome (CFZ) is a rare autosomal recessive phenotype characterized by short stature, hypotonia, Moebius sequence, Pierre Robin sequence, and myopathic facies. The molecular basis is unknown and only 17 cases have been reported. We describe 2 Qatari sisters born to first-degree consanguineous healthy parents, clinically diagnosed with CFZ syndrome. The younger sister, age 8 years, has short stature, dolichocephaly and turricephaly, downslanting palpebral fissures, ptosis, cleft palate, kyphoscoliosis, and hypotonia. Her sister, age 18 years, has short stature, cleft palate, and hypotonia; 3 male sibs are unaffected. We performed WES in the 2 affected sisters and one unaffected brother and identified a homozygous W284S variant in *STAC3* in both sisters which was heterozygous in the unaffected brother. Horstick et al. (2013) previously reported homozygosity of this same *STAC3* variant (*STAC3*-W284S) as the cause of Native American Myopathy (NAM) in 5 Lumbee Native American families with 18 affected individuals whose characteristic features included susceptibility to malignant hyperthermia, kyphoscoliosis, cleft palate, and myopathic facies. They showed that *STAC3* encodes a protein involved in excitation-contraction coupling in skeletal muscles through membrane voltage studies in a zebrafish mutant. Given the phenotypic similarity of individuals affected with CFZ syndrome and NAM, we performed Sanger sequencing on *STAC3* in 4 additional unpublished CFZ probands and 3 individuals (2 families) previously reported by Carey et al. (1982) and Ryan et al. (1999) for a total of 7 individuals with CFZ from 6 families. We did not identify rare (MAF < 0. 5%) coding variants in *STAC3* in these individuals. While we cannot rule out the possibility that non-coding or regulatory *STAC3* variants are present in all these individuals (n = 7), our results suggest that despite their phenotypic similarity, CFZ syndrome and NAM are caused by mutations in different genes. We conclude that our 2 affected Qatari sisters actually have NAM caused by homozygous *STAC3*-W284S variants and are the first non-Native American cases. Given their phenotypic similarity, however, we suggest that individuals with a clinical diagnosis of CFZ syndrome should be tested for variants in *STAC3* and speculate that the gene responsible for CFZ may encode a protein that functions in the same network as that encoded by *STAC3*.

2922T

Sequential sequencing approach elucidates high proportion of monogenic obesity cases in a consanguineous population. *S. Saeed[1], A. Bonnefond[2,3,4], J. Philippe[2,3,4], E. Durand[2,3,4], J. Manzoor[5], H. Crouch[1], O. Sand[2,3,4], M. Arslan[6,7], P. Froguel[1,2,3,4].* 1) Genomics of Common Disease, Imperial College London, London, United Kingdom; 2) European Genomic Institute for Diabetes (EGID), Lille, France; 3) CNRS-UMR8199, Lille Pasteur Institute, Lille, France; 4) Lille University, Lille, France; 5) Department of Paediatric Endocrinology, Children's Hospital, Lahore; 6) Department of Biological Sciences, Forman Christian College, Lahore, Pakistan; 7) Centre for Research in Molecular Medicine, The University of Lahore, Lahore, Pakistan.

Objective: Although 40-70% of human obesity is attributable to underlying genetic factors, most of the genes predisposing an individual to increased adiposity remain elusive. Monogenic forms of obesity have provided critical insights into mechanisms governing energy homeostasis, mainly by controlling appetite. In outbred populations, only 3-5% of severely obese patients have been identified with monogenic obesity, but its prevalence in consanguineous populations has not been systematically ascertained. We searched for obesity-associated mutations through a step-wise multilevel sequence analysis in children with severe obesity from Pakistani consanguineous families. Methods: The present investigation is based on 145 unrelated children of consanguineous parents, with a weight-for-age percentile >97. All subjects were, initially screened for mutations in the coding regions of leptin (*LEP*) and melanocortin 4 receptor (*MC4R*) genes by Sanger sequencing. Subjects negative for loss-of-function mutations in these genes were screened using microdroplet PCR-enrichment followed by next generation sequencing (NGS). Genomic structural variations were assessed by genome-wide genotyping. Samples from probands in which pathogenic mutations in known obesity-associated genes could not be identified using the aforementioned procedures, were analysed by whole exome sequencing (WES). Results: Using Sanger and microdroplet-based sequencing and genotyping we identified 44 obese probands carrying 14 different loss-of-function mutations in 5 genes: *LEP* (n=28), leptin receptor (*LEPR*) (n=8), *MC4R* (n=4), Bardet–Biedl syndrome-2 (*BBS-2*) (n=1) and *BBS-10* (n=2). We also foundtwo neighboring homozygous deletions that result in removing large segment of the *LEPR*. All mutations were in homozygous state. Lastly, WES has so far provided indication of an involvement of homozygous mutations insix susceptible genes from seven individuals. Although regarded as strong candidate genes for obesity through genome-wide association studies (GWAS) and/or animal studies, their involvement in monogenic obesity in human is reported here for the first time. Conclusion: The study shows a high prevalence (36%) of known and novel pathogenic mutations causing severe obesity in a consanguineous population of Pakistan. These results also underscore the importance of systematic genetic analysis of other inbred populations to unravel novel genetic variants and signalling pathways modulating energy homeostasis.

2923F

A loss-of-function mutation in *JAK1* is associated with epidermodysplasia verruciformis. *R. Wang[1], J. liu[2], L. Zhang[1], J. Li[1], S. Shu[1], D. Ma[2], X. Zhang[1,2].* 1) McKusick-Zhang Center for Genetic Medicine, Chinese Academy of Medical Sciences-Peking Union Medical College; 2) Peking Union Medical College Hospital, Chinese Academy of Medical Sciences-Peking Union Medical College, Beijing, China.

Epidermodysplasia verruciformis (EV) is a rare genodermatosis characterized by persistent, refractory and disseminated skin lesions on sun-exposed areas, and it has a high risk of skin carcinoma that results from an increased susceptibility to a specific group of human beta papillomavirus, such as the oncogenic HPV-5 and HPV-8. EV is often inherited in an autosomal recessive manner, and approximately 75% of EV patients are caused by mutations in *EVER1* and *EVER2*, which encode proteins involved in intracellular zinc transportation. Recently, we recruited a four-generation Chinese family showing an autosomal dominant transmission of EV. Genomic DNA samples were obtained from 5 affected individuals and 9 unaffected individuals after informed consent. Mutations in *EVER1* and *EVER2* have been previously excluded by Sanger sequencing. We performed whole-exome sequencing on four affected individuals, and identified a novel heterozygousdonor-splice site mutation (NM_002227. 2 c. 3258+1 G>A) in the Janus kinase 1 gene (*JAK1*) on chromosome 1p31. 3. The splicing variant co-segregated with the phenotype of all the family members expect III-21. This variant is not present in dbSNP142, the 1000 Genomes draft database and NHLBI Exome Sequencing Project and was not detected in chromosomes from 173 ethnically matched control individuals. Sanger sequencing of the cDNAobtained from peripheral lymphocytes of the affected individuals revealed that the splicing mutation disrupted normal splicing, skipped exon 23, then disrupted the reading frame and led to a premature termination at codon 1050. Quantitative RT-PCR on total RNA isolated from peripheral lymphocytes of the family members showed that all affected and carrier individuals had significantly reduced JAK1 expression compared to the unaffected individuals, suggesting that the splicing mutation might destabilize the mRNA transcript via the nonsense-mediated mRNA decay. Furthermore, Western blot analysis revealed that mutant JAK1 produced a truncated protein and a reduced expression level compared to the wild-type JAK1. In additional 12 sporadic cases with EV, we also sequenced the 25 exons and their flanking intronic sequences of *JAK1* but found no mutation. In summary, we have identified a germline splicing mutation in *JAK1*and provided supportive evidence for its causality in EV.

2924W

The molecular anatomy of multiple Noonan Syndrome *PTPN11* mutations in the testes of normal men using Deep Sequencing. *J. Eboreime, P. Calabrese, N. Arnheim.* Molecular and Computational Biology, University of Southern California, Los Angeles, CA.

Some nucleotide sites with exceptionally high *de novo* germline mutation rates cause sporadic cases of different Mendelian diseases. Noonan syndrome (NS) is one such disease characterized by short stature, heart defects, intellectual disability, craniofacial abnormalities and a predisposition to certain cancers. NS has features (RAMP) in common with related genetic diseases: 1) Recurrent mutation 1,000 fold higher than the genome-average, 2) Autosomal dominant transmission, 3) Male-biased *de novo* mutations and 4) Paternal age effect (PAE). The increasing birth rates in older couples highlights the importance of understanding the PAE mechanism. Previously, we found the most common NS *PTPN11* mutation (c. 922A>G) was not uniformly distributed in the testes of normal men as hypermutation would predict. Rather the mutations were highly clustered with almost all of them found in small regions of the testis, suggesting RAMP mutations occur rarely but, when they arise, confer a selective advantage to mutant testis stem cells over wild type ones. We speculated that the high frequency of sporadic NS cases (~1/2,000 births) might be due to a small number of disease sites that confer such an advantage leading to high numbers of sperm carrying different disease mutations. To test this idea we dissected the testes of 3 normal men into 192 pieces each, and targeted a 171 bp *PTPN11* DNA segment. For each of the 576 libraries created, we tagged each original target DNA strand with a unique identifier to lower the impact of NGS errors. We measured the mutation frequency at all the sites in the segment, 10 of which are recurrent NS mutation sites. Due to heat-induced deamination and oxidation during the early PCR cycles this assay has a C>T and G>T background frequency of 9x10-5, compared to 4x10-6 for the other four mutation types. We found that for some, but not all, of the NS sites the mutations were clustered, supporting the selective advantage hypothesis. In one testis, e. g. , there were two adjacent pieces where the c. 188A>G mutation frequencies were greater than 10-3, while at this site all the other testis pieces had frequencies near or below the background frequency. Possible explanations for NS sites not showing clustering are: these mutations do not confer a selective advantage, the advantage is too slight to be observed with this assay, or due to the stochastic process of rare mutation followed by selective growth the clusters were not present in these testes.

2925T

Identification of 4 novel mutations in a cohort of tuberous sclerosis patients from Russia. *L. I. Shagam, M. Yu. Dorofeeva, A. R. Sadykov, A. V. Polyakova, D. M. Shatalova, V. S. Sukhorukov.* Research Institute of Pediatrics at Pirogov Russian Medical University, Moscow, Russian Federation.

Tuberous sclerosis is a rare autosomal dominant disorder characterized by highly variable clinical manifestations including seizures, developmental delay, intellectual disability and autism. About 80-85% of patients are known to bear mutation in either tuberin or hamartin gene. Our research was aimed at identification of mutations that cause tuberous sclerosis in a cohort of 8 patients from Russian population. The patients were diagnosed according to (Northrup et al. , 2013). We examined tuberin and hamartin genes of the study participants for germline mutations by targeted resequencing of coding exons and adjacent 5 bp intronic regions. The targets were amplified by means Ion AmpliSeq™ technology and sequenced on the Ion Torrent Personal Genome Machine. We have identified 6 mutations with 4 of them to the best of our knowledge being first described here. Herein we provide a brief account of the clinical features accompanying the mutations in the 4 unrelated cases. Patient #1, a 3-year-old girl is clinically characterized by drug-resistant epileptic seizures, polycystic kidney disease, facial angiofibroma and multiple hypomelanotic macules. Her parents are healthy. Family trio analysis has revealed a *de novo* frameshift deletion of tuberin gene *TSC2* (NM_000548): c. 1198delT, p. (Phe400fs*25). Patient #2, a 5-year-old boy characterized by presence of hypomelanotic macules, multiple cardiac rhabdomyomas, supraventricular extrasystole, vicarious hypertrophy of the right kidney, hepatic angiomyolipoma. His mother is healthy whereas his father also suffers tuberous sclerosis with hypomelanotic macules and multiple cardiac rhabdomyomas. The patient bears a hamartin gene nonsense mutation *TSC1* (NM_000368): c. 2527C>T, p. (Gln843Ter) inherited from his father. Patient #3, a 6-year-old boy characterized by drug-resistant epilepsy, subependymal giant cell astrocytoma, developmental delay; forehead fibrous plaque, hypomelanotic macules, facial angiofibroma; cardiac rhabdomyoma, renal angiomyolipomatosis, polycystic kidney disease and retinal hamartoma. His parents are healthy. The patient bears a *de novo* tuberin gene nonsense mutation *TSC2*: c. 1073G>A, p. (Trp358Ter). Patient #4, a 14-year-old boy characterized by symptomatic focal epilepsy, intellectual disability, autism spectrum disorder. He bears a tuberin gene mutation *TSC2*: c. 4183C>T, p. (Gln1395Ter).

2926F

Mutation spectrum in a cohort of patients with *ABCA4*-related retinal dystrophies. *Y. Xie[1], W. Lee[1], S. H. Tsang[1], G. A. Fishman[2], F. T. Collison[2], R. Riveiro-Alvarez[3], C. Ayuso[3], T. Gambin[4], J. R. Lupski[4], R. Allikmets[1].* 1) Department of Ophthalmology, Columbia University, New York, NY; 2) The Pangere Center for Hereditary Retinal Diseases, The Chicago Lighthouse for People Who Are Blind or Visually Impaired, Chicago, IL; 3) Department of Genetics, Instituto de Investigacion Sanitaria-University Hospital Fundacion Jimenez Diaz (IIS-FJD), Madrid, Spain; 4) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Mutations in the *ABCA4* gene are responsible for a wide variety of retinal dystrophy phenotypes from autosomal recessive Stargardt disease (STGD1) to cone–rod dystrophy (CRD) and, in some advanced cases, generalized choriocapillaris dystrophy (GCCD) and retinitis pigmentosa (RP). Currently there are more than 800 diseases-associated *ABCA4* variants identified; the combined clinical and allelic heterogeneity has substantially complicated genetic analyses of *ABCA4*-associated retinal diseases. In a fraction of patients (10-15%) with clinical diagnosis of phenotypes compatible with *ABCA4*-associated diseases, no causal mutations were found after complete sequencing of the *ABCA4* gene, in 50 families and 31 sporadic cases. Here we describe a genetic analysis to determine the causal gene of the disease in these patients. Affected and available unaffected members from these families and sporadic cases, in total 175 individuals, were subjected to whole exome sequencing (WES). Possibly disease-associated variants were determined by filtering based on minor allele frequency and predicted pathogenicity. Variants were verified by Sanger sequencing followed by segregation analysis. The definite causal gene was identified in 30/50 families (60%). Analysis of 17/50 (34%) families resulted in several candidate genes, while no plausible candidate was found in only 3 families (6%). Of the 31 sporadic cases, 13 were solved, 9 had multiple plausible candidate genes, and 9 were not solved. The causal genes for the solved cases belong to the following categories: 1) phenotypic expansion in known retinal disease genes (11 cases; *CRB1, CRX*, etc.); 2) revised clinical diagnosis or ambiguous phenotypes (12; *RS1, PRPH2, RP1L1*, etc.); 3) new or re-assessed mutations in *ABCA4* (10); 4) new genes (3; *RDH11, RAB28*); 5) unusual clinical outcome (1; *MMACHC*); 6) other RP genes (9). The genetic causes of *ABCA4*-like phenotypes range from new genes to unusual clinical manifestations of known genes. Determining all genetic variation underlying retinal diseases is necessary for precise molecular genetic diagnosis and improved prognosis of these conditions.

2927W

Exome sequence analysis suggests genetic burden contributes to phenotypic variability and complex neuropathy. *T. Harel[1,17], C. Gonzaga-Jauregui[1,2,17], T. Gambin[1], M. Kousi[2], L. B. Griffin[3,4], L. Francescatto[2], B. Ozes[5], E. Karaca[1], S. N. Jhangiani[6], M. N. Bainbridge[6], K. S. Lawson[7], D. Pehlivan[1], Y. Okamoto[1], M. Withers[1], P. Mancias[8], A. Slavotinek[9], P. J. Reitnauer[10], M. T. Goksungur[11], M. Shy[12], J. Willer[2], B. N. Flores[3], W. Wisznewski[1], Y. Parman[11], A. Antonellis[3,13,14], D. M. Muzny[5], N. Katsanis[2], E. Battaloglu[5], E. Boerwinkle[6,7], R. A. Gibbs[1,6], J. R. Lupski[1,6,15,16], Baylor-Hopkins Center for Mendelian Genomics.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030; 2) Center for Human Disease Modeling, Duke University, Durham, NC 27710; 3) Cellular and Molecular Biology Program, Univeristy of Michigan Medical School, Ann Arbor, MI 48109; 4) Medical Scientist Training Program, Univeristy of Michigan Medical School, Ann Arbor, MI 48109; 5) Department of Molecular Biology and Genetics, Bogazici University, Istanbul, Turkey; 6) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030; 7) Human Genetics Center and Institute of Molecular Medicine, Univeristy of Texas- Houston Health Science Center, Houston, TX 77030; 8) Department of Neurology and Pediatrics, Division of Child & Adolescent Neurology, University of Texas Medical School at Houston, Houston, TX 77030; 9) Division of Genetics, Department of Pediatrics, University of California, San Francisco, CA 94158; 10) Pediatric Teaching Program, Cone Health System and UNC-Chapel Hill, Greensboro NC 27401; 11) Department of Neurology, Istanbul University, Istanbul Medical Faculty, Istanbul, Turkey; 12) Department of Neurology, Carver College of Medicine, University of Iowa, Iowa City, IA 52242; 13) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI 48109; 14) Department of Neurology, University of Michigan Medical School, Ann Arbor, MI 48109; 15) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030; 16) Texas Children's Hospital, Houston, TX 77030; 17) These authors contributed equally to this work.

Charcot-Marie-Tooth (CMT) disease is a clinically and genetically heterogeneous distal symmetric polyneuropathy, with over 70 causative genes identified to date. Whole exome sequencing (WES) of 40 individuals from 37 unrelated families with CMT-like peripheral neuropathy, in whom extensive genetic evaluation had failed to establish a molecular diagnosis, identified apparent causative mutations in ~45% (17/37) of families. Molecular diagnosis led to retrospective review and redefinition of the phenotype in eight subjects, underscoring the clinical utility of WES. Three novel candidate disease genes are proposed, functionally linked to well-established neuropathy genes and supported by *in vivo* studies. Remarkably, aggregate analysis of mutation data revealed a significantly increased number of rare variants across neuropathy associated genes in subjects versus controls, even after accounting for the apparent causative mutations. The elevated mutation burden was confirmed in a second ethnically discrete neuropathy cohort, suggesting its potential contribution to phenotypic variability. Neuropathy genes implicated by burden in families were shown to interact synergistically in a zebrafish assay, wherein pairwise co-injection of sub-effective doses of selected morpholinos yielded a neuronal phenotype beyond that expected for each morpholino alone. Our findings suggest that the combinatorial effect of rare variants contributes to disease burden and variable expressivity.

2928T

Identification of a 2nd family with a MYH14 p. Arg941Leu mutation confirms MYH14 as a gene causing Charcot-Marie-Tooth neuropathy and hearing loss. *C. S. Smith[1], G. Pfeffer[1,2], R. E. Lamont[1,3], F. P. Bernier[1,3,4], A. M. Innes[1,3,4], J. S. Parboosingh[1,3], Care4Rare Canada Consortium.* 1) Department of Medical Genetics, University of Calgary, Calgary, Alberta, Canada; 2) Department of Clinical Neurosciences, University of Calgary, Calgary, Alberta, Canada; 3) Alberta Children's Hospital Research Institute, Calgary, Alberta, Canada; 4) Clinical Genetics, Alberta Children's Hospital, Calgary, Alberta, Canada.

Mutations in myosin heavy chain 14 (*MYH14*), a non-muscle myosin, are generally associated with autosomal dominant nonsyndromic deafness (OMIM 608568). Herein, we report a three generation family where a dominant neuromuscular phenotype comprised of a CMT-like neuropathy and hearing loss segregates a heterozygous NM_001077186.1(*MYH14):* c. 2822G>T (p. Arg941Leu) variant identified by exome sequencing. A previous report associated this same variant with a complex neuromuscular phenotype consisting of neuropathy, myopathy, deafness and vocal hoarseness (PNMHH, OMIM 608568) in a large multigenerational Korean family. This suggests that the p. Arg941Leu mutation might cause an appreciably different phenotype from other reported mutations in *MYH14*. However, the reasons for these phenotypic differences have yet to be explored. Based on homology modeling of the MYH14 structure, the p. Arg941Leu mutation is predicted to interfere with the essential dimerization function of a coiled-coil domain in the tail segment of the protein. However, there are also nonsyndromic deafness mutations in the tail domain of the protein and thus, location of the variant within the broader tail domain fails to entirely explain the difference in phenotype. With this second report of a *MYH14* mutation causing a deafness-neuropathy phenotype, we suggest that this gene is one of the growing number that cause this combined phenotype. We intend to characterize the effect of this mutation on protein stability, localization and function *in vitro* using cell culture in order to explore why it causes a pleiotropic phenotype.

2929F

Mutations in phosphatidylinositol 4-kinase-beta cause autosomal-dominant non-syndromic hearing loss. *M. Gong[1], XL Su[2], J Zhang[3], S Rahman[4], F Rüschendorf[5], H Cui[2], J liang[6], L Fang[5], H Hu[7], S Froehler[9], Y Yu[5], G Patone[5], N Hubner[5], K Raile[1], M Gross[8], S Bähring[1], FC Luft[1], S Bähring[1], K Hussain[4], W Chen[9].* 1) 8 Experimental and Clinical Research Center (ECRC), a joint cooperation between the Charité Medical Faculty and the Max-Delbrueck-Center for Molecular Medicine in the Helmholtz Association (MDC), Lindenberger Weg. 80, 13125, Berlin, Germany; 2) Clinical Medicine Research Center of the Affiliated Hospital, Inner Mongolia Medical University, No 1 TongDao Northern Street, 010050, Huhhot, Inner Mongolia, China; 3) Affiliated Hospital of Guangdong Medical College, 57 Renmin Dadao Nan, 524001 Zhanjiang, China; 4) Genomic Medicine Programme, UCL Institute of Child Health and Great Ormond Street Hospital for Children, 30 Guilford Street, London WC1N 1EH, UK; 5) Max-Delbrueck-Center for Molecular Medicine (MDC), Robert-Roessle-Str. 10, 13125, Berlin, Germany; 6) Affiliated People Hospital, Inner Mongolia Medical College, Huhhot, 010050, Inner Mongolia, China; 7) Department of Human Molecular Genetics, Max Planck Institute for Molecular Genetics, Ihnestrabe 63-73, 14195 Berlin, Germany; 8) Deparment of Audiology and Phoniatrics, Charité-Universitàtsmedizin, Charitéplatz 1, 10117, Berlin, Germany; 9) Berlin Institute for Medical Systems Biology, Max-Delbrueck-Center for Molecular Medicine (MDC), Robert-Roessle-Str. 10, 13125, Berlin, Germany.

Early-onset or childhood autosomal-dominantly inherited non-syndromic sensorineural hearing impairment (ADNSSHI) is a heterogeneous genetic disorder. We combined linkage analysis and exome sequencing to identify an ADNSSHI locus (DFNA62) in a large multi-generational Chinese family. We identified a highly conserved missense mutation in the gene encoding phosphatidylinositol 4-kinase, catalytic-beta (*PI4KB)*. We next studied a small 57-member Chinese hereditary non-syndromic sensorineural hearing impairment (NSSHI) cohort from the same geographic area and found three highly conserved missense mutations in 5 patients that were absent from >1200 control Chinese chromosomes. We then observed that PI4KB protein expression in the organ of Corti of 11 and 15 week-old human fetal inner-ear sections. Transfected Hela-cell functional analyses showed cytoplasmic dilating vesicles, compared to wild-type PI4KB transfected cells. We also found that kinase activity in mutated PI4KB transfected HEK293 cells was increased compared to controls. Finally, up-regulation of Cx26 in the mutated PI4KB-transfected HEK293 cells suggested that rare alleles in *PI4KB* could be involved in the pathogenesis of ADNSSHI.

2930W

A dominant Limb Girdle Muscular Dystrophy with severe respiratory involvement; *HNRNPLD* as the new player. *S. E. Lipinski[1], M. J. Lipinski[2], R. N. Kurtzke[3], O. Goker-Alpan[1].* 1) O and O Alpan, LLC, Fairfax, VA; 2) Medstar Washington Hospital Center, Washington, DC; 3) Neurology Center of Fairfax, LTD, Fairfax, VA.

Limb-girdle muscular dystrophies are divided into autosomal dominant (LGMD 1) and autosomal recessive (LGMD 2) types. Little evidence exists about respiratory involvement in LGMD 1. More than 70% of cases of LGMD 2C–2F (sarcoglycanopathies) have respiratory muscle involvement with reduced FVC, and those with γ and α sarcoglycanopathy have severe respiratory insufficiency, with FVC reduced to <40% predicted. In LGMD 2A (calpainopathy), there is late respiratory muscle involvement with sparing of cardiac muscle. A 54 year old male was seen fordisproportionately severe respiratory muscle weakness with mild limb weakness and elevated creatinine kinase. At 52, he developed dyspnea on exertion, then at rest, with severe orthopnea and nocturnal dyspnea. Deltoid muscle biopsy showed denervation atrophy and non-specific non-inflammatory myopathic features. Exam was significant for respiratory accessory muscle use, gasping with 3 word dyspnea, mild weakness of proximal muscles and trace peripheral edema. Pulmonary function tests showed forced vital capacity (FVC) of 36%, forced expiratory volume (FEV1) of 43%, and FEV1/FVC of 106%. Sleep study demonstrated numerous oxygen desaturations, and CPAP ventilation was recommended. Chest Xray showed diaphragmatic paralysis and atelectasis. Echocardiography showed normal ventricular functions and chamber sizes with no valvular disease. Right heart catheterization was pursuedto rule out pulmonary hypertension as respiratory acidosis was noted. All pressures were normal. Whole Exome Sequencing found a candidate gene, *HNRNPLD,* as well as two genes associated with hypertrophic cardiomyopathy,*MYL3* and *ACTN2*. The patient is heterozygous for a deletion/insertion mutation in *HNRNPDL* (c. 110_113delAGCT), a deletion in the *MYL3* gene (c. 457delC), and for a variant of uncertain significance in *ACTN2* (c. 2147 C>T). The *MYL3* gene encodes the essential light chain of myosin and the *ACTN2* gene encodes alpha-actinin which anchors the sarcomeric actin filaments. Given the normal cardiac evaluation, these were concluded not to play a role in the phenotype. *HNRNPLD* mutations were reported previously in two families with Limb Girdle Muscular dystrophy. In zebrafish, this gene plays an important role in muscle development and thus, its dysfunction would be consistent with a myopathy. This is the first report relating *HNRNPLD* mutations at the heterozygote state, with an adult onset muscular dystrophy and severe respiratory muscle involvement.

2931T

Beyond the exome: Improving genetic diagnoses in Mendelian disease with whole genome and RNA sequencing. *B. B. Cummings[1,2,3], T. Tukiainen[2,3], M. Lek[2,3], F. Zhao[2,3], B. Weisburd[2,3], L. Waddell[4], A. Topf[5], S. Donkervoort[7], V. Straub[6], C. Bonnemann[7], NF. Clarke[4], ST. Cooper[4,8], DG. MacArthur[2,3].* 1) Harvard Medical School, Boston, MA, USA; 2) Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA; 4) Institute for Neuroscience and Muscle Research, The Children's Hospital at Westmead, Sydney, New South Wales, Australia; 5) Institute of Genetic Medicine, Newcastle University, United Kingdom; 6) Institute of Human Genetics, University of Newcastle upon Tyne, United Kingdom; 7) Neuromuscular and Neurogenetic Disorders of Childhood Section, National Institute of Neurological Disorders and Stroke/NIH, Bethesda, MD; 8) Discipline of Paediatrics and Child Health, University of Sydney, New South Wales, Australia.

Exome sequencing is a powerful and cost-effective tool that has become increasingly routine in Mendelian disease diagnosis; however, the current diagnostic rate for exome analysis across a variety of rare diseases is approximately 25-50%. Where exome data falls short, whole genome and RNA sequencing can offer additional insight. Whole genome sequencing provides more uniform coverage and improved identification of structural variants whereas RNA sequencing gives insight into the transcriptional landscape of the affected tissue allowing, for instance, the detection of aberrant splicing and allelic imbalance, all types of events rarely detectable from genotype data alone. Such analyses can empower molecular diagnosis by validating the transcriptional effects of variants proposed by exome data or by identifying novel variants. Here we describe an integrated approach of trio whole genome and patient muscle RNA sequencing in over 50 exome-unsolved cases with severe neuromuscular diseases. Our data consist of individuals for whom exome sequencing has prioritized variants predicted to alter gene expression or splicing as well as those for which there are no candidates from exome sequencing. We demonstrate the power of whole genome and RNA sequencing to validate candidate splice-disrupting mutations and to identify splice-altering variants in both exonic and deep intronic regions. We also apply both technologies to a set of undiagnosed Duchenne muscular dystrophy cases and demonstrate the value of both technologies for detecting and functionally validating complex structural rearrangements (such as inversions) invisible to standard diagnostic methods. Finally, we describe an analysis framework focused on the detection of transcript level changes that are unique to the patient, relative to a database drawn from over 150 skeletal muscle RNA-seq samples from the GTEx project, as well as query heterozygous predicted deleterious variants to identify evidence of allelic imbalance. Together, our results demonstrate the value of both whole-genome and tissue transcriptome data for the detection and interpretation of variants missed by standard exome-based approaches.

2932F

Deciphering of a rare autosomal recessive type of Osteogenesis Imperfecta (OI) in an Iranian family by Next Generation Sequencing (NGS). *G. Shariati[1,2], A. Saberi[1,2], M. Hamid[2,3], N. Mazaheri[2,4], J. Zeighami[2], M. Sarvari[2], H. Galehdari[2,4].* 1) Ahvaz Jundishapur University of Medical Sciences, Ahvaz, KHUZESTAN, Iran; 2) Narges Genetic diagnosis and PND Lab, Ahvaz, Iran; 3) Department of Biotechnology, Pasteur Institute, Tehran, Iran; 4) Department of Genetics, Shahid Chamran University, Ahvaz, Iran.

Introduction:Osteogenesis Imperfecta (OI) is a heterogeneous group of heritable skeletal disorder that characterized by bone fragility. Disease caused by alteration in the synthesis and folding of type I collagen and usually inherited in an autosomal dominant (AD) mode, with incidence in 6–7 per 100000. There are at least five recognized forms of Autosomal dominant OI. Type I is the mildest and most common form of OI with mild bone fragility, relatively few fractures, and minimal limb deformities. It accounts for 50 percent of the total OI population. Type II and III is the most severe type at infants and among children respectively but people with type IV, V and VI are moderate in severity nonetheless type VI is extremely rare. These types are result of mutations in the collagen encoding genes (*COL1A1* or *COL1A2*), but rarely recessive inherited OI has been also discovered . Types VII and VIII are recessively inherited types of OI that caused by mutations in two genes , *CRTAP* and *LEPRE1* respectively. Since then, another recessive form, Osteogenesis Imperfecta, Type XI [OI11 (MIM 610968)] that caused by homozygote mutation in an collagen chaperone protein, FK506-binding protein 10 gene (*FKBP10*) also has been shown. several mutations included missense and frame shift indels are identified for this type of OI. Therefore since different genes have been involved for several types of disease, Sanger sequencing is less efficient way to reach a molecular diagnosis. In order to evade these difficulties Next Generation Sequencing (NGS) technology has been installed in clinical laboratories as an ideal approach for the mutation analysis in a short time with low cost. Methods:Here we describe a large pedigree included three closely related Iranian families with OI-like bone fragility, sever to moderate bone deformity, short stature, scoliosis, joint laxity and osteopenia. For mutation analysis NGS was applied to test clinical sample from one severe patient as proband and validated by sanger sequencing. Results and Conclusions:NGS of proband sample resulted in detection of a mutation in *FKBP10* gene. This mutation was successfully validated by Sanger sequencing in other four patients and related parents. Therefore, we conclude that this new *FKBP10* mutation are a cause of recessive Osteogenesis Imperfecta in this related families and NGS provides an efficient strategy to reach to a molecular diagnosis in such inherited disorders.

2933W

Computational approach improves identification of *RPGR* ORF15 mutations for inherited retinal disease patients by Next-generation sequencing. *Z. Ge[1,2], L. Zhao[1,2,3], M. Xu[1,2], E. Jones[1,2], R. Chen[1,2,3,4,5].* 1) Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Structural and Computational Biology and Molecular Biophysics Graduate Program, Houston, TX; 4) The Verna and Marrs Mclean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX; 5) Program in Developmental Biology, Baylor College of Medicine, Houston, TX.

Alterations in *RPGR* are major cause of inherited retinal diseases including Retinitis Pigmentosa (RP) and Cone Rod Dystrophy (CRD). Previous research showed that mutations in *RPGR* account for ~70% of X-linked RP and ~ 50% of X-linked CRD cases. In addition, ~15% of simplex male RP patients carry *RPGR* mutations. In particular, *RPGR* open reading frame (ORF) 15 is a mutation hotspot with about 60% mutations identified in *RPGR* located within ORF15. However, due to highly repetitive sequence and complex arrangement within *RPGR* ORF15, it is challenging to identify mutations in this region. This can lead to false negative diagnosis for some of the RP cases. To address this issue, we developed a computational algorithm, which is optimized for identification of mutations within the *RPGR* ORF15 region. When applied to 110 RP and 58 CRD patients previously unsolved, 9 patients were found to carry frameshift mutations in *RPGR* that were missed by previous analysis method. Further evaluation and optimization of our approach is current underway by testing on a large cohort of patient samples.

2934T

Impact of whole exome sequencing on identification of disease causing mutations among Iranian patients with Autosomal Recessive Retinitis Pigmentosa. *M. Beheshtian[1], S. Saee Rad[2], H. Hashemi[3], M. Babanejad[1], M. Mohseni[1], A. Eshghabadi[3], F. Hajizadeh[3], MR. Akbari[4,5], K. Kahrizi[1], M. Riazi Esfahani[3], H. Najmabadi[1].* 1) Genetics Research Center, University of Social Welfare and Rehabilitation Scienc, Tehran, Iran; 2) Vice dean for research affairs, Alborz University of Medical Sciences, Karaj, Iran; 3) Noor Ophthalmology Research Center, Noor Eye Hospital, Tehran, Iran; 4) Women's College Research Institute, Women's College Hospital, University of Toronto, Toronto, ON, Canada; 5) Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada.

Non-syndromic autosomal recessive retinitis pigmentosa (arRP) is a genetically heterogeneous disorder which results from mutations in a large number of genes and which can lead to blindness. We aimed to determine the power of next-generation sequencing (NGS) in identification of the genes responsible for non-syndromic arRP among Iranian patients. We used whole exome sequencing (WES), followed by Sanger sequencing to identify the underlying gene mutations causing non-syndromic arRP. Our study revealed disease-causing mutations in known arRP genes for 10 of the 13 families studied (76. 9%). These included two frameshift insertion/deletion mutations in *CRB1* and *ABCA4*, one splicing mutation in *PDE6B*, four missense mutations in *RP1*, *CRB1*, *PANK2* and *IFT140*, and three stop codon mutations in *RDH12*, *PRCD*, and *C2orf71*. The three remaining families harbored no mutation in previously known RP genes. Of the 10 disease causing mutations identified among the investigated Iranian patients with non-syndromic arRP, eight variants had not been reported previously, including *CRB1* (c. 1252dup; p. Ser418Phefs*3), *ABCA4* (c. 2927del; p. Leu976Argfs*55), *PDE6B* (c. 1060-1G>T), *RP1* (c. 679T>G; p. Phe227Val), *CRB1* (c. 2711G>A; p. Cys904Tyr), *PANK2* (c. 419T>C; p. Phe140Ser), *IFT140* (c. 3827G>A; p. Gly1276Glu), and *C2orf71* (c. 712A>T; p. Lys238*). We confirmed segregation of all 10 mutations with disease phenotypes in our studied population. This study supports the genetic heterogeneity of non-syndromic arRP in Iranian patients, and provides an opportunity to show the effectiveness of WES in identification of pathogenic mutations in 77% of the tested Iranian patients with non-syndromic arRP born to consanguineous parents.

2935F

Modifiers genes in Metabolic Pathway modulate the severity of Incontinentia Pigmenti phenotype. *F. Fusco[1], M. I. Conte[1], M. Paciolla[1], A. Pescatore[1], M. F. Branda[1], M. C. Girone[1], M. B. Lioi[2], M. V. Ursini[1].* 1) IGB ABT CNR, Naples, Italy; 2) University of Basilicata, Potenza 85100, Italy.

Incontinentia pigmenti (IP, MIM308300, 1/10. 000) is a rare X-linked dominant and male-lethal multisystemic disorder whose only genetic cause is the mutation of *NEMO* gene (*IKBKG/Nuclear Factor kappaB, Essential MOdulator*). *NEMO* encodes for the regulatory subunit of the IKK complex required for the activation of the NF-kB survival pathway. The skin defects are hallmarks of IP disease and the extra cutaneous alterations (ocular, dental, hair, nail and central nervous system-CNS- defects) may occur at variable frequency. Despite a genetic homogeneity, a highly heterogeneous inter- and intrafamilial clinical presentation has been reported in IP. The severity is related to the presence of neurologic impairment (seizures, encephalopathy, encephalomyelitis, ischemic stroke) in 30% of IP patients with the same genomic alteration (*NEMOdel4-10*). To identify genetic factors that modify clinical variability in IP, we performed a high-resolution exome sequencing of three IP families-trios belonging to our IP-Genetic Biobank (IPGB), to identify genetic variants that beyond the *NEMO* common loss-of-function mutation (*NEMOdel4-10*), can contribute to the severity of the CNS phenotype. Complex filtering analysis of exome variants allowed us to identify as candidate genes, variations in proteins belonging to the one-carbon metabolism pathway. These variations are reported to produce mild/severe increase of homocysteine plasma concentration and Folic Acid and VitamimB12 deficiency in carrier patients. The in-deep analysis of clinical metabolic phenotype of sequenced probands suggested the presence of a mild deficit in cobalamin metabolic pathway. To confirm the damaging effect of homocysteine in cells of CNS we used *Nemo(-)* Neuronal Precursor Cells (NPC) from conditional *Nemo* knockout mouse brain finding that in the absence of *Nemo*, the NPC are more sensible to cell death induced by low levels homocysteine. Collectively our data suggest that genes belonging to the one-carbon metabolism represent modifier factors in IP disease able to modulate the IP phenotype severity by producing mild elevation of homocysteine levels that trigger brain injury by affecting survival of CNS cells lacking *NEMO*/*IKBKG* gene.

2936W

A novel missense mutation in *NLRC4* gene causes an autoinflammatory disorder associated with hypercytokinemia, chronic meningitis and hearing impairment. *H. Oda[1,2], K. Nakagawa[2], K. Izawa[2], T. Yasumi[2], R. Nishikomori[2], T. Heike[2], H. Kawashima[3], A. Hijikata[4], T. Shirai[4], M. Saito[5], M. Yamagishi[1], O. Ohara[1, 6].* 1) Integrative genomics, RIKEN-IMS, Yokohama, Japan; 2) Pediatric Dept. Kyoto University, Kyoto, Japan; 3) Pediatric Dept. Tokyo Medical University, Tokyo, Japan; 4) Nagahama Institute for Bio-Science and Technology, Shiga, Japan; 5) Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, Japan; 6) Kazusa DNA Research Institute.

[Background] NLR-family CARD domain-containing protein 4 (NLRC4) is one of the cytosolic receptors for bacterial components. When stimulated, NLRC4 forms a multiprotein complex called inflammasome and triggers the production of inflammatory cytokines such as IL1b and IL18. To date, three amino acid alterations of NLRC4, namely, p. Val341Ala, p. Thr337Ser, and p. His443Pro, have been reported in patients with autoinflammatory symptoms. However, the clinical courses of these patients vary substantially, in terms of the presence or absence of life-threatening systematic inflammation and/or persistent colitis. [Case presentation] The patient, 20-year old male, showed periodic episodes of fever, arthralgia, erythema and chronic meningitis (CSF cell count; 850/ul) from neonatal period. His white blood cell count is 16300/ul (neutrophil; 94%), CRP 21. 1 mg/dl, serum amyloid A 453 ug/ml, IL-1b 31 pg/ml, IL-6 16. 8 pg/ml (elevated to 1160 pg/ml during attacks), TNFa <5 pg/ml, and sTNFR1 1330 pg/ml. His intelligence quotient (IQ) was 49 (WISC-III). He showed bilateral hearing impairment (8 kHz, 50 dB). While he had suffered from multiple episodes of systemic inflammation, which sometimes necessitated steroid-pulse therapy or even hemodiafiltration, anti-IL1 (Anakinra) therapy dramatically improved his clinical course. [Methods and Results]We performed trio-based whole exome sequencing and detected c. 529A>G (p. Thr177Ala) variant in the patient, which is not reported in dbSNP, HGVD or ExAC databases. *In silico* structural analysis revealed that Threonine 177 is located close to ATP-binding site of NLRC4, which is also the case with other previously reported NLRC4 mutations. Transfection of NLRC4 T177A construct into THP1, a human monocytic cell line, induces increased cell death compared to WT-NLRC4, suggesting the pathogenicity of T177A mutation. [Conclusion]We identified a novel missense mutation of NLRC4 in a patient with hypercytokinemia and CNS symptoms. Further molecular investigation is needed to define the genotype-phenotype correlation of NLRC4-associated autoinflammatory disorders.

2937T

Changes in red blood cell membrane structure in G6PD deficiency and thalassemia: an atomic force microscopy study. *W. Jiang[1], J. TANG[1], C. R. JIANG[1], X. XIAO[2], Z. S. FANG[1], L. LI[2], L. H. HAN[1], A. Q. MEI[1], Y. FENG[1], Y. B. GUO[1], H. Y. LI[1].* 1) Department of Medical Genetics, Zhongshan School of Medicine, Sun Yat-sen Univer, Guang Zhou, China; 2) Department of Clinical Laboratory Medicine, the third affiliated Hospital of Guangzhou medical University, Guangzhou,China.

Anemia is a major source of morbidity and mortality worldwide. Anemia may be caused by a variety of pathologic conditions, including such congenital disease as thalassemia and G6PD deficiency. The hemolytic anemia in G6PD deficiency is usually triggered by oxidative stress, but the mechanism remains uncertain. Moreover, the relationship between thalassemia and RBCs membrane ultrstructure remains to be clarified. We have used atomic force microscopy (AFM) for studying changes in red blood cell membrane to provide new insights on the hemolysis mechanism of G6PD deficiency and thalassemia. G6PD activity assay and molecular genetic tests were used for molecular diagnosis, AFM for analyzing the ultra structure of RBC membranes from 33 G6PD deficient patients, five alpha-thalassemia and five beta-thalassemia individuals. We analyzed the ultrastructure of RBC membranes with or without Primaquine pretreatment and observed the protective effects of vitamin C by AFM. AFM imaging and quantitative analysis showed that G6PD deficient and thalassemias erythrocytes became heterogeneous roughness measurements of RBC membranes are increased, and different mutations may relate with roughness parameters. The width of thalassemias erythrocyte membranes is decreased, and the height increased, when compared with the control. Primaquine induces an increase roughness and height of erythrocyte membrane in normal and G6PD deficient erythrocytes, which could be prevented by vitamin C treatment in normal RBCs, but not in G6PD deficient erythrocytes. The ultra structures damages of erythrocyte membrane in G6PD deficient and thalassemia patients will give valuable information to explore the mechanism of hemolytic anemia.

2938F

From the Bcl-2-deficient mice to the human phenotype : an uncomplete demonstration requiring data sharing. *L. Duplomb-Jego[1], N. Droin[2], J. Thevenon[1], O. Bouchot[1], AL. Bruel[1], C. Thauvin-Robinet[1], E. Solary[2], L. Faivre[1].* 1) Dijon University Hospital, Dijon, France; 2) Gustave Roussy Institute, Villejuif, France.

Protooncogen BCL2 is a major repressor of apopotosis in diverse cell types. Mice knockouts for Bcl2 complete embryonic development, but present with a multisystemic progressive disorder with growth retardation and early mortality postnatally. Progressive hematological manifestation are reported with initially normal hematopoiesis and lymphocyte differentiation, evolving in massive apoptotic of thymus and spleen. Severe polycystic kidney disease characterized by dilated proximal and distal tubular segments and hyperproliferation of epithelium and interstitium lead to renal failure. bcl-2-/-mice turn gray with the second hair follicle cycle, implicating a defect in redox-regulated melanin synthesis. To date, no human disorder was demonstrated to be caused by BCL2 haploinsufficiency. A 31-year-old patient presented the unique association of polycystic kidney, lymphopenia with a personal history of T cell lymphoblastic lymphoma in early infancy, white hair since the age of 20 years. Additionally, he developed an abdominal aortic dissection, and ascending thoracic aortic aneurysm, scoliosis and learning difficulties. Only the PKD was found in the father and sister, and an inherited PKD2 mutation was evidenced. Based on the striking similarities shared by the patient and the bcl2-/- mice, we hypothesized that the patient phenotype could be due to the BCL2 haploinsufficiency. Repeated in vitro experiments performed on cultured patients lymphocytes demonstrated a severe apoptosis as compared to controls. Apoptosis phenotype could be rescued by anti-CD3 antibodies, suggesting BCL2 mediated apoptosis. In cultured lymphocytes, BCL2 expression was very low when quantified using real-time PCR and immuno-imprint. Intermediate results were found for both parents, suggesting an autosomal recessive inheritance. Direct sequencing of the exons, promotors and 3'UTR region failed in identifying any *BCL2* pathogenic variant, as well as mRNA sequencing of *BCL2*. Array-CGH was normal. Exome sequencing of the patient, the parents and sister failed in identifying a variant in a gene interacting with *BCL2*. The methylation of *BCL2* promotor was normal, and miRNA sequencing revealed an overexpression and underexpression of miRNA interacting with *BCL2*. Data sharing is required in order to identify other patients with a similar clinical or cellular phenotype. The hypothesis of the presence of a modifying factor modulating the expression of the *PKD2* variant cannot be excluded.

2939W

A potential founder variant in *RLTPR* in 3 Norwegian families with warts, molluscs and T cell dysfunction. *H. Sorte[1], L. T. Osnes[2], B. Fevang[3,4], H. C. Erichsen[5], T. G. Abrahamsen[3,5], T. Øverland[5], P. Samarakoon[1], T. Gambin[6], Z. H. C. Akdemir[6], O. K. Rødningen[1], R. Lyle[1], J. R. Lupski[6,7,8,9], A. Stray-Pedersen[6,10].* 1) Department of Medical Genetics, Oslo University Hospital and University of Oslo, Norway; 2) Department of Immunology, Oslo University Hospital, Oslo, Norway; 3) Institute of Clinical Medicine, University of Oslo, Oslo, Norway; 4) Section of Clinical Immunology and Infectious Diseases, Oslo University Hospital, Oslo, Norway; 5) Department of Pediatrics, Oslo University Hospital, Oslo, Norway; 6) Baylor-Hopkins Center for Mendelian Genomics (BHCMG) of the Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, US; 7) Human Genome Sequencing Center of Baylor College of Medicine, Houston, TX, US; 8) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, US; 9) Department of Pediatrics, Baylor College of Medicine, Houston, TX, US; 10) Norwegian National Unit for Newborn Screening, Oslo University Hospital, Oslo, Norway.

RLTPR (RGD-, leucine-rich repeat-, tropomodulin domain-, and proline-rich domain containing protein) is a protein of undetermined function. However a role in T cell activation has been suggested and *RLTPR* expression has been shown to be downregulated in psoriatic skin. The mouse protein homolog (Rltpr) is essential for costimulation of T cell activation via CD28 and the development of regulatory T cells. Four affected individuals from three different families were exome sequenced as part of a larger study for detecting genetic causes of primary immunodeficiencies. No disease-causing variants were identified in known primary immunodeficiency genes or in OMIM genes. The same homozygous missense variant was identified in all 4 patients. In each family the variant was located within a narrow region of homozygosity, representing a potential region of autozygosity. All 3 families are of Norwegian origin, but without known inter- or intra-consanguinity. All four patients had a common skin phenotype of warts, molluscs and eczema since early childhood. Gastrointestinal disease, childhood asthma, BK virus induced cystitis were observed to a variable degree. All had eosinophilia. Immuno-phenotyping demonstrated reduced number of CD4+ memory and follicular CD4+ in 3/4, and reduced regulatory T cells in 2/4 affected individuals. Normal NK cell numbers and function. Low number of IgM memory and class switched B-cells, and/or hypogammaglobulinemia were observed in 3/4 individuals. We report a novel primary immunodeficiency, and a differential molecular diagnosis to *CXCR4* (WHIM), *DOCK8*, *GATA2*, *MAGT1*, *TMC6* and *TMC8* related diseases. The specific variant may represent a Norwegian founder variant segregating on a population specific haplotype.

2940T

Mendelian Genomics approach identifies novel disease causal genes for human immunological disorders. *Y. Zhang[1,2], A. Oler[1], X. Yu[1,3], H. Matthews[1,4], M. Similuk[1], J. McElwee[5], A. Freeman[6], J. Milner[1,3], H. Su[1,2], M. Lenardo[1,4], NIAID Clinical Genomics Program.* 1) NIAID Clinical Genomics Program, National Institutes of Health, Bethesda, MD; 2) Laboratory of Host Defenses, National Institutes of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD; 3) Laboratory of Allergic Diseases, National Institutes of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD; 4) Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD; 5) Merck Research Laboratories, Merck & Co. Inc. , Boston, MA; 6) Laboratory of Clinical Infectious Diseases, National Institutes of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD.

Genomic sequencing approaches have been widely used to identify causative rare variants in families with Mendelian diseases. The discoveries of new genes and pathways for human immune disorders provide new insights into our understanding of the regulation of human immune system, as well as improving clinical diagnosis and targeted therapy for drug development. The NIAID Clinical Genomics program, in conjunction with the NIAID Primary Immune Deficiency Clinic, was set up to exploit new genomics technologies for identification of novel disease genes in patients who have a spectrum of diagnoses without known molecular etiology. The immunological disorders under our investigation include primary immunodeficiencies, immune homeostasis disorders, autoimmune conditions, and allergic disorders. We have preformed about 400 whole exome sequencing (WES) for more than 100 enrolled patients and family members. WES revealed deleterious mutations for previous established diagnosis in 15% of the sequenced patients. Most of those patients were noted to have atypical clinical features comparing to previous reported cases. Potential disease causal variants were identified in novel candidate genes in about 40% of the WES families. With functional validation and characterization, we have successfully discovered more than 10 new immunological genetic disorders, including discoveries for expanding our knowledge of previous known immunology essential genes (CTLA4, PIK3CD, STAT3 GOF etc.), discoveries of new functions of previously known immunology relevant genes (TPPII etc.) and discoveries of novel genes without previous association or involvement in immunology (MAGT1, PGM3 etc.). Those discoveries have already led to several treatment protocols applied to our patients (Magnesium for XMEN patients, N-acetylglucosamine/triacetyluridine for PGM3 patients, PIK3 inhibitor trial etc.). We are currently expanding our Mendelian genomics efforts by collaborating with investigators and clinicians across multiple countries, as well as industry. Recently, we also switched our sequencing platform to whole genome sequencing (WGS) and have performed 313 WGS from 120 families. The WGS samples include previous WES families without a clear deleterious variants candidate and new enrolled families/patients. Those data sets provide a comprehensive comparison of WES vs. WGS about efficiency and sensitivity for detecting potential disease causal mutations.

2941F

Frequency and effect of α-thalassemia mutations on the hematological phenotype in Mexican mestizo patients with β-thalassemia. *L. delC. Rizo de la Torre[1,2], V. M. Rentería López[2], B. Ibarra[1,3], F. J. Perea Díaz[1,2].* 1) Doctorado en Genética Humana, CUCS, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico; 2) División de Genética, Centro de Investigación BioMédica de Occidente (CIBO), Instituto Mexicano del Seguro Social (IMSS), Sierra Mojada No. 800, Guadalajara Jalisco, México; 3) Instituto de Genética Humana "Enrique Corona Rivera", Centro Universitario de Ciencias de la Salud (CUCS), Universidad de Guadalajara (UdG), Sierra Mojada No. 950, Guadalajara Jalisco, México.

Introduction. β-thalassemia (β-thal) is one of the most frequent monogenic disorders worldwide, it is caused by reduction (β+) or absence (β0) of β-globin chains synthesis by mutations in *HBB*, leading to microcytic hypochromic anemia. β-thal is a clinically heterogenous disease. There are other known genetic factors that modify the phenotype such as reduction of α-globin chains (-α/αα, -α/-α or --/αα), therefore co-inheritance of α-thalassemia (α-thal) is benefical for patients with β-thal. Objective. To estimate the frequency of α-thal mutations in Mexican mestizo patients with β-thal and it's effect on the hematological phenotype. Methodology. DNA of 109 β-thal heterozygote patients was studied. β-thal mutations were identified by Amplification Refractory Mutation System and DNA sequencing, α-thal mutations by Gap-PCR, besides one patient's α-thal allele was identified by MLPA. Hematological phenotype of β-thal carriers according to α-thal genotype was compared using IBM SPSS Statistics for Macintosh, Version 21. 0. Results. Four α-thal alleles (-α3. 7, --SEA, --FIL, --MEX2) and one triplication (ααα3. 7) were identified. Ten patients (9. 1%) had α-thal, the following genotypes were observed: five (4. 6%) -α3. 7/αα, two (1. 8%) --SEA/αα, one (0. 9%) -α3. 7/-α3. 7, one (0. 9%) --FIL/αα and one (0. 9%) -α3. 7/--MEX2, four other patients (3. 7%) were carriers of the ααα3. 7 allele. No α-thal mutations were found in 93 (85. 3%) patients, two (1. 8%) patients remained unidentified. Regarding β-thal, a total of 14 different *HBB* mutations were observed (eight b0 and six b+). b0 alleles were observed in 78. 9% of the patients, while 21. 1% were b+. The most frequent ones were c. 118C>T (39. 4%), c. 92+1 G>A (20. 2%) and c. 93-21 G>A (14. 7%). According to the hematological phenotype, double heterozygotes b0 and -α/-α or --/αα showed less severe phenotye than simple heterozygotes (Hb=12. 0±2. 6 *vs* 11. 6±9. 6 g/dL; MCV=65. 5±2. 1 *vs* 61. 1±7. 0 fL; HCM=21. 2±0. 4 *vs* 19. 6±2. 6 pg). Also, b+ carriers showed a slight increment on MCV (65. 5±9. 2 *vs* 63. 9±4. 3 fL) and MCH (21. 9 ± 3. 1 *vs* 20. 7 ± 1. 6 pg) when co-inherit -α/αα alleles. However, in all cases the comparative analysis did not show statistical significance ($p > 0.05$). Due to the reduced number of patients with α-thal, genotype-phenotype correlation could not be established. Conclusion. Presence of α-thal mutations is frequent in Mexican mestizo patients with β-thal. We observed no association of α-thal mutations with the hematological phenotype.

2942W

Molecular analysis of a novel mutation (CD16 GGC →GGT) in the exon 1 region of the human β-globin gene, associated with an unexpectedly severe β thalassemia intermedia phenotype. *F. Taghavifar[1], G. Shariati[2], M. Hamid[2,3], A. Metzenberg[1].* 1) Department of Biology, California State Univeristy, Northridge; 2) Narges Medical Genetics & PND Laboratory; 3) Department of Molecular Medicine, Biotechnology Research Center, Pasteur Institute of Iran.

β thalassemia is an autosomal recessive disorder, which is caused by mutations in the HBB gene that codes for β–globin chains of the hemoglobin protein. β thalassemia is associated with a reduced or absent synthesis of the β–globin chains, which resulst in insufficient hemoglobin production, mild to severe anemia and ultimately organ damage and poor overall growth of the affected individuals. A novel mutation, CD16 (GGC→GGT), in the exon 1 region of the β-globin gene has been found by DNA sequencing in a mother and daughter of an Iranian family. The father was heterozygous for a common β thalassemia allele (CD36/37 –T) and he displayed a classic minor β thalassemia phenotype. The mother was asymptomatic for β-thalassemia, but had hemotological parameters that were borderline. Their daughter, who was a compound heterozygote for both the maternal novel and the paternal mutations, showed an unexpectedly severe intermediate β thalassemia phenotype. Physical examination of the proband showed paleness and mild splenomegaly, and she had a history of transfusion dependence. Therefore, the novel CD16 mutation affected the normal expression of the β-globin gene, either alone or in combination with another mutation. To investigate the molecular basis of the observed β thalassemia phenotype in the family, we analyzed the structural and functional effects of the CD16 mutation on HBB gene expression, using bioinformatics and RT-qPCR experiment**.** Analysis showed that the C→T transition in the novel gene created a cryptic splice site, with homology to the 6 middle nucleotides of the donor consensus sequence at the exon 1- intron 1 boundary. It has been reported that this homologous sequence in exon 1 causes alternative splicing, reducing the efficiency of the normal donor splice site. RT-qPCR analysis revealed that the novel mutation was associated with a down-regulation of the HBB gene. This study describes the first report of a novel mutation in the HBB gene, which creates a cyptic splice site, down-regulating the β-globin gene expression. We are currently investigating the effects of this mutation on the express of other globin and non-globin genes, using the HiSeq 2000 system from Illumina. These data will provide important information for genetic counceling centers working on the premarital and prenatal diagnosis of thalassemia.

2943T

Uncovering the etiology of ICF syndrome: new disease genes and functional insight in the molecular processes underlying the immunodeficiency. *S. M. van der Maarel[1], P. E. Thijssen[1], Y. Ito[2], G. Grillo[3], G. Velasco[3], H. Nitta[2], M. Unoki[2], M. Yoshihara[4], M. Suyama[4], R. J. L. F. Lemmers[1], J. C. de Greef[1], A. Gennery[5], P. Picco[6], B. Kloeckener-Gruissem[7], T. Güngör[8], I. Reisli[9], C. Picard[10], K. Kebaili[11], B. Roquelaure[12], T. Iwai[13], I. Kondo[14], T. Kubota[15], M. M. van Ostaijen-Ten Dam[16], M. J. D. van Tol[16], C. Weemaes[17], C. Francastel[3], H. Sasaki[2], H. van Attikum[1].* 1) Human Genetics, Leiden University Medical Center, Leiden, Netherlands; 2) Division of Epigenomics and Development, Department of Molecular Genetics, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan; 3) CNRS UMR7216, Epigenetics and Cell Fate, Université Paris Diderot, Sorbonne Paris Cité, Paris, France; 4) Division of Bioinformatics, Department of Multi-scale Research Center for Medical Science, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan; 5) Department of Paediatric Immunology, Newcastle Upon Tyne Hospital, Newcastle Upon Tyne, United Kingdom; 6) Division of Pediatrics and Pediatric Reumathology, G. Gaslini Scientific Institute, Genova, Italy; 7) Institute of Medical Molecular Genetics, University of Zurich, Schlieren, Switzerland; 8) Department of Oncology, University Children's Hospital, Zurich, Switzerland; 9) Department of Pediatric Immunology and Allergy, Necmettin Erbakan University, Meram Medical Faculty, Konya, Turkey; 10) Centre de Référence Déficits Immunitaires Héréditaires, Paris, France; 11) Centre de Référence Déficits Immunitaires Héréditaires, Institut d'Hématologie et d'Oncologie Pédiatrique, Lyon, France; 12) Service d'hépato-gastro-entérologie et nutrition, endocrinologie et néphrologie Pédiatriques, Hôpital de la Timone, Marseille, France; 13) Department of Pediatric Hematology and Oncology, Shikoku Medical Center for Children and adults, Kagawa, Japan; 14) Department of Pediatrics, Ooida Hospital, Kochi, Japan; 15) Department of Epigenetic Medicine, Faculty of Medicine, Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi, Yamanashi, Japan; 16) Laboratory for Immunology, Department of Pediatrics, Leiden University Medical Center, Leiden, The Netherlands; 17) Department of Pediatric Infectious Diseases and Immunology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands.

The immunodeficiency, centromeric instability, and facial anomalies (ICF) syndrome is clinically characterized by recurrent and often fatal infections as a consequence of hypo- or a-gammaglobulinemia in the presence of B cells. ICF syndrome has been considered an exemplar epigenetic disease with half of the patients carrying loss of function mutations in the *de novo* DNA methyltransferase 3B gene (*DNMT3B,* referred to as ICF1), resulting in genome-wide hypomethylation, predominantly affecting the pericentromeric satellite repeats of chromosomes 1, 9 and 16. ICF syndrome is a genetically heterogeneous disorder with another 30% of patients carrying mutations in the zinc finger and BTB domain containing 24 (*ZBTB24*) gene (ICF2). Here we report mutations in the cell division cycle associated 7 (*CDCA7*) and the helicase, lymphoid-specific (*HELLS*) genes in the majority of unexplained ICF cases. By transient depletion of each of the individual ICF genes in mouse embryonic fibroblasts we demonstrate that all four ICF genes have a role in the establishment and/or maintenance of CpG methylation at pericentromeric satellite repeats. While these and other studies have firmly established a causal relationship between the ICF gene mutations and the major epigenetic changes observed in ICF syndrome, the molecular mechanisms underlying the immune defect have thus far remained largely elusive. We now demonstrate that ZBTB24 has a role in the non-homologous end-joining-dependent repair of DNA double strand breaks during class switch recombination, explaining the immunodeficiency in ICF2 patients. This, for the first time, provides a mechanistic link between the chromatin modifiers mutated in ICF syndrome and the associated immunodeficiency phenotype.

2944F

Carriers of the complex allele *HFE* c. [187C>G;340+4T>C] have increased risk of iron overload in São Miguel Island population (Azores, Portugal). *L. Mota-Vieira[1,2,3], C. T. Gomes[1], L. De Fez[1], S. Bulhões[1], M. J. Brilhante[1], T. Pereirinha[1], R. Cabral[1], C. C. Branco[1,2,3].* 1) Molecular Genetics and Pathology Unit, Hosp of Divino Espirito Santo of Ponta Delgada, Azores Islands, Portugal; 2) Instituto Gulbenkian de Ciência; Oeiras, Portugal; 3) BioISI - Biosystems & Integrative Sciences Institute, Faculty of Sciences, University of Lisboa, Lisboa, Portugal.

Iron overload is associated with acquired and genetic conditions, being the most common the hereditary hemochromatosis (HH) type I, caused by *HFE* mutations. Despite recent advances, a better understanding of molecular basis of iron overload is needed, to improve patient's outcome through early diagnosis and treatment. Here, we performed a hospital-based case-control study of 41 patients from São Miguel Island (Azores, Portugal), 6 belonging to a family with HH type-I pseudodominant inheritance, and 35 unrelated individuals fulfilling the biochemical criteria of iron overload compatible with HH type-I. To that end, we genotyped the most common *HFE* mutations –c. 845G>A [p. Cys282Tyr], c. 187C>G [p. His63Asp], and c. 193A>T [p. Ser65Cys] –, by TaqMan real-time PCR. Results revealed that 10 (24. 4%) patients were homozygous for the c. 845A mutation, and 5 (12. 2%) were c. 845A/c. 187G compound heterozygous. Sanger sequencing of *HFE* exons, where these mutations are located (exons 2, 4, and their intron-flanking regions), did not reveal other mutation but allowed the identification of c. 340+4C [IVS2+4C] splice variant in 26 (74. 3%) patients. In terms of functionality, the c. 340+4C may generate alternative splicing by *HFE* exon 2 skipping, and, consequently, a protein missing the α1 domain, which is essential for the interaction of HFE with the transferrin receptor 1. Moreover, we found that 21 (60. 0%) patients have at least two *HFE* mutations/variants, placed *in-cis* or *in-trans* position, that may explain the iron overload. Further, we investigated these mutations/variants configuration with iron overload, by determining the *HFE* haplotypes and genotypic profiles. Results evidenced that carriers of the *HFE*-c. 187G allele also carry the -c. 340+4C, suggesting that they are *in-cis* configuration. This data was corroborated by the association analysis, where carriers of the complex allele *HFE*-c. [187C>G;340+4T>C] have an increased risk for iron overload (RR=2. 08, 95% CI=1. 40-2. 94, p<0. 001). Therefore, homozygous for this complex allele are at risk to have iron overload, since they will produce two altered proteins – the p. 63Asp [c. 187G] protein, and the protein lacking 88 amino acids encoded by exon 2. In summary, we provide evidence that the complex allele *HFE*-c. [187C>G;340+4T>C] has a role, as genetic predisposition factor, on iron overload in the São Miguel population (Azores, Portugal). Independent replication studies in other populations are needed to confirm this association.

2945W

Coordinated regulation of splicing events within *SLC12A3* and implications to Gitelman Syndrome. *C. Mercado, X. Wang, Y. C. Chang.* School of Medicine, University of Maryland, Baltimore, MD 21201, USA.

The majority of human multi-exon genes undergo alternative splicing (AS), resulting in structurally different transcripts and functionally distinct protein isoforms that are differentially expressed across tissues. *SLC12A3* encodes Na+/Cl- cotransporter (NCC), which is involved in salt reabsorption along the distal nephron. Inactivating mutations within *SLC12A3* cause Gitelman syndrome (GS), an autosomal recessive disorder characterized by chronic salt wasting and hypotension. Previous findings indicate that AS of *SLC12A3* exons 7 and 8 leads to transcript isoforms missing either exon 7 (Dexon7) or both exons (Dexon7-8). Isoform Dexon7 contains a premature stop codon in exon 9 and, presumably, activates nonsense-mediated decay, leading to little or no NCC synthesis. However, -exon7-8 results in an in-frame deletion of a highly conserved extracellular domain of NCC (p. (Ala285_Lys365del)). All 3 isoforms are found in normal human blood cells and kidneys, with Dexon7 and Dexon7-8 much less abundant in both tissues. Importantly, a splicing mutation (IVS7-1G>A; 965_976>CCGAAAATTTT) leads to a higher abundance of isoform Dexon7-8 and is thought to be pathogenic in GS patients. However, the exact mechanism as to how this mutation, changes in isoform abundances, or the small NCC protein isoform leads to GS is not known. In this study, we performed RNA-seq to analyze the human and mouse kidney transcriptomes. Several human- and kidney-specific *SLC12A3* AS isoforms were detected and confirmed by conventional RT-PCR. In addition to the known AS involving exons 7 and 8, we identified a novel AS event that occurs within the 3'UTR. Splicing in this region results in a shorter 3'UTR lacking an internal 171 bp segment. Using isoform-specific primers, we established that isoform Dexon7-8 with the larger 3'UTR is abundantly expressed in kidney. In contrast, isoform Dexon7-8 with the smaller 3'UTR is not detected, suggesting that AS in these 2 regions is coordinately regulated. Interestingly, the spliced 3'UTR segment harbors a putative microRNA-binding site for hsa-miR-4722-3p. Therefore, isoform Dexon7-8 might be targeted for miRNA-mediated mRNA degradation or translational suppression, resulting in little or no NCC with p. (Ala285_Lys365del). In summary, microRNA binding to isoform Dexon7-8 can potentially suppress the translation of an abnormal NCC protein. In this case, the GS-causing splicing mutation is functionally a null mutation due to coordinated splicing at 2 sites.

2946T

Novel defects of *FOG2/ZFPM2* including 3. 2 Mb deletion lead to isolated 46, XY disorders of sex development and 46, XX ovarian dysfunction. *D. Baetens[1], N. Reynaert[3], F. de Zegher[3], E. De Baere[1], M. Cools[2].* 1) Center for Medical Genetics, Ghent University, Ghent, Belgium; 2) Department of Pediatrics, Division of Pediatric Endocrinology, Ghent University Hospital and Ghent University, Ghent, Belgium; 3) Department of Pediatric Endocrinology, University Hospitals Leuven, Leuven, Belgium.

Background: *FOG2* (Friend of GATA2; also known as *ZFPM2*) mutations and copy number variants (CNVs) have been associated with congenital cardiac anomalies and diaphragmatic hernias for more than a decade. In addition, *FOG2* has recently been implicated as novel disease gene for isolated 46, XY disorders of sex development (DSD). Here, we report the first CNV encompassing *FOG2* and other genes in a patient with isolated 46, XY DSD and his mother with 46, XX ovarian dysfunction. In addition, a 46, XY DSD cohort was screened for coding *FOG2* mutations. Results: ArrayCGH revealed a heterozygous 3. 2 Mb deletion (chr8: 104348705-107509840) encompassing *FOG2* and nine other genes in a neonate with penoscrotal hypospadias and bilateral cryptorchidism. This deletion is of maternal origin. Endocrinological investigations showed low testosterone, anti-Møllerian hormone (AMH) and inhibine B levels, suggestive of gonadal dysgenesis. Cardiac ultrasound was normal. The mother had a low serum AMH value (0. 57 µg/l [0. 9-9. 5]), suggestive of ovarian dysfunction. We screened a cohort of 58 patients with 46, XY DSD for coding *FOG2* mutations or CNVs using a flexible targeted next-generation sequencing approach (MiSeq) and qPCR (Roche) respectively. This uncovered a novel heterozygous missense variant in a single case: c. 2432T>G p. (Val811Gly). This missense variant is predicted to be deleterious by several algorithms and affects an amino acid that is highly conserved up to chicken. The physicochemical distance between Val and Gly is 109, suggesting a possible effect on protein function. The variant is not known as a SNP and is absent in the ExAc browser. Segregation testing and functional characterization of this variant is ongoing. Conclusion: There are several reports of *FOG2* mutations and translocations disrupting *FOG2* in patients with congenital cardiac defects with or without gonadal abnormalities. Recently missense mutations of *FOG2* were identified in two patients with isolated 46, XY DSD without cardiac involvement. To the best of our knowledge, we present here the first genomic rearrangement encompassing*FOG2* as the underlying cause of isolated 46, XY DSD and of 46, XX ovarian dysfunction. This is emphasizing its dosage sensitivity and its strict tissue-specific regulation. Finally, we identified a novel missense variant in 46, XY DSD. Taken together, we have expanded the molecular spectrum of *FOG2* mutations in isolated 46, XY DSD and in 46, XX ovarian dysfunction.

2947F

Phenotypic interaction between GH-releasing hormone (GHRH) receptor mutation detected by exome sequencing and non-classical congenital adrenal hyperplasia. *I. Arnohold[1], Q. MA[2], F. CORREA[1], M. FRANÇA[1], Q. FANG[2], T. BACHEGA[1], A. RODRIGUES[1], J. LI[2], B. MENDONCA[1], A. JORGE[1], S. CAMPER[2], L. CARVALHO[1].* 1) Hospital das Clinicas da Faculdade de Medicina da Universidade de Sao Paulo, Sao Paulo, SP, Brazil; 2) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI.

Background: Isolated Growth Hormone Deficiency (IGHD) is the most common pituitary hormone deficiency and can result from genetic or acquired causes. Patients with GH deficiency usually have a retarded bone age. **Objective:** To diagnose the genetic cause of IGHD and clarify the unusual clinical presentation of advanced bone age in one patient born to consanguineous parents. **Patient:** A 7. 5 year-old boy presented with severe short stature (height 102. 5 cm, SDS-3. 7), high-pitched voice, blue sclerae and prominent forehead. Genital examination was prepubertal and bone age was 6 years. Parents were second-degree cousins. Clonidine and combined pituitary stimulation test (insulin, TRH and GnRH), revealed GH deficiency (GH peak 0. 6 ng/ml) and was compatible with partial ACTH deficiency (cortisol peak=16. 1 mcg/dl). Pituitary MRI was normal. The patient was successfully treated with rGH (33 mcg/kg/day) with a first year growth velocity of 11. 7 cm/yr. Surprisingly, at 10. 8 years his bone age had advanced to 13 years, but physical exam, LH and testosterone levels remained prepubertal. **Methods:** Sanger-sequencing of *GH1*, *GHRH, GHRH receptor* and *CYP21A2* (using an active gene-specific exon 3 primer) followed by whole exome sequencing (WES). MLPA of the *CYP21* unit. **Results:** An ACTH stimulation test disclosed a basal 17-hydroxyprogesterone of 9. 4 ng/ml and peak of 52 ng/ml indicating non-classical adrenal hyperplasia due to 21-hydroxylase deficiency and treatment with cortisone acetate was added. Sanger sequencing of *CYP21A2* revealed homozygous c. 844G>T, p. Val282Leu mutation, but *GH1*, *GHRH* and *GHRH receptor* were read as normal. MLPA indicated a *CYP21A1P* and *C4B* duplication, but no major deletion or conversion. WES revealed a homozygous mutation in *GHRHR* (c. T431A, p. Leu144His) which had been overlooked by initial Sanger sequencing but could be identified when the electropherogram was re-examined. This recurrent mutation had been previously identified in Brazilian, Spanish and US patients with IGHD. Interestingly, WES indicated the CYP21A2 p. Val282Leu mutation in heterozygous state, probably due to the presence of a normal sequence in at least one copy of the pseudogene. **Conclusion:** A patient born to consanguineous parents presented with two recessive endocrine diseases modifying the phenotype: GH deficiency due to a *GHRH receptor* mutation diagnosed by WES and non-classical CAH diagnosed correctly only by Sanger sequencing specific to the active gene *CYP21A2*.

2948W

Phenotypic spectrum from cerebral midline defects to hypopituitarism: examples from HESX1 mutations. *Q. Fang[1], Q. Ma[1], A. Sadeghi-Nejad[2], I. J. P. Arnhold[3], F. Correa[3], A. B. Ozel[1], J. Z. Li[1], S. A. Camper[1], L. R. S. Carvalho[3].* 1) University of Michigan, Department of Human Genetics, Ann Arbor, MI; 2) Tufts University School of Medicine, Boston, MA; 3) University of Sao Paulo, Clinicas Hospital, Brazil.

Congenital hypopituitarism occurs approximately 1/4000 births and can be life threatening if not identified and treated effectively. About a dozen different genetic causes have been reported, but most of the patients have no molecular diagnosis. Mutations in the transcription factors POU1F1 and PROP1 are typically nonsyndromic and associated with recessive inheritance, while mutations in several other causal genes present as dominant with incomplete penetrance with variable clinical features, sometimes including craniofacial and eye defects. HESX1 is a member of the *paired* (*prd*) class of homeodomain proteins and functions as a transcriptional repressor. Mutations in *HESX1* can be dominant or recessive and cause septo-optic dysplasia (SOD), combined pituitary hormone deficiency (CPHD) or isolated growth hormone deficiency (IGHD). There are no obvious genotype-phenotype correlations. The diverse effects on head development are not surprising because *Hesx1* is expressed in embryonic stem cells, the anterior midline during gastrulation, the ventral prosencephalon, the ventral diencephalon and the pituitary primordium. Here we report mutations in *HESX1* found by exome sequencing that explain hypopituitarism in two unrelated patients. The familial pituitary aplasia case is a compound heterozygote for missense mutations in the homeodomain: p. R159W/p. R160H. Cell culture studies reveal that these mutations diminish the repressive activity of HESX1 on PROP1. In previous reports homozygosity for p. R160H caused CPHD, not aplasia, while homozygosity for p. R160C caused SOD. We also identified a homozygous p. I26T missense mutation in the second case from consanguineous parents with evolving CPHD and eutopic posterior pituitary. A patient with the same mutation was previously reported to have evolving CPHD with ectopic posterior pituitary. These cases illustrate the variable features that can result from loss of function mutations in the same gene. We propose that the severity of the clinical presentation is related to the genetic load of deleterious variants in other genes that contribute to head development. We are testing this idea by exome-sequencing and analyzing more patients.

2949T

Germline mutations identification in familial isolated prolactinoma through whole-exome sequencing. *F. Melo[1], R. CARDENAS[2], L. BASTOS-RODRIGUES[3], P. COUTO[1], A. BALE[4], J. NG[5], T. CURRAN[5], P. DE MARCO[1], S. PENA[2], E. FRIEDMAN[6], L. DE MARCO[1].* 1) Department of Surgery, Universidade Federal de Minas Gerais, Belo Horizonte; 2) Departments of Biochemistry, Universidade Federal de Minas Gerais, Belo Horizonte; 3) Department of Basic Sciences, Universidade Federal de Juiz de Fora, Brazil; 4) Department of Genetics, Yale University School of Medicine, New Haven, USA; 5) Cell & Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia; 6) The Susanne Levy Gertner Oncogenetics Unit, Chaim Sheba Medical Center, Tel-Hashomer, Israel.

Introduction: Prolactinomas are the most common type of pituitary tumors, being associated with hyperprolactinemia that can lead to hypogonadism and mass-effect related symptoms. Rarely, prolactinoma is diagnosed in a familial context either as part of Multiple Endocrine Neoplasia Type I syndrome (MEN 1) or as an independent entity, that can be referred to as Familial Isolated Prolactin-secreting Pituitary Adenoma. Germline mutations in *MEN1, AIP* and *PRLR* genes have been identified in a subset of prolactinoma-affected families as disease causative mutations. However, in the majority of familial prolactinoma cases the causal gene mutation remains unknown. In order to identify potential isolated Prolactin-secreting pituitary adenoma candidate genes, DNA from a Brazilian affected family were subjected to whole exome sequencing (WES). **Materials and methods**: The family is comprised of ten siblings (ages ranging from 44 to 64 years old) and three were diagnosed with hyperprolactinemia and isolated Prolactin-secreting pituitary adenomas. The mother and seven siblings were clinically asymptomatic. The father died of unrelated disease. **Results**: Direct DNA sequencing of *MEN1*, *AIP* and *PRLR* genes revealed no mutations. Genomic DNA from all affected siblings and their mother were subjected to whole-exome (WES) capturing and sequencing using the Roche NimbleGen V2 chip and the Illumina HiSeq2000 sequencing platform. Two different analytical approaches to identify rare genetic determinants were used: *Ingenuity® Variant AnalysisTM* and *Mendel, MD* (unpublished). Variant calling from WES of the genotyped DNA resulted in 19,134 common single nucleotide variations (SNV) in 8,729 genes. The mean base call quality was 1546 and average read depth was 82. For the filtering approach, we excluded variants with a minor allele frequency > 0. 01 in public datasets and identified a remaining set of functional variants, predicted to be deleterious using Sift, PolyPhen, Condel and CADD. **Conclusion**: A final list of 12 genes was generated and the mutations are being validated by Sanger sequencing approach. This work gives a new insight into unveiling FIPA-causing mutations for those patients that are negative for *MEN1*, *AIP* and *PRLR* gene molecular alterations.

2950F

Examining the genetics of Hereditary Pancreatitis (HP) in adult patients with Chronic Pancreatitis. *H. Azimi[1,3,4,5], S. Ghavimi[7], J. Brown[2], J. Sleza[6], S. Emamian[8].* 1) PsychoGenome, Ottawa, Ontario, Canada; 2) Toronto University School of Public Health, Toronto, Ontario, Canada; 3) All Saints School of Medicine, Dominica; 4) Carleton University, Ottawa, Ontario, Canada; 5) Shahid Beheshti University School of Medicine, Tehran, Iran; 6) Ottawa University, Ottawa, Ontario, Canada; 7) Howard University Hospital, Dept. of Medicine, Washington, DC, USA; 8) Indiana University, Bloomington, Indiana, USA.

Objective: To study the genetics of hereditary pancreatitis (HP) in adult patients with Chronic pancreatitis. Methods: This outpatient study was conducted on 100 Patients who came to the clinic of Tajrish's Hospital and wore diagnosed with Chronic Pancreatitis. We examined the patients and their parents for Hereditary Pancreatitis. HP defined as those with trypsinogen gene (PRSS1) mutation on the long arm of chromosome seven (7q35). DNA was extracted from peripheral blood leukocytes, and exons 2 and 3 of the gene were individually amplified by polymerase chain reaction and sequenced. Results: Of 100 adult patients with a pancreatitis, 28 (28%) were diagnosed with adult onset of Hereditary pancreaitis. From the 28 patient with positive mutation on the trypsinogen (PRSS1) gene, 13 of them had one parent with mutation of the trypsinogen gene (PRSS1). The route of inheritance is Autosomal Dominant. From the 13 adult patients who had parents with the mutations of the trypsinogen gene (PRSS1), 11 (85%) recalled of their babies or young children, with having been diagnosed with epigastric pain, severe vomiting and nausea both after birth and during early childhood. From the 28 patients who wore diagnosed with having the mutation and (HP), 6 (21%) indicated that one of their parents has been deceased or currently diagnosed with Pancreatic Cancer. From the 28 patients with (HP) and gene mutation 8 (29%) also had Type 1 Diabetes Mellitus. Conclusions: Adult patients born to mothers or fathers with pancreatitis have an increase chance of having hereditary pancreatitis. Also babies born to parents whom are diagnosed with adult onset of hereditary pancreatitis have a high chance of having hereditary pancreatitis. It is clear that hereditary pancreatitis is autosomal dominant. The natural history of HP follows a similar pattern to alcohol-associated chronic pancreatitis, but there are important differences, for example, HP has an earlier age of onset of pancreatitis although malabsorption and diabetes mellitus occur at a later stage in the disease. Families are defined as having (HP) if the phenotype is consistent with highly penetrant autosomal dominant inheritance. We can conclude that majority of patients that develop hereditary pancreatitis in early childhood or adulthood had a parent with hereditary pancreatic disorder if diagnosed or pancreatic cancer in older age. Also HP patients should avoid environmental factors that cause pancreatic cancer.

2951W

Quantitative missense mutation profiling points to the root of phenotypic variability in Autosomal Dominant Polycystic Kidney Disease (ADPKD). *V. G. Gainullin[1], B. M. Paul[1], S. L. R. Klein[2], C. M. Heyer[1], P. C. Harris[1].* 1) Division of Nephrology and Hypertension, Mayo Clinic, Rochester, Minnesota, USA; 2) Macalester College, St. Paul, Minnesota, USA.

As whole exome sequencing is increasingly used to resolve monogenic and complex diseases, determining the significance of missense variants is of great importance. High-throughput cellular analyses of protein stability, folding, trafficking and protein-protein interactions are pivotal for determining likely pathogenicity and even to estimate penetrance. Here we investigate the pathogenic significance of 58 unique missense variants of uncertain significance to *PKD1* and *PKD2,* the genes causing the common nephropathy, autosomal dominant polycystic kidney disease (ADPKD). ADPKD generally results in renal failure in late middle age, but the phenotype is highly variable ranging from in utero onset to only a few cysts in old age, with allelic effects implicated. The two proteins encoded by the ADPKD genes, polycystin-1 (PC1) and PC2 interact while folding in the ER, and PC1 maturation and cell surface localization requires the ER chaperone function of PC2. Although it is known that truncating mutations to either gene result in ER-retained protein, the effect of the many missense mutations has not been quantified. Here, we use a quantitative flow-cytometry-based approach in combination with biochemical characterization of glycosylation and processing to analyze maturation and surface delivery of mutant polycystins. Computing the cumulative pathogenicity scores we found that 9% of *PKD1* and 53% of *PKD2* missense mutations fully disrupted protein folding, 45% of *PKD1* and 29% of *PKD2* were incompletely penetrant and 45% of *PKD1* and 18% of *PKD2* did not disrupt protein folding or localization. Furthermore, 83% of the *PKD1* and 85% of *PKD2* folding mutants were temperature-sensitive: rescued to wild-type status with 30°C incubation, suggesting molecular chaperone potential in treating some but not all patients inheriting missense mutations. Hence, this quantitative analysis can gauge relative penetrance of single alleles as well as allele combinations in ADPKD patients. Taken together, our comprehensive analyses show the potential of high-throughput screening methods to evaluate and estimate the penetrance of pathogenic missense variants.

2952T

Enzymatic properties of mouse chitotriosidase expressed in *Escherichia coli*. *M. Kimura, K. Ishikawa, K. Sekine, S. Yoshikawa, S. Wakita, A. Sato, A. Kashimura, M. Sakaguchi, Y. Sugahara, F. Oyama.* Department of Chemistry and Life Science, Kogakuin University, Hachioji, Tokyo, Japan.

Chitotriosidase (Chit1) was the first mammalian chitinase to be purified and cloned. Chit1 have attracted considerable attention due to their increased expression in individuals with different pathological conditions. Chit1 rise to Gaucher disease, chronic obstructive pulmonary disease (Chronic Obstructive Pulmonary Disease, COPD), Alzheimer's disease, in smokers, and Niemann-Pick disease and atherosclerotic tissue. These data strongly suggest that Chit1 plays important roles in many pathophysiological conditions. However, the contribution of Chit1 to the pathophysiology of these diseases remains to be determined. To understand its pathophysiological roles, we cloned mouse Chit1 cDNA into the vector pEZZ18 for intracellular expression in *E. coli* as a fusion protein of Protein A-Chit1-V5-His. The recombinant protein showed a robust peak of activity with a maximum activity at pH 5. 0, where the optimal temperature was 50°C. This recombinant protein can be used to elucidate detailed biomedical functions of the mouse Chit1.

2953F

Expression and characterization of catalytic domain of mouse AMCase in *Escherichia coli* . *F. Oyama, A. Kashimura, M. Kimura, K. Okawa, H. Suzuki, A. Ukita, S. Wakita, K. Okazaki, M. Ohno, M. Sakaguchi, Y. Sugahara.* Dept Chemistry and Life Science, Kogakuin Univ, Hachioji, Tokyo, Japan.

Acidic mammalian chitinase (AMCase) has been implicated in asthma and other allergic inflammations, and food processing. Mouse AMCase is composed of an N-terminal catalytic domain (CatD) and a C-terminal chitin-binding domain (CBD). Here, we report the expression of the CatD of mouse AMCase as a recombinant fusion protein with Protein A and V5-His in *E. coli* and its functional properties. The chitinolytic activity of the recombinant CatD against 4-nitrophenyl *N, N'*-diacetyl-β-D-chitobioside was equivalent to that of the full-length AMCase with regard to its specific activity, pH and temperature optima as well as pH and thermal stabilities. CatD bound to chitin beads; cleaved the hexamer of N-acetylglucosamine, colloidal and crystalline chitin as well as shrimp shell; and released primarily N, N'-diacetylchitobiose fragments at pH 2. 0. Thus, the CatD of mouse AMCase can recognize chitin substrates and degrade them in the absence of CBD.

2954W

A partially inactivating mutation in the sodium-dependent lysophosphatidylcholine transporter MFSD2A causes a non-lethal microcephaly syndrome. *V. Alakbarzade[1, 2], A. Hameed[3], D. Q. Y Quek[4], B. A Chioza[1], E. L Baple[1,5,6], A Cazenave-Gassiot[7], L. N Nguyen[4], M. R Wenk[7], A. Q Ahmad[8,9], A Sreekantan-Nair[1], M. N Weedon[1], P Rich[1,11], T. T Warner[2], D. L Silver[4], A. H Crosby[1].* 1) Institute of Biomedical and Clinical Science, University of Exeter Medical School, RILD Wellcome Wolfson Centre, Exeter, UK; 2) Reta Lila Weston Institute of Neurological Studies, Department of Molecular Neurosciences, University College London Institute of Neurology, London, UK; 3) Institute of Biomedical and Genetic Engineering (IBGE), Islamabad, Pakistan; 4) Signature Research Program in Cardiovascular and Metabolic Disorders, Duke-National University of Singapore Graduate Medical School, Singapore; 5) Human Genetics and Genomic Medicine, Faculty of Medicine, University of Southampton, Southampton, UK; 6) Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton, UK; 7) Life Sciences Institute, National University of Singapore, Singapore; 8) Department of Physical Medicine and Rehabilitation, Indiana University-Purdue University Indianapolis (IUPUI), Indianapolis, Indiana, USA; 9) Rehabilitation Hospital Indiana, Indianapolis, Indiana, USA; 10) Department of Neuroradiology, St. George's Hospital, London, UK; 11) Southwest Thames Regional Genetics Service, St George's Healthcare National Health Service (NHS) Trust, London, UK.

The major pathway by which the brain obtains essential omega-3 fatty acids such as docosahexaenoic acid (DHA) from the circulation is through a sodium-dependent lysophosphatidylcholine (LPC) transporter (MFSD2A), expressed in the endothelium of the blood-brain barrier, which is essential for normal brain growth and cognitive function. We investigated an extensive Pakistani pedigree with multiple interlinking nuclear families with individuals affected by an autosomal recessive progressive neurological condition involving microcephaly, intellectual disability, spasticity, absent speech and dysmorphic features. A whole genome SNP scan of family members identified a single notable homozygous region peculiar to all affected family members on chromosome 1p34. 2. Whole exome sequencing identified single likely deleterious sequence variant located within the disease locus, in *MFSD2A* (chr1:40433304C>T), affecting a highly conserved residue (p. Ser339Leu). The variant cosegregated with the disease phenotype and was absent in online genomic databases and regional controls. Our subsequent studies determined that the p. Ser339Leu alteration does not affect protein or cell surface expression of MFSD2A but rather significantly reduces, although not completely abolishes, transporter activity. Notably, affected individuals displayed significantly increased plasma concentrations of LPCs containing mono- and polyunsaturated fatty acyl chains, indicative of reduced brain uptake, confirming the specificity of MFSD2A for LPCs having mono- and polyunsaturated fatty acyl chains. Together, these findings indicate an essential role for LPCs in human brain development and function and provide the first description of disease associated with aberrant brain LPC transport in humans.

2955T

Transforming growth factor beta receptor stability and signaling in LTBP4-related cutis laxa. *Z. Urban[1], J. W. Huang[2], C. K. Chiang[2], E. C. Lawrence[1], B. Dabovic[3], C. T. Su[1].* 1) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Division of Nephrology, Department of Internal Medicine, National Taiwan University Medical College and Hospital, Taipei, Taiwan; 3) Department of Cell Biology, New York University School of Medicine, New York, NY.

Latent transforming growth factor beta binding protein 4 (LTBP4) mutations cause autosomal recessive cutis laxa associated with severe pulmonary, gastrointestinal and urinary involvement. To elucidate the molecular mechanisms of LTBP4-related cutis laxa, we investigated the consequences of LTBP4 loss on transforming growth factor beta (TGFß) signaling. *LTBP4* mutant human dermal fibroblasts showed elevated extracellular TGFß activity, however, downstream signaling molecules of the TGFß pathway, including pSMAD2 and pERK, were reduced. In addition, TGFß receptors 1 and 2 (TGFBR1 and TGFBR2) were diminished. TGFb1 supplementation led to an initial rapid increase followed by a sustained deficit in receptor abundance and activity. TGFBR1 was co-localized with lysosomes, and inhibitors of TGFBR1 kinase activity, endocytosis or lysosomes, normalized the levels of TGFBR1 and TGFBR2 in mutant cells. Co-immunoprecipitation demonstrated a molecular interaction between TGFBR2 and LTBP4. TGFb receptor abundance and signaling was reduced by LTBP4 depletion in normal cells and treatment with recombinant LTBP4 enhanced these measures in mutant cells. TGFb receptor levels and signaling were reduced, but normalized upon TGFBR1 kinase inhibition in Ltbp4 knockout mice. Thus, LTBP4 is required for TGFb receptor stability both *in vitro* and *in vivo*. TGFBR1 kinase inhibition stabilizes the receptors and normalizes signaling in *LTBP4* mutant cells and animals, suggesting a possible treatment approach for LTBP4-related cutis laxa.

2956F

The spectrum of *NOTCH1* mutations in Adams-Oliver syndrome highlights an important role for the ligand-binding domain in disease etiology. *L. Southgate[1,2], M. Sukalo[3], A. S. Karountzos[4], E. J. Taylor[4], W. Wuyts[5], M. Zenker[3], R. D. Machado[4], R. C. Trembath[1].* 1) Barts & The London School of Medicine & Dentistry, Queen Mary University of London, Charterhouse Square, London, United Kingdom; 2) Division of Genetics and Molecular Medicine, King's College London, Guy's Hospital, London, United Kingdom; 3) Institute of Human Genetics, Otto-von-Guericke-Universitét Magdeburg, University Hospital Magdeburg, Magdeburg, Germany; 4) School of Life Sciences, University of Lincoln, Brayford Pool, Lincoln, United Kingdom; 5) Department of Medical Genetics, University and University Hospital of Antwerp, Edegem, Belgium.

Adams-Oliver syndrome (AOS) is a rare developmental disorder characterized by terminal transverse limb defects and scalp aplasia cutis congenita (ACC). Notable cardiac and/or neurological involvement is also apparent in a proportion of cases. Recent studies exploring the genetic basis of AOS have highlighted a key role for the Notch signaling pathway, in particular the NOTCH1 receptor. The purpose of this study was to examine the mutation spectrum and functional impact of identified *NOTCH1* mutations. We screened a cohort of 74 unrelated subjects diagnosed with either AOS or isolated ACC. Using whole-exome (n=21) or Sanger (n=53) sequencing, we identified 12 *NOTCH1* mutations, accounting for 16% of our cohort. Taken together with previous reports, there are now 31 independent *NOTCH1* mutations described in AOS. Of these, the majority (n=19; 61%) are missense mutations, approximately one third of which are cysteine substitutions, anticipated to disrupt disulfide bonds critical for maintaining protein structure. The remainder predominantly comprise frameshift (n=6; 19%) or nonsense (n=4; 13%) mutations, with only a single splice variant and one large deletion reported to date. Of note, 19% of all described mutations are located within the ligand-binding domain, defined by EGF-like repeat domains 11-13, which only represents 4. 5% of the full-length NOTCH1 receptor. When extended to the flanking domains EGF10-14 (7. 5% of the protein), previously reported to stabilize the ligand-receptor interaction, the proportion of AOS-related mutations rises to 32%, highlighting a potentially key role for the NOTCH1 ligand-binding domain in the etiology of this disorder. Indeed, we have previously demonstrated by real-time PCR that AOS mutations within this region of the receptor lead to dysregulated Notch signaling, with a significant reduction of *HEY1* and *HES1* transcript levels. To further examine the functional impact of missense mutations within the ligand-binding domain of NOTCH1, we have generated mutagenized constructs for use in X-ray crystallography studies. Bioinformatic predictions indicate that a significant proportion of AOS-related mutations will disrupt key amino acids, leading to a destabilization of protein structure and calcium binding. Experimental analysis of this critical domain is now required to aid interpretation of the pathogenicity of identified variants and to further develop our understanding of the cellular processes predisposing to this condition.

2957W

Mutations in human homologue of chicken *talpid-3* gene (*KIAA0586*) cause a hybrid ciliopathy with overlapping features of Joubert and Jeune syndromes. *M. Gunay-Aygun[1,7], T Vilboux[1,3], J Stephen[1], D Maglic[1], L Mian[1], D Konzman[1], J Guo[2], D Yildirimli[1], J Bryant[1], R Fischer[1], WM Zein[4], J Snow[5], M Vemulapalli[6], A Young[6], J Mullikin[6], C Toro[3], BD Solomon[2,7], J Niederhuber[3], WA Gahl[1,2,7], MC Malicdan[1,2].* 1) MGB, NIH/NHGRI, Bethesda, MD; 2) NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, 20892 Maryland, USA; 3) Inova Translational Medicine Institute, Falls Church, 22042 Virginia, USA; 4) Ophthalmic Genetics & Visual Function Branch, National Eye Institute, National Institutes of Health, Bethesda, 20892 Maryland, USA; 5) Office of the Clinical Director, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, 20892 Maryland, USA; 6) NIH Intramural Sequencing Center (NISC), National Human Genome Research Institute, National Institutes of Health, Bethesda, 20892 Maryland, USA; 7) Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, 20892 Maryland USA.

Statement of Purpose: In chicken, loss of *talpid-3* results in nonfunctional cilia and short-rib polydactyly syndrome. This phenotype is caused by a frameshift mutation in the chicken ortholog of the human *KIAA0586* gene, which encodes a novel coiled-coil domain protein essential for primary ciliogenesis, suggesting that *KIAA0586* can be associated with ciliopathy in humans. **Methods:** In our patients with ciliopathy (enrolled in NIH protocol "Clinical and Molecular Investigations into Ciliopathies", www. clinicaltrials. gov: NCT00068224, approved by the Institutional Review Board of the National Human Genome Research Institute), we performed whole exome sequencing on DNA from affected individuals and their parents. We analyzed gene expression on fibroblast cell line, and determined the effect of gene mutation on ciliogenesis in cells derived from patients. **Results:**We identified biallelic null mutations in the human talpid-3 ortholog, *KIAA0586*, in six children with findings of overlapping ciliopathies. Jeune syndrome, also known as asphyxiating thoracic dystrophy, is a skeletal ciliopathy characterized by a small thorax due to short ribs, disproportionate short stature due to short long bones, polydactyly, and renal, hepatic and retinal involvement. Joubert syndrome is a ciliopathy characterized by distinctive midbrain and cerebellar malformations that result in the "molar tooth sign" on axial brain images as well as hypotonia and developmental delay. Fibroblasts cultured from one of the Jeune-Joubert syndrome patients exhibited more severe cilia defects than fibroblasts from patients with only Joubert syndrome; this difference was reflected in *KIAA0586* RNA expression levels. Rescue of the cilia defect with full-length wild type *KIAA0586* indicated a causal link between cilia formation and KIAA0586 function**Conclusions**: Our results show that null allele mutations in *KIAA0586* lead to Joubert syndrome with or without Jeune asphyxiating thoracic dystrophy. Furthermore, our results confirm that KIAA0586/Talpid3 is essential in cilia formation in humans, expand the *KIAA0586* phenotype to include features of Jeune syndrome, and provide a pathogenetic connection between Joubert and Jeune syndromes, based upon aberrant ciliogenesis.

2958T

N-linked glycosylation in renal and cardiac diseases: the role of ALG13 short isoform 2. *T. Esposito[1], G. De Stefano[1], I. Di Lorenzo[1], F. Napolitano[1], A. M. Saleem[2], F. Gianfrancesco[1].* 1) Institute of Genetics and Biophysics "ABT" National Research Council of Italy, N, Naples, Italy, Naples, Italy, Italy; 2) Academic Renal Unit, School of Clinical Sciences, University of Bristol, Bristol, UK.

N-glycosylation, a post-translational modification process, has an important role in protein folding, intracellular trafficking and membrane targeting, as well as in the regulation of protein function. Mutations affecting N-linked glycosylation sites of some podocyte (e. g. , Nephrin and TRPC6) and cardiomyocyte (e. g. ,TRPM4b) proteins have a very important role in focal segmental glomerulosclerosis (FSGS) and progressive cardiac conduction defect (PCCD) respectively. Recently, we identified two missense mutations (p. R113W and p. T141L) in two novel X-linked genes (*NXF5* and *ALG13*) associated with FSGS and PCCD in an extended Australian pedigree. *ALG13* codifies for a glycosyltransferase that catalyses the second sugar addition of the highly conserved oligosaccharide precursor in the endoplasmic reticulum (ER). The most conserved isoforms of *ALG13* gene are the long isoform 1 (1137 aa) and the short isoform 2 (165 aa, ALG13-is2). Mutations, which occur in the C-terminal region of the long isoform, cause congenital disorders of glycosylation type I, epileptic encephalopathies and nonsyndromic intellectual disability. We demonstrate that the mutation c. 421-422 AC>TT [p. T141L], identified in the FSGS/PCCD patients, alters the C-terminal alpha-helix structure of the short isoform 2 (ALG13-is2), a region highly conserved in mouse (100% identity at nucleotide level) which is important for the localization of the protein in the ER. ALG13-is2 shows a specific high level of expression in human and mouse podocytes and cardiomyocytes which increases during differentiation, both at mRNA and protein level. Knock-down of the ALG13-is2, in human podocyte cells at 12 days of differentiation, results in the alteration of the N-glycosylation pattern of Nephrin, as demonstrated by the presence of an additional immune-staining band of about 130 kD in western blot, perturbed organization of the actin of the cytoskeleton and altered localization of Nephrin on the cellular membrane observed by immunofluorescence. We also demonstrate that the altered pattern of glycosylation induces over-expression of Bip and Calreticulin, suggesting ER stress. These results provide preliminary evidence that ALG13-is2 could be an important player in renal and cardiac defects.

2959F

ASXL3 regulates H2A deubiqutination and gene transcription in Bainbridge-Ropers syndrome. *S. Bielas[1], A. Srivastava[1], R. KC[1], Y. C. Tsan[1], A. M. Chinnaiyan[2,3], D. Martin[1,4], S. Bielas[1].* 1) Department of Human Genetics, University of Michigan Medical School, Ann Arbor, MI; 2) Howard Hughes Medical Institute, Michigan Center for Translational Pathology, University of Michigan Medical School, Ann Arbor, MI; 3) Department of Pathology, University of Michigan Medical School, Ann Arbor, MI; 4) Department of Pediatrics, University of Michigan Medical School, Ann Arbor, MI.

Bainbridge-Ropers Syndrome (BRS) is a phenotypically and genetically heterogeneous disorder characterized by failure to thrive, global developmental delay, primary microcephaly, intellectual disability, dysmorphic features and feeding problems. We identified a novel *de novo* nonsense mutation c. 1448dupT, T484NfsX5 in *ASXL3 (additional-sex combs like 3)* in a BRS patient. *De novo ASXL3* nonsense mutations described for BRS localize within a ~700 nucleotide interval of exon 11, coined the "mutational cluster region". To investigate the impact of *ASXL3* mutations on histone modifications and transcription regulation, we established an *ASXL3* patient fibroblast cell line. A stably expressed truncated protein is not observed in *ASXL3* patient cells, implicating haploinsufficiency in the pathogenicity of this dominant disorder. Similar to ASXL1 and ASXL2, we found that ASXL3 interacts with BAP1, the ligase that removes the mono-ubiquitin from Histone 2A (H2AK119Ub1) as part of the Polycomb repressive deubiquitination (PR-DUB) complex. Consistent with this interaction, we found H2AK119Ub1 is increased 5-fold in *ASXL3* patient fibroblasts. This was not accompanied by a change in the level of histone H3 trimethylation at lysine 27 (H3K27me3). To investigate the impact of this finding on transcriptional regulation, we performed RNA-sequencing on *ASXL3* patient fibroblasts and two unrelated fibroblast controls. Out of 700 differentially expressed genes, 35% were upregulated and 65% were downregulated in *ASXL3* patient fibroblasts. Among these, Polycomb repressive complex 1 (PRC1) target genes were significantly misregulated. These findings indicate that *ASXL3*, as part of the PR-DUB complex, removes ubiquitin from H2AK119Ub1, an important repressive histone post-translational modification generated by PRC1. Covalent histone modifications are an important layer of transcriptional regulation that occur at the level of chromatin structure. While histone ubiquitination was discovered more than three decades ago, its functions in transcriptional regulation and epigenetic repression are still less well understood than other histone modifications. This study suggests a functionally important role for dynamic H2AK119Ub1 remodeling in transcriptional regulation and the pathophysiology of BRS.

2960W

Identification of the Molecular Pathways Affected by Mutations in the Lysine Acetyltransferase, KAT6A and associations with phenotypic variability. *V. A. Arboleda[1], H. L. Lee[1], Z. C. Chen[2], H. A. Amartino[3], K. L. Low[4], R. N. Newbery-Ecob[4], S. N. Nelson[1,2].* 1) Pathology and Laboratory Medicine, UCLA, Los Angeles , CA; 2) Department of Human Genetics, David Geffen School of Medicine, UCLA; 3) Department of Pediatric Neurology, Hospital Universitario Austral, Buenos Aires, Argentina; 4) Department of Clinical Genetics, Bristol Royal Hospital for Children.

Histone acetyltransferase and deactylase enzymes affect fundamental cellular processes including the cell-cycle, cell differentiation, metabolism, and apoptosis. Nonsense mutations in *KAT6A*, a lysine acetyltransferase, result in global developmental delay, microcephaly and craniofacial dysmorphism. We recently reported the first description of *de novo* heterozygous nonsense mutations in *KAT6A* (a. k. a. *MOZ, MYST3*) that were identified by clinical exome sequencing (CES) in 4 families. Similar mutations have now been identified in more than 20 patients. From our initial report, the core features are consistently observed, however there is a larger phenotypic variability that is now apparent in regard to cardiac, gastrointestinal, and hematologic manifestations. For example, while the majority of reported patients have structurally normal brains, one patient was found to have dilated perivascular Virchow Robin spaces. Other patients have variable allergies, neutropenia, cardiac anomalies, or neurosensory deficits. In order to better understand the mechanism of mutated KAT6A, we have cultured dermal fibroblasts from three individuals with *KAT6A* mutations. Basal levels of *KAT6A* mRNA and protein were low in untreated fibroblasts. However, with Adriamycin treatment, immunofluorescence shows a nuclear localization. We performed RNA-seq analysis under resting and Adriamycin-treated conditions to stimulate nuclear KAT6A expression. In the untreated cells, there were minimal differences between control and *KAT6A*-mutant fibroblasts. However, upon stimulation with adriamycin, we observed significant differences in 300 300 genes between control and *KAT6A* mutated cells. Surprisingly, *KAT6A* mRNA levels were markedly decreased in all treated cells, suggesting that nuclear up-regulation of KAT6A is through post-translational regulation. Clinical exome sequencing is now a primary mode for diagnosis of children with syndromic congenital conditions, such as *KAT6A*. Since there is a growing body of collected phenotypic information, there is an opportunity to identify strong modifier alleles, which may contribute to the variability in phenotype. Additionally, KAT6A acts indirectly on gene regulation through alterations in histone acetylation which may have varied effects. Overlay of the expression patterns perturbed by KAT6A mutations with inherited DNA variants in those genes will provide insights to assess how *KAT6A* mutations affect downstream molecular pathways.

2961T

Autosomal Recessive Ehlers Danlos Syndrome in Three Patients with Congenital Adrenal Hyperplasia due to 21-hydroxylase deficiency. *A. Perritt[1], R. Morissette[1], W. Chen[2], V. Sachdev[3], H. Hannoush[3], A. Arai[3], D. P. Merke[1,4].* 1) NIH Clinical Center, Bethesda, MD; 2) PreventionGenetics, Marshfield, WI; 3) The Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD; 4) National Heart, Lung and Blood Institute, Bethesda, MD.

Background: Congenital Adrenal Hyperplasia (CAH) due to 21-hydroxylase (21-OH) deficiency is an autosomal recessive endocrine disorder resulting in cortisol deficiency and androgen excess. *CYP21A2*, the gene encoding 21-OH, is located on the short arm of chromosome 6 and is flanked by *TNXB*, which encodes the extracellular matrix (ECM) protein tenascin-X (TNX). Approximately 30% of *CYP21A2* mutations are deletions. Both *CYP21A2* and *TNXB* have highly homologous pseudogenes, and during meiosis unequal crossover can occur resulting in either a *CYP21A1P*/*CYP21A2* or *TNXA*/*TNXB* chimera. We found that ~10% of CAH patients have a TNX chimera, termed CAH-X syndrome, which leads to Ehlers Danlos syndrome (EDS) phenotypes. We have identified two different autosomal dominant *TNXA*/*TNXB* chimeras. CAH-X CH1 is characterized by a 120 bp deletion in exon 35 in *TNXB* leading to TNX haploinsufficiency. CAH-X CH2 is characterized by the c. 1274C>G (C4058W) variant in exon 40 leading to a dominant negative effect. Three variants have been shown to travel with this mutation, two of which are predicted to be damaging. We have now identified three families with autosomal recessive CAH-X; therefore, we sought to further define the TNX genetic and protein status in these novel CAH-X patients. **Methods:** Dermal fibroblasts and tissue from patients and controls were used for wound healing assays, Western blots, and immunohistochemistry experiments. Southern blotting, multiplex ligation-dependent probe amplification, Sanger and next-generation sequencing were used to study genetic defects. **Results:** We identified three patients with autosomal recessive (AR) CAH-X with a more severe phenotype than autosomal dominant (AD) CAH-X patients. Each proband had unique *TNXB* mutations on both alleles. The AR patients have more severe EDS phenotypes than AD patients, including higher Beighton scores, joint and skin laxity, congenital heart defects, and multiple subluxations. Growth curves showed that AR patients have slower cell growth than AD and CAH control patients. TNX expression was unchanged in AD and AR patients. Immunohistochemical experiments testing collagens and other ECM markers are underway. **Conclusion:** Autosomal recessive CAH-X presents with more severe phenotypes than autosomal dominant. Slower growth rate suggests reduced cell proliferation in AR CAH-X patients. Investigations into the three variant cluster and the impact of mutations on both alleles on TNX's function in the ECM are underway.

2962F

Learning pathophysiology of Alport syndrome from RNA-seq of podocytes differentiated from amniotic-fluid derived patient progenitors. *A. M. Pinto[1,2], S. Da Sacco[3], C. Fallerini[1], S. Furini[4], I. Meloni[1], M. Baldassarri[1,2], E. Frullanti[1], F. Ariani[1], L. Perin[3], A. Renieri[1,2].* 1) Medical Genetics, University of Siena, Siena, Italy; 2) Genetica Medica, Azienda Ospedaliera Universitaria Senese, Siena, Italy; 3) Saban Research Institute, Children's Hospital University of Southern California, Los Angeles; 4) Department of Medical Biotechnologies, University of Siena, Siena, Italy.

Alport syndrome is a clinically heterogeneous nephropathy caused by mutations in collagen IV genes. It is characterized by haematuria, proteinuria, progressive renal failure and ultrastructural lesions of the glomerular basement membrane (GBM), such as thickening, thinning and lamellation often associated with sensorineural deafness and ocular lesions. The role of podocytes' abnormalities in the disease onset and in its progression hasn't been well characterized yet and a transcriptome profiling of Alport derived-cell lines has never been previously reported. We performed a trascriptome profiling on RNA from amniotic fluid-derived Alport renal progenitors and progenitors differentiated into podocyte like-cells using our established protocol to better understand the pathogenic mechanism of the disease and to shed light on the ultrastructural GBM lesions, a hallmark of this disorder. The corresponding renal progenitors and differentiated podocytes cell lines from amniotic fluid of healthy fetus were used as control. RNA-seq data were analyzed for expression levels'changes, splicing pattern changes and differential polyadenylation sites usage. Expression of several transcripts, markers of podocytes' differentiation, such as COL4 genes, drastically increases during the differentiation process in both control and Alport podocyte like-cells. Expression levels'analysis detected 413 and 49 genes down-regulated in Alport progenitors and podocyte like-cells, respectively. The same analysis identified 396 and 76 genes up-regulated in the two cell lines compared to the corresponding controls. Importantly, a gene ontology on the genes down-regulated revealed, in both cell lines, a statistically significant enrichment for transcripts encoding for proteins expressed in the extracellular compartment and implied in cell adhesion. This finding may explain the massive disruption of the glomerular filtration barrier and the GBM typical Alport lesions. Notably, in both Alport progenitors and podocyte like-cell we noticed a consistent up-regulation of IL-32 (21,2 fold change, P=0. 00025) expressed at the same level in control progenitors and podocytes. These data were confirmed by qRT-PCR. IL-32 induces IL-6 and IL-8 expression, both found upregulated in Alport cell lines supporting an implication of this pathway in the disease pathogenesis. IL-32 binds to PR-3, a serin-proteinase, which degrades extracellular matrix proteins such as elastin, fibronectin and collagen IV. Thus, we might hypothesize that IL-32 up-regulation could contribute to the GBM lesions. Although preliminary, our findings open up the possibility of new therapeutic strategies for Alport disease based on the usage of IL-32 inhibitors, already in phase of development for autoimmune disorders.

2963W

Genotype-dependent alterations in TGFβ signaling in *ELN*-related cutis laxa. *S. Alkan, Z. URBAN.* HUMAN GENETICS, UNIVERSITY OF PITTSBURGH, PITTSBURGH, PA.

Mutations in the elastin protein (*ELN*) cause autosomal dominant cutis laxa characterized by loose, inelastic and redundant skin that get worse by age. Previous studies indicate elevated canonical (SMAD-dependent), but unaffected non-canonical (SMAD- independent) TGFb signaling. Our goal was to evaluate the impact of *ELN* mutations on cellular signaling by TGFb, a growth factor known to be involved in a range of connective tissue diseases. Dermal fibroblasts from four patients with *ELN* mutations in exon 34 or exon 30 and four age and sex-matched controls were used in immunoblotting to assay for key components of the transforming growth factor beta (TGFb) pathway under baseline conditions as well as following TGFb1 or TGFb1 with TGFb receptor 1 (TGFBR1) inhibitor treatment. Under baseline conditions (serum and serum-free), SMAD2 phosphorylation was increased in exon 30 mutant cells. No significant difference was observed in pSMAD2 and pERK levels following TGFb1 and TGFb1 with TGFBR1 inhibitor treatments. The abundance of pERK was decreased in mutant cells at baseline (serum-free), but no significant difference was observed for this molecule between mutant cells and controls under any other treatment conditions. TGFBR1 protein levels were increased in *ELN* mutant cells at baseline conditions and remained increased relative to the controls upon TGFb supplementation. No significant difference was observed in the TGBFR1 levels upon TGFBR1 inhibitor treatment. On the other hand, no significant difference was observed in the levels of pSMAD2, SMAD2, ERK and pERK levels in any conditions in exon 34-mutant cells. Our results showed elevated canonical, but unaffected non-canonical TGFb signaling despite unaltered extracellular TGFb activity in exon 30-mutant cells, whereas cells with *ELN* mutations in exon 34 had normal TGFb signaling. Elevated TGFb signaling in exon 30 mutant cells was associated with increased levels of TGFBR1 but not TGFBR2. Our results indicate mutation-specific TGFb signaling changes in *ELN*-related cutis laxa patients, which may contribute to disease severity.

2964T

Knockout of *RP2* decreases GRK1 and rod transducin subunits and leads to photoreceptor degeneration in zebrafish. *M. Liu[1], F. Liu[1], S. Yu[1], J. Chen[1], X. Shu[2], J. Zou[3].* 1) Department of Genetics and Developmental Biology, Huazhong University of Science and Technolog, Wuhan, Hubei, China; 2) Department of Life Sciences, Glasgow Caledonian University; 3) Institute of Translational Medicine, Zhejiang University.

Retinitis pigmentosa (RP) affects about 1. 8 million individuals worldwide. X-linked retinitis pigmentosa (XLRP) is one of the most severe forms of RP. Nearly 85% of XLRP cases are caused by mutations in the X-linked retinitis pigmentosa 2 (*RP2*) and *RPGR*. RP2 has been considered to be a GTPase activator protein for ARL3 and to play a role in the traffic of ciliary proteins. The mechanism of how *RP2* mutations cause retinitis pigmentosa is still unclear. In this study, we generated an *RP2* knockout zebrafish line using TALEN technology. Progressive retinal degeneration could be observed in the *RP2* knockout zebrafish. The degeneration of rods' outer segments is predominant, followed by the degeneration of cones' outer segments. These phenotypes are similar to the characteristics of RP2 patients, and also partly consistent with the phenotypes of *RP2* knockout mice and morpholino mediated *RP2* knockdown zebrafish. For the first time, we found that *RP2* deletion leads to decreased protein levels and abnormal retinal localizations of GRK1 and rod transducin subunits (GNAT1 and GNB1) in zebrafish. Furthermore, the distribution of the total farnesylated proteins in zebrafish retina is also affected by *RP2* ablation. These molecular alterations observed in the *RP2* knockout zebrafish might probably be responsible for the gradual loss of the photoreceptors' outer segments. Our work identified the progression of retinal degeneration in *RP2* knockout zebrafish, provided a foundation for revealing the pathogenesis of retinitis pigmentosa caused by *RP2* mutations, and would help to develop potential therapeutics against retinitis pigmentosa in further studies.

2965F

The p. E229K variant in *CYP1B1*: mutation or polymorphism? *G. Chavarria-Soley.* Biology, University of Costa Rica, San José, Costa Rica.

The *CYP1B1* gene is mutated in over 80% percent of familial cases of primary congenital glaucoma (PCG; MIM 231300), which shows autosomal recessive inheritance. Around 150 missense and nonsense substitutions have been reported for *CYP1B1*. While some of these variants are either clearly disease causing mutations of clearly frequent polymorphisms, there are some with an unclear clinical significance. The p. E229K variant falls into the last category. It has been routinely reported in the literature as a mutation in *CYP1B1*, and has been found in several different ethnic gropus. The goal of the present study is to analyze all evidence available in an effort to clarify the role of p. E229K. As a first step, a PubMed search was performed in order to find all references to the variant in the literature. When family data was available, it was determined whether the variant segregates with the disease, and whether it is found in a homozygous or compound heterozygous form. All functional tests and models found in the literature for this variant were analyzed. Next, in order to determine the degree of conservation of the 229 position, a multiple sequence alignment was performed for the protein sequence of *CYP1B1* in 12 different animals. A prediction of the functional effect of the p. E229K missense variant was obtained with six different prediction tools. Finally, data for the frequency of the variant was obtained from the ExAC Browser. After analyzing the literature, no example of clear segregation of the variant with autosomal recessive PCG could be found. A structural analysis from the literature predicts a deleterious effect for the variant, while a functional study where mutated *CYP1B1* was expressed in yeast found no change in activity of the enzyme, but a reduced abundance. In the multiple sequence alignment, the Glu in the 229 position of *CYP1B1* is conserved in mammals, but not in the bird or fish included. Four out of the six tools for the prediction of functional effects, predicted p. E229K to be a neutral variant. In the data from the ExAC browser, an overall frequency of 1,4% was reported for all populations. Interestingly, in the South Asian population the variant has a frequency of 5,6%, and 35 individuals (out of 7245) are homozygous. Taken together, the different lines of evidence do not support the idea of p. E229K as a disease causing mutation in PCG. At most, the variant could be considered as a hypomorphic allele.

2966W

The miR-34c is upregulated in α1-antitrypsin deficient livers, induces β-catenin activation and impairs liver zonation. *P. Piccolo[1], A. Barbato[1], P. Annunziata[1], S. Attanasio[1], B. Granese[1], N. Pastore[1], L. R. Soria[1], C. Mueller[2], J. Teckman[3], N. Brunetti-Pierri[1,4].* 1) Telethon Institute of Genetics and Medicine, Pozzuoli, Italy; 2) Boston University Center for Regenerative Medicine (CReM), Boston, MA, USA; 3) Department of Pediatrics, Saint Louis University School of Medicine, Cardinal Glennon Children's Medical Center, Saint Louis, MO, USA; 4) Department of Translational Medicine, Federico II University, Naples, Italy.

Deficiency of α1 anti-trypsin (AAT) is the most common genetic cause of liver disease in children and liver transplantation is currently the only available treatment for severe cases. The vast majority of patients with AAT deficiency harbor a single amino acid substitution which results in protein misfolding, polymerization, and accumulation within the endoplasmic reticulum of hepatocytes. Intracellular retention of aberrantly folded mutant AAT in the liver results in hepatitis, cirrhosis, and hepatocellular carcinoma. PiZ transgenic mice expressing mutant human AAT recapitulate the features of the hepatic disease. On PiZ livers, we performed next generation sequencing of miRNAs which revealed 70 miRNAs differentially expressed between controls and PiZ mice. Among the dysregulated miRNAs, we focused on miR-34c that was increased in both liver and plasma of PiZ mice and its levels correlated with the abundance of hepatic mutant AAT. Importantly, miR-34c was also increased in livers of human patients with advanced hepatic disease. In both human (Huh-7) and mouse (Hepa1-6) hepatocytes, transfection of miR-34c down-regulated expression of murine AAT genes and HNF4A, a master regulator of hepatocyte spatial organization of metabolic pathways along porto-central axis of liver lobules, known as liver zonation. Interestingly, PiZ mice showed down-regulation of HNF4A and increased levels of active β-catenin, the other major determinant of liver zonation. PiZ mice exhibited impairment of liver zonation, as shown by alteration of glutamine synthase expression pattern, down-regulation of urea cycle enzymes, and defective ammonia metabolism. Injection of an AAV2/8 vector expressing miR-34c in wild-type C57BL6 mice resulted in down-regulation of murine AAT gene expression, activation of β-catenin, and impairment of liver zonation. In conclusion, we identified miR-34c as a potential novel biomarker of liver disease induced by AAT deficiency and a novel regulator of liver zonation.

2967T

Unique case of congenital diarrheal disorder with very short transit associated compound heterozygous variants of unknown significance in a gene related to myosin stability and function, *UNC45A*. *A. Bourchany[1,2], J. THEVENON[2,3], R. MAUDINAS[1], E. SAVAJOLS[4], F. HUET[1], A. BRUEL[2], N. GIGOT[2], Y. DUFFOURD[2], J. ST-ONGE[2], JB. RIVIERE[2,3,5], L. FAIVRE[2,3].* 1) Service de Pédiatrie 1 Hôpital d'Enfants CHU Dijon, et Université de Bourgogne, Dijon, France; 2) Equipe d'Accueil 4271, Génétique des Anomalies du Développement, Université de Bourgogne, Dijon, France; 3) Centre de Génétique et Centre de Référence Anomalies du Développement et Syndromes Malformatifs, FHU TRANSLAD, Hôpital d'Enfants, CHU Dijon et Université de Bourgogne, Dijon, France; 4) Département de Pédiatrie 2, Hôpital d'Enfants, CHU Dijon et Université de Bourgogne, Dijon, France; 5) Laboratoire de Génétique Moléculaire, FHU-TRANSLAD, Plateau technique de Biologie, CHU, Dijon, France.

Congenital diarrheal disorders (CDD), a group of rare inherited enteropathies with early-onset in life, lead to a diagnostic challenge for clinicians, geneticists and pathologists. They are responsible for chronic diarrhea of variable prognosis and of sufficient severity to require a parenteral nutrition support, even occasionally intestinal transplantation. The molecular bases of CDDs have been predominantly identified in the last years. Several mechanisms of the diarrhea linked to different type of intestinal cells failure have been described but none was related to intestinal muscle cells. Cells involved in diarrheal disorders pathophysiology are enterocytes, enteroendocrine or immune cells. Classifications retain defects in digestion, absorption and transport of nutrients and electrolytes ; disorders of enterocyte differentiation and polarization ; defects of enteroendocrine cell differentiation ; and disregulation of the intestinal immune response as potential 4 groups of mechanisms. We herein report a 3 year-old girl presenting with severe intractable diarrhea diagnosed the 4th day of life, characterized by a very short transit, and still requiring exclusive parenteral nutrition. Repeated intestinal biopsies did not show any villous atrophy. *MYO5B* and *EpCam* sequencing did not evidence any pathogenic variant. Severe bilateral deafness secondarily appeared. Exome sequencing identified compound heterozygotes variants in *UNC45A*, 1 *de novo* missense (p. Val423Asp) and one inherited stop-gained mutation(p. Arg262*). The Hsp90 co-chaperone protein Unc45 isoform a is widely expressed, present in intestinal tissue, and shows selectivity for the smooth muscle myosin motor domain. Indeed, *UNC45A* may be necessary for proper folding of myosin. Smooth muscle cells take action in intestinal wall during both patterns of intestinal peristaltism. Data-sharing will be needed in order to confirm or not the implication of *UNC45A* in this rare phenotype. Exome sequencing also revealed 2 *MYO15A* mutations that may explain the patient deafness.

2968F

The role of Filamin A in gastrointestinal developmental defects. *M. M. Alves[1], D. Halim[1], H. van der Linden[1], A. S. Brooks[1], A. J. Burns[1,2], R. M. W. Hofstra[1,2].* 1) Dept Clinical Genetics, Erasmus University Medical Center, Rotterdam, Netherlands; 2) Birth Defects Research Centre, UCL Institute of Child Health, London, UK.

Filamin A (*FLNA*) encodes a cytoskeletal protein that regulates cell shape by cross-linking actin filaments. Mutations in *FLNA* have been associated with a wide spectrum of disorders, and more recently with an X–linked form of Congenital Short Bowel Syndrome (CSBS). As CSBS patients frequently present with intestinal pseudo-obstruction (IPO), an involvement of the Enteric Nervous System (ENS) has been suggested to contribute to the reduced intestinal motility. However, histological analysis performed on bowel tissues of these patients show contradictory results. While some CSBS patients have signs of an ENS defect, others show abnormal gut smooth muscle. To understand the molecular mechanisms associated with *FLNA* mutations underlying IPO development in CSBS patients, we performed expression studies at different human embryonic stages, and a combination of *in vitro* and *in vivo* studies. We show that FLNA is expressed in the muscular layer of the small intestine from early embryonic stages. Furthermore, *FLNA* mutations associated with CSBS block expression of one of its two isoforms, possibly explaining why these patients only have intestinal complaints. Finally, by generating a transgenic zebrafish line using TALENs, we show that the longer FLNA isoform is required for intestinal development. *FLNA* mutant fish are phenotypically indistinguishable from wild-type fish, except for gut length, which is significantly reduced (20%). In conclusion, our results suggest that the intestinal motility defects associated with CSBS are caused by smooth muscle defects due to the absence of the longer FLNA isoform, together bringing new insights to IPO and CSBS pathogenesis.

2969W

Identification of *de novo* mutations in very early-onset inflammatory bowel disease (VEO-IBD). *N. Dawany[1], P. Mamula[1,2], C. M. Grochowski[1], K. E. Sullivan[1,2], R. N. Baldassano[1,2], D. Piccoli[1,2], M. Devoto[1,2], J. R. Kelsen[1,2].* 1) The Children's Hospital of Philadelphia, Philadelphia, PA. , USA; 2) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA. , USA.

Very early-onset inflammatory bowel disease (VEO-IBD) is considered to be a distinct subset of IBD. The severe presentation and early onset suggest a strong genetic etiology. The causes of and therapies for VEO-IBD are highly variable in children. While VEO-IBD has been associated with higher rates of affected first-degree family relatives, supporting the strong genetic component, we aimed to identify the contribution of *de novo* mutations to disease development. We analyzed whole exome sequencing data from 125 VEO-IBD patients whose age of onset ranged from 3 weeks to 4 years. We have previously examined the role of inheritance, including variants inherited through a compound heterozygous and a homozygous recessive model, and found possible causal variants in genes such as *DAPK1*, *IL10RA*, *CR2* and *POLE*. In this study we focused our analyses on patients whose parents had also been sequenced, including 1 quad, 67 trios and 18 mother/son duos. We then assessed the role of *de novo* occurrences of single base variants and copy number variations in the development of VEO-IBD. *De novo* variants identified in patients were filtered to include only those with an allele frequency less than 1% in the 1000 Genomes Project, EVS and ExAC, a minimum CADD score of 10 and a minimum depth of 20 in all members of a family. The percent of bases with the *de novo* variant and the presence of duplicate reads were examined to exclude possible artifacts. Copy number variation analysis was performed using the eXome-Hidden Markov Model (XHMM). Our analyses revealed candidate *de novo* mutations including a whole gene deletion in *XIAP* in a male who presented with severe IBD at 3 weeks of age and has been refractory to medical and surgical therapy. We also identified a *de novo* variant in *GPR35* (p. Pro74Ser) in a male diagnosed at 6 months of age with severe ileocolonic VEO-IBD. *GPR35*, a G coupled protein, is expressed on immune and intestinal epithelial cells and has been implicated in IBD. GPCRs, including *GPR35*, have recently been associated with gut homeostasis through regulation of metabolism and subsequent anti-inflammatory T cell regulatory response. While inherited variants have been previously implicated in the pathogenesis of VEO-IBD, in this study we have shown that in a subset of patients, *de novo* mutations are also critical in disease development.

2970T

Clinical spectrum of disease associated with *POLG* mutations. *P. Arumugam[1], R. Kothari Sonam[2], V. Jyothi[1], PS. Bindu[2], N. Gayathri[2], K. Thangaraj[1].* 1) Centre for Cellular and Molecular Biology, Hyderabad, India; 2) NIMHANS, Bangalore, India.

Mitochondrial diseases are a heterogeneous group of disorders caused by genetic defects in mitochondrial DNA or in nuclear genes. Human mitochondrial DNA (mtDNA) polymerase γ (pol γ) is the sole enzyme required to replicate the mitochondrial genome. Pol γ comprises two subunits - a catalytic p140 subunit encoded by the *POLG* gene and a smaller p55 accessory subunit encoded by the *POLG2* gene. Mutations in the *POLG* gene have emerged as one of the most common causes of inherited mitochondrial diseases in children and adults. This study sequenced the exons and intron-exon boundaries of the *POLG* gene from approximately 300 unrelated patients with clinical presentations suggestive of *POLG* related mitochondrial disease and DNA samples of 2000 healthy Indian controls were used to determine the carrier frequencies of common *POLG* (p. A467T, G848S, W748S) mutations. In this study, we identified several pathogenic reported (A467T, L304R, W748S + E1143G, H110Y and R1187W) and novel (Y837H, Q51H, K633T, R41Q and R42Q) mutations in the *POLG* gene. Their pathogenicity of novel mutations were assessed based upon (a) absence in 100 normal controls, (b) structurally non-conservative substitution of an evolutionary conserved amino acid, (c) location in protein regions of structural/functional importance, (c) mtDNA deletion and/or depletion. Several novel synonymous variants and insertions and deletions in the CAG repeats of *POLG* also were observed. The frequency of common *POLG* mutations is 0. 3% for A467T, 0. 15% for W748 and 0. 00% for G848S in India. In conclusion, the identification of *POLG* mutations is very important for diagnosis, proper medical management and appropriate genetic counseling.

2971F

CCG•CGG interruptions within CTG•CAG expansion mutation increase disease penetrance in SCA8. *B. A. Perez[1-3], Y. Ikeda[4], T. Ashizawa[1, 5], J. W. Day[6], S. H. Subramony[1, 5], L. P. W. Ranum[1, 2].* 1) Center for NeuroGenetics, University of Florida, Gainesville, FL; 2) Department of Molecular Genetics and Microbiology, University of Florida, Gainesville, FL; 3) Genetics Institute, University of Florida, Gainesville, FL; 4) Department of Neurology, Gunma University, Maebashi, Japan; 5) Department of Neurology, University of Florida, Gainesville, FL; 6) Department of Neurology, Stanford University, Stanford, CA.

Spinocerebellar ataxia type 8 (SCA8) is a dominantly inherited neurodegenerative disorder caused by a bidirectionally transcribed CTG•CAG repeat expansion. Both RNA and protein gain-of-function (GOF) mechanisms have been reported in SCA8: expanded transcripts can sequester MBNL proteins within RNA foci (RNA GOF) and expressnovel proteins in the absence of an AUG initiation codon by repeat-associated non-ATG (RAN) translation (protein GOF). A puzzling feature of SCA8 is its reduced penetrance, in which many individuals that carry the expansion mutation do not develop disease. Although SCA8 is transmitted in a dominant manner, the reduced penetrance of the disease makes SCA8 most often appear sporadic with no apparent family history. Among the 75 families we have collected samples from, 77% appeared sporadic, 16% recessive and only 7% dominant. Families with more than three affected individuals have CCG•CGG interruptions within the SCA8 CTG•CAG repeat tract at a higher frequency than families with lower disease penetrance. In the large MN-A family, members of a branch of the family showing high penetrance have variable numbers of CCG•CGG interruptions. In contrast, individuals with expanded alleles from an unaffected branch of the family have uninterrupted CTG•CAG repeats. In another family, an asymptomatic parent with a pure CTG•CAG expansion has both an asymptomatic child carrying an uninterrupted CTG•CAG expansion and an affected child with a CTG•CAG expansion with a *de novo* mutation of 3 CCG•CGG interruptions. To investigate the molecular effects of these sequence interruptions we generated vectors expressing pure or interrupted CAG expansions followed by C-terminal epitope tags in each reading frame to detect individual RAN proteins. Initial experiments show different patterns of cellular localization. We are performing additional experiments to assess changes these interruptions confer on the mutant RNAs and RAN proteins. Taken together, these data suggest CCG•CGG interruptions increase disease penetrance in SCA8 and alter cellular properties of the resulting protein products. Further investigations may provide significant insight into genetic counseling for SCA8 patients and for understanding the contribution of RAN protein toxicity in microsatellite expansion disorders.

2972W

Complete *APTX* deletion in a patient with ataxia with oculomotor apraxia type 1. *A. M. W. van den Ouweland[1], M. Guitart[2], C. Escofet[2], G. Yoon[3], P. Elfferich[1], G. M. Bolman[1], R. van der Helm[1], R. van de Graaf[1], R. van Minkelen[1].* 1) Clinical Genetics, Erasmus Medical Center, Rotterdam, Netherlands; 2) Genetic Laboratory, UDIAT-Centre Diagnòstic, Neuropediatrics Unity, Corporació Sanitéria Universitéria Parc Taulí, Sabadell, Spain; 3) Division of Clinical and Metabolic Genetics, Department of Pediatrics, The Hospital for Sick Children and University of Toronto, Canada.

*Background*Ataxia with oculomotor apraxia type 1 is an autosomal-recessive neurodegenerative disorder characterized by a childhood onset of slowly progressive cerebellar ataxia, followed by oculomotor apraxia and a severe primary motor peripheral axonal motor neuropathy. Ataxia with oculomotor apraxia type 1 is caused by homozygote or compound heterozygote mutations in *APTX* (chromosome 9p13. 3). *Case presentation*Our patient has a clinical presentation that is typical for ataxia with oculomotor apraxia type 1 with no particularly severe phenotype. Multiplex Ligation-dependent Probe Amplification analysis resulted in the identification of a homozygous deletion of all coding *APTX* exons (3 to 9). SNP array analysis using the Illumina Infinium CytoSNP-850K microarray indicated that the deletion was about 61kb. Based on the SNP array results, the breakpoints were found using direct sequence analysis: c. -5+1225_*44991del67512, p. ?. Both parents were heterozygous for the deletion. Homozygous complete *APTX* deletions have been described in literature for two other patients. We obtained a sample from one of these two patients and characterized the deletion (156kb) as c. -23729_*115366del155489, p. ?. The more severe phenotype reported for this patient is not observed in our patient. It remains unclear whether the larger size of the deletion (156kb vs 61kb) plays a role in the phenotype; no extra genes are deleted besides the non-coding APTX exons 1A and 2. *Conclusion*Here we described an ataxia with oculomotor apraxia type 1 patient who has a homozygous deletion of the complete coding region of *APTX* instead of homozygosity or compound heterozygosity for *APTX* mutations. We were unable to confirm a more severe phenotype for ataxia with oculomotor apraxia type 1 patients with a complete deletion of *APTX*, however, more research is needed to study the exact breakpoints/sizes of these kind of *APTX* deletions.

2973T

Novel Mutation Identified in the SCN1A (Channelopathy) Gene Associated with Acute Onset Ataxia, Dystonia, and Afebrile Seizures. *M. Walke, A. Rajadhyaksha, I. Miller, P. Jayakar.* Nicklaus Children's Hospital , Miami , FL.

Channelopathies including mutations in theSCN1Agene are inherited in an autosomal dominant manner. Heterozygous mutations in these genesmay cause a wide range of phenotypes, from febrile seizures to severe intractable epilepsy, known as Dravet Syndrome (Miller et al, 2014). A rare presenting feature of chanelopathiesis acute onset ataxia. Case: Currently,a 4 year old,previously healthy patient presented at 30 months of age for an evaluation of acute ataxia, dystonia, and 3 afebrile seizures that started at 27 months. No triggering factors were noted on history. Extensive genetic, metabolic work up, and neuroimaging were all normal. Whole Exome showed a heterozygous de novo mutation in exon 6 of the SCN1A gene, p. Val250Ile (GTA>ATA): c. 748G>A (NM_001165936. 1). This de novo mutation has not been reported previously in literature databases. However, it has been detected in another patient with seizures at thetest performing laboratory. Therefore, this mutation was interpreted as a disease-causing mutation. Furthermore, in silico analysis predicted the mutation to be adamaging missense mutation, due to its location in the pore region. The phenotype of Dravet syndrome includes intractable epilepsy and comorbidities ofcognitive impairment,which are usually seen by one year of age followinginitial febrile seizures. The ataxia is predicted to be caused by mutations in the NaV1. 1 channels, impairing the sodium currents and action potential firing in hippocampal GABAergic inhibitory neurons. This also affects the Purkinje neurons of the cerebellum (Catterall, et al.). Truncating mutations in the pore region of the SCN1Ageneare associated with a more severe phenotype of Dravet Syndromeincluding severe myoclonic epilepsy of infancy (SMEI) cognitive impairment and progressive ataxia. Although, missense mutations in the pore region have not been seen with this particularpresentation. Conclusion: Patients may havea variable phenotype and mutation testing for channelopathies is recommended for atypical presentationofpatients presenting with ataxia, dystonia anda history of afebrile seizures. References:Oakley, et al; Insights into pathophysiology and therapy from a mouse model of Dravet syndrome. Epilepsia. 2011 Apr;52 Suppl 2:59-61. Miller, et al; *SCN1A*-Related Seizure Disorders. Gene reviews. Catterall, et al; NaV1. 1 channels and epilepsy. J Physiol. 2010 Jun 1; 588(Pt 11): 1849–1859.

2974F

Severe Optic Nerve and Chiasm Hypoplasia, Blindness, Autism and Cognitive Handicap Associated with BMP4 Mutation. *P. Bitoun[1,2], M. legendre[3], E. Pipiras[2], S. Amselem[3], B. Benzacken[2].* 1) GROUPE MEDICAL JARENTE, Genetique Medicale, Paris, France,; 2) Hopital Jean Verdier, Paris XIII University, APHP, Embryo-cyto Génétique Bondy, France,; 3) Hopital Armand Trousseau, Paris VI University, APHP, Genetique, Paris, France,.

Introduction : BMP4 mutations are associated with anterior segment dysgenesis, coloboma and retinal dystrophy, deletions with microanophthalmia and pituitary anomaly. The authors report a male patient with BMP4 mutation. **Material & methods:** Patient had a brain MRI , endocrine evaluation, array CGH and multiple gene tests including HESX1, OTX2, SOX2, SOX3, PITX2, PITX3, BCOR, BMP4. **Results:** 1. 5 year old patient seen for blindness, 4th of 6 otherwise healthy children born uneventfully,birth weight of 3420g, height of 50cm to non-consanguinous healthy parents of north African origin. He had bilateral microcornea, sclerocornea , aniridia and large eyebulbs with staphyloma. Brain MRi at 1. 5 years showed: severe malformation of the visual pathways, near agenesis of optic nerves and chiasm, very severe hypoplasia of the proximal optic tracts not seen at the brain stem level, geniculate bodies were also not seen. The pituitary was small with continuous stalk. A colpocephaly was present w/o gyration or myelination anomaly. Patient sat at 10 months , walked at 4 years, with cognitive delay and absent speech. Bone age delay 5 years/ 8 years. Endocrine evaluation showed normal thyroid and Cortisol levels. Arginine Insuline GH stimulation test peak at 8,4 ug/l : in favor of a GH deficiency but repeat glucagon test showed a normal GH peak at 23 ng/ml. Patient had severe hypotrophy with reduced muscle mass and hyperlaxity. He was reassessed at the age of 21 and showed severe cognitive delay, lack of speech, hypotrophy at 35 kg and height of 170cm. All gene tests and array CGH was normal. BMP4 sequencing confirmed the c. 451G>A predicted to cause p. Glu151Lys mutation identified by NGS analysis using the Illumina platform …. The mother was not carrying the mutation and the father was not available for sequencing. **Discussion**: The G to A mutation occurred in a CG dimer and the Glu 1511 is highly conserved across vertebral evolution and was absent from dbSNP, Ensembl, 1000 Genomes, ExAC databases. A previously reported truncating mutation c. 226del2 mutation (p. Ser76fs104X) was associated with microanophthalmia, retinal dystrophy, polydactyly and syndactyly, brain atrophy, widened sulci and partial agenesis of the corpus callosum. **Conclusions**: A BMP4 likely de novo missense mutation was associated with severe hypoplasia of the optic nerve, chiasm and tracts with cognitive motor and language delay thus expanding the phenotype of BMP4 mutations.

2975W

A study of the role of the *FOXP2* and *CNTNAP2* genes in persistent developmental stuttering. *T. Han[1], J. Park[2], C. Domingues[3], D. Moretti-Ferreira[3], E. Paris[1], E. Sainz[1], J. Guiterrez[1], D. Drayna[1].* 1) National Institute on Deafness and Other Communication Disorders, NIH, Bethesda, MD; 2) National Institute of Neurological Disorders and Stroke, NIH, Bethesda, MD; 3) Department of Genetics, Institute of Bioscience São Paulo State University, Botucatu, São Paulo, Brazil.

A number of speech disorders including stuttering have been shown to have important genetic contributions, as indicated by high heritability estimates from twin and other studies. We studied the potential contribution to stuttering from variants in the *FOXP2* gene, which have previously been associated with developmental verbal dyspraxia, and from variants in the *CNTNAP2* gene, which have been associated with specific language impairment (SLI). DNA sequence analysis of these two genes in a group of 703 cases, all with familial persistent developmental stuttering, revealed no excess of potentially deleterious coding sequence variants in the cases compared to a matched group of 671 well characterized neurologically normal controls. This was compared to the distribution of variants in the *GNPTAB*, *GNPTG*, and *NAGPA* which have previously been associated with persistent stuttering. Consistent with previous results, *GNPTAB* and *GNPTG* showed significantly different mutant frequencies in stuttering cases and normal controls ($P = 0.0005$ for *GNPTAB*, $P = 0.0095$ for *GNPTG*). No differences in mutation frequency in the *FOXP2, CNTNAP2,* or *NAGPA* genes were observed between cases and controls. To examine the pattern of expression of these five genes in human brain, real time quantitative reverse transcription PCR was performed on RNA purified from 27 different human brain regions. The expression patterns of *FOXP2* and *CNTNAP2* were generally different from those of *GNPTAB*, *GNPTG* in terms of relatively lower expression in cerebellum. This study provides an improved estimate of the contribution of mutations in the *GNPTAB* and *GNPTG* to persistent stuttering, and suggests that variants in the *FOXP2* and *CNTNAP2* are not involved in the genesis of familial persistent stuttering. This, together with the different brain expression patterns of *GNPTAB* and *GNPTG* compared to that of *FOXP2* and *CNTNAP2*, suggests that the genetic neuropathological origins of stuttering differ from those of verbal dyspraxia and SLI.

2976T

Cell adhesion and migration defects in severe myopia associated with retinal dystrophy and cystic cerebellar dysplasia due to biallelic *LAMA1* mutations. *M. C. Malicdan[1,2], T. Vilboux[2,11], Y. M. Chang[2], P. M. Zerfas[3], J. Guo[2], A. R. Cullinane[2], M. Bryan[2], D. Yildirimli[2], J. Bryant[2], R. Fischer[2], B. P. Brooks[4], W. M. Zein[4], E. A. Wiggs[5], C. Z. Zalewski[6], A. Poretti[7], M. Vemulapalli[8], A. Young[8], B. Maskeri[8], J. Mullikin[8], M. Kirby[9], S. M. Anderson[9], C. Toro[1], M. Huizing[2], W. A. Gahl[1,2,10], M. Gunay-Aygun[2,10].* 1) NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, 20892 Maryland, USA; 2) MGB, NHGRI, National Institutes of Health, Bethesda, 20892, Maryland, USA; 3) Diagnostic and Research Services Branch, Office of Research Services, National Institutes of Health, Bethesda, 20892 Maryland, USA; 4) Ophthalmic Genetics & Visual Function Branch, National Eye Institute, National Institutes of Health, Bethesda, 20892 Maryland, USA; 5) Office of the Clinical Director, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, 20892 Maryland, USA; 6) Audiology Unit, Otolaryngology Branch, National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, 20892 Maryland, USA; 7) Section of Pediatric Neuroradiology, Division of Pediatric Radiology, Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University, Baltimore, 21218 Maryland, USA; 8) NIH Intramural Sequencing Center (NISC), National Human Genome Research Institute, National Institutes of Health, Bethesda, 20892 Maryland, USA; 9) Flow Cytometry Core, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA; 10) Office of the Clinical Director, National Human Genome Research Institute, National Institutes of Health, Bethesda, 20892 Maryland USA; 11) Inova Translational Medicine Institute, Falls Church, 22042Virginia, USA.

Statement of Purpose: Laminins are heterotrimeric glycoprotein complexes, consisting of α, b and γ subunits, that form a major component of basement membranes and extracellular matrix. Laminin complexes have different, but often overlapping, distributions and functions, and laminin defects have been associated with many different disorders. In this work, we define the clinical phenotype of LAMA1 deficiency with anxiety, cerebellar dysplasia with cysts, and severe myopia with retinal dystrophy. **Methods:** Under our NIH protocol, NCT00068224, we evaluated three individuals in two families presenting with anxiety, tics, cerebellar dysplasia with cysts, and myopia. We performed whole exome sequencing on affected individuals and available family members. To determine the effect of gene mutation in cells, we analyzed RNA and protein expression in fibroblasts derived from patient cells, as well as explore the mechanism that may explain the cell defect. **Results:** All affected individuals have null mutations in *LAMA1*, a gene that encodes an important component of the basement membrane. Our functional studies in cultured fibroblasts of subjects with LAMA1 deficiency identified impaired adhesion, reduced migration, abnormal morphology and increased apoptosis. Importantly, LAMA1-knockdown in human neuronal cells also showed abnormal morphology and filopodia formation, supporting the importance of LAMA1 in neuronal migration, and marking these cells as useful for further therapeutic explorations. **Conclusions:** We describe detailed clinical, neuroimaging, and functional studies of lamininopathy associated with *LAMA1* mutations, characterized by anxiety, cystic cerebellar dysplasia with cysts, and severe myopia associated with retinal degeneration. Whether tics, obsessive compulsive traits, and anxiety are a primary part of this phenotype will likely be determined after description of other patients with *LAMA1* mutations. Based upon population studies (1000 Genomes and NHLBI Exome Sequencing Project) potential null variants (stop gain, frame shift or splice variants) in *LAMA1* have an estimated frequency of 0.53%; this translates into a prevalence of 1 in 35,000 for this new disease. *LAMA1* should be considered a candidate gene in patients with anxiety, tics, cerebellar dyspasia with cysts and severe myopia with retinal dystrophy.

2977F

Identification of homozygous mutations in the *CHRNG* genes by whole exome sequencing in siblings presented with congenital arthrogryposis multiplex. *G-H. Kim[1], J-M. Kim[1], J. Cho[2], E. Kang[2], J-H. Choi[2], B. Lee[1,2], H-W. Yoo[1,2].* 1) Med Gen Clinics Center, Asan Med Ctr, Seoul, South Korea; 2) Department of Pediatrics, Asan Medical Center Children's Hospital, University of Ulsan College of Medicine, Seoul, South Korea.

Multiple pterygium syndrome(MPS) is a genetically heterogeneous disorder caused by germline mutations in the *CHRNG* gene. It is characterized by congenital contractures of multiple joints, scoliosis, muscle weakness, typical facial features, and multiple pterygium. This study describes two siblings with MPS presented with multiple joint contractures and their molecular characteristics. A 7-year-old girl presented to our institute for the evaluation of multiple joint contractures. She was born at term with a birth weight of 3. 4 kg with no perinatal problems to non-consanguineous parents. The patient manifested multiple joint contractures of both metacarpal joints after birth and experienced gait disturbance due to knee joint contractures as she grow older. Her younger brother was born at 37 weeks of gestation with a birth weight of 2. 6 kg. Polyhydramnios was detected by prenatal ultrasonography. Cardiac pacemaker was implanted due to sick sinus syndrome at the age of 3 months. He had both inguinal hernia, which were corrected surgically during infancy. He presented with multiple joint contractures, cervical spine kyphosis, both camptodactyly, vertical talus, and rocker bottom feet at the age of 5 years. Whole exome sequencing (WES) analysis of the trio family identified that the proband was homozygous for a p. Pro143Arg of *CHRNG*, which was predicted to be detrimental by PolyPhen-2 and SIFT. Also, this variant was not found in the 1000Genomes database (http://browser. 1000genomes. org/) and ClinVar (http://www. ncbi. nlm. nih. gov/clinvar/). Both parents were heterozygous for the variant. The variant was confirmed by Sanger sequencing and her younger brother also harbored homozygous mutation in p. Pro143Arg in *CHRNG*. The MPS is an extremely rare and phenotypically heterogeneous disorder from lethal to non-lethal forms. Both of them are caused by mutations in *CHRNG*, encoding the gamma subunit of the embryonic acetylcholine receptor which has a role in the muscle-relaxant effect. This study reports two siblings with MPS harboring homozygous mutation in *CHRNG* for the second time. The first two Korean MPS cases also carried the p. Pro143Arg mutation identified by WES, representing possible founder effect in Korean population. WES analysis is helpful to make rapid, confirmative diagnosis of ultra-rare disorder.

2978W

A family with Joubert and Meckel-Gruber Syndrome phenotypes associated with *TMEM231* mutations. *D. Maglic[1], T. Vilboux[1,5], J. Stephen[1], D. Konzman[1], J. Bryant[1], M. Vemulapalli[3], B. Maskeri[3], A. Young[3], J. Mullikin[3], W. A. Gahl[1,2,4], M. C. Malicdan[1,2], M. Gunay-Aygun[1].* 1) Medical Genetics Branch, National Institutes of Health, Bethesda, MD, MD; 2) NIH Undiagnosed Diseases Program, Common Fund, Office of the Director, National Institutes of Health, Bethesda, Maryland, USA; 3) NIH Intramural Sequencing Center and Comparative Genomics Unit, Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892, USA; 4) Office of the Clinical Director, NHGRI, National Institutes of Health, Bethesda, MD USA; 5) Inova Translational Medicine Institute, Falls Church, Virginia, USA.

Statement of Purpose: Joubert syndrome is defined based on a distinctive brain malformation (molar tooth sign) on brain imaging. Most Joubert syndrome patients display hypotonia, developmental delay, abnormal eye movements, and an abnormal respiratory pattern in infancy. Variable features include fibrocystic kidney disease, congenital hepatic fibrosis, retinal degeneration, retinal colobomas, and polydactyly. Meckel-Gruber syndrome, which is typically fatal perinatally, is associated with polycystic kidneys, occipital encephalocele and polydactyly. Both syndromes are genetically heterogeneous with multiple causative genes, some of which may cause either syndrome. In this paper we report a family with overlappingJoubert and Meckel-Gruber syndrome. **Methods:** Under our NHGRI ciliopathy study (www. clinicaltrials. gov, trial NCT00068224), we have evaluated over 100 patients with Joubert syndrome. Evaluations at the NIH Clinical center included family history and physical examination and comprehensive biochemical and imaging studies. Whole exome sequencing (WES) was performed on all affected and available members of the family. **Results:** Here, we report a family with 4 children with Joubert syndrome. None of first three children with Joubert syndrome had polydactyly and their kidney ultrasonograpy was normal at ages 4. 8 (twins) and 1. 2 years. The fourth sibling displayed polycystic kidneys on prenatal ultrasonography and unfortunately passed soon after birth. Interestingly, the children's paternal uncle and his wife had a fetal demise due to Meckel-Gruber syndrome. By WES we identified a homozygous missense variant (c. 712G>A, p. D238N) in *TMEM231*, a previously reported variant associated with Joubert syndrome, in all four affected siblings. Since the mother was the only parent heterozygous for this mutation, we suspect the affected children to be hemizygotes and the father to be a carrier of a big deletion not detectable by WES. We suspect that this deletion in *TMEM231* variant is also present in the fetus with Meckel-Gruber syndrome likely in combination with another severe variant (c. 334T>G, p. Trp112Gly) inherited from the mother. **Conclusion:** We describe a three-generation, non-consanguineous family who harbor several variants in *TMEM231* that resulted in a complex overlap of Joubert and Meckel-Gruber syndromes.

2979T

Novel autosomal recessive intellectual disability syndrome with manganese deficiency, muscular hypotonia, and cerebellar atrophy caused by mutation of the SLC39A8 transporter. *R. Abou Jamra[1,5], C. H. Beaulieu[2], A. Mhanni[3], O. H. Gebril[4], K. Kernohan[2], A. E. Chudley[3], H. Tawamie[5], F. Radwan[5], J. Schwatzentruber[6], . FORGE Canada Consortium[2], J. Majewski[6], S. Uebe[5], A. Ekici[5], A. Reis[5], A. M. Innes[7], S. Parboosingh[7], K. M. Boycott[2].* 1) Centogene, Rostock, MV, Germany; 2) Children's Hospital of Eastern Ontario Research Institute; University of Ottawa, Ottawa, Canada; 3) Section of Genetics and Metabolism; Children's Hospital and the Department of Pediatrics and Child Health; University of Manitoba, Winnipeg, Canada; 4) Department of Research on Children with Special Needs; National Research Centre (NRC), Cairo, Egypt; 5) Institute of Human Genetics; FAU Erlangen-Nürnberg, Erlangen, Germany; 6) McGill University and Genome Quebec Innovation Centre, Montreal, Canada; 7) Department of Medical Genetics; Alberta Children's Hospital and University of Calgary, Calgary, Canada.

We ascertained two Egyptian siblings of consanguineous parents both presenting with severe intellectual disability, hypotonia and hyporeflexia, strabismus, short stature, early-onset epilepsy or absences, and normal head circumference with brain atrophy in one child. Autozygosity mapping and exome sequencing revealed a candidate mutation in SLC39A8 (encoding the protein ZIP8) in a homozygous status that leads to an alteration on protein level; p. Gly38Arg. In silico simulations and molecular modelling showed that this is a highly conserved residue and that the alteration is probably pathogenic. Independently, further patients were identified within the FORGE (Finding of Rare Disease Genes) Canada project; five children of 5 families of Hutterite descent were also found homozygous for the p. Gly38Arg alteration. The patients presented severe developmental delay, severe hypotonia, short stature, cerebellar atrophy with normal head circumference, mild skeletal dysplasia, and connective tissue abnormalities. SLC39A8 is a member of the solute carrier gene (SLC) superfamily. The encoded protein ZIP8 is a Zn/HCO symporter, and also transports Mn and Cd across the plasma membrane. Evaluation of trace element levels in affected patients revealed low blood Mn levels in all the Egyptian and the Canadian patients, and low Zn levels in the Canadian patients, further supporting the pathogenicity of the alteration. Cell cultures of several tissues taken from hypomorphic Slc39a8(neo/neo) mice exhibit diminished zinc and iron levels, and die lately 48 h postnatally. Defects include severely hypoplastic spleen, hypoplasia of liver, kidney, lung, and lower limbs. We conclude that the variant we identified in SLC39A8 is pathogenic, and that it probably leads to a distinguishable autosomal recessive, severe intellectual disability syndrome with Mn deficiency, short stature, and muscular hypotonia. Further symptoms are brain atrophy with normal head circumference, seizures, strabismus, and osteopenia. Based on the severity of the symptoms of our patients and on the results of the hypomorphic mouse model, we think - at this stage - that the variant we identified in SLC39A8 influences the functions of the protein, but is not a loss of function. This is the first gene associated with a human Mn deficiency syndrome and this finding also provides insight into the role of Mn homeostasis in development and health.

2980F

Increased *STAG2* Dosage Defines a Novel Cohesinopathy with Intellectual Disability and Behavioural Problems. *J. Gecz[1], R. Kumar[1], M. A. Corbett[1], B. W. M. van Bon[2], A. Gardner[1], L. A. Jolly[1], K. L. Friend[3], c. Tan[1], H. van Esch[4], M. Ryanaud[5], M. Field[6], M. Lefler[6], B. Budny[7], M. Wisniewska[8], M. Badura-Stronka[8], A. Latos-Bielenska[8], J. Batanian[9], J. A. Rosenfeld[10,14], L. Basel-Vanagaite[11], R. Ullmann[12,16], H. Hu[12], M. I. Love[13], S. A. Haas[13], P. Stankiewicz[14], S. W. Cheung[14], A. Baxendale[15], J. Nicholl[3], E. M. Thompson[1,15], E. Haan[1,15], V. M. Kalscheuer[12].* 1) Dept Paediatrics, The University of Adelaide, Women's & Children's Hosp, Adelaide, South Australia, Australia; 2) Radboud University Medical Center, 6525 GA, Nijmegen, The Netherlands; 3) Genetics and Molecular Pathology, SA Pathology, North Adelaide, SA 5006, Australia; 4) Center for Human Genetics, University Hospitals Leuven, 3000 Leuven, Belgium; 5) Centre Hospitalier Régional Universitaire, Service de Génétique, 37000 Tours, France; 6) Genetics of Learning Disability Service, Hunter Genetics, Waratah, NSW 2298, Australia; 7) Department of Endocrinology, Metabolism and Internal Diseases, Poznan University of Medical Sciences, Poznan 60-355, Poland; 8) Department of Medical Genetics, Poznan University of Medical Sciences, Poznan 60-355, Poland; 9) Department of Pediatrics, Saint Louis University, St Louis, MO 63104, USA; 10) Signature Genomic Laboratories, Spokane, WA 99207, USA; 11) Raphael Recanati Genetic Institute and Felsenstein Medical Research Center, Rabin Medical Center, Beilinson Campus, Petah Tikva 49100, Israel; 12) Department of Human Molecular Genetics, Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany; 13) Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestrasse 73, 14195 Berlin, Germany; 14) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; 15) South Australian Clinical Genetics Service, SA Pathology, North Adelaide, SA 5006, Australia; 16) Bundeswehr Institute of Radiobiology, 80937 Munich, Germany.

Next generation genomic technologies have made a significant contribution to the understanding of the genetic architecture of human neurodevelopmental disorders. Copy number variants (CNVs) play an important role in the genetics of intellectual disability (ID). For many CNVs, and copy number gains in particular, the responsible dosage-sensitive gene(s) have been hard to identify. We have collected 18 different interstitial microduplications and one microtriplication of Xq25. There were 6 familial and 13 singleton cases (33 affected males and 10 carrier females). The critical overlapping region involved the *STAG2* gene, which codes for a subunit of the cohesin complex that regulates cohesion of sister chromatids and gene transcription. We demonstrate that *STAG2* is the dosage-sensitive gene within these CNVs, as gains of *STAG2* mRNA and protein deregulate disease-relevant neuronal gene networks in cells derived from affected individuals as determined by RNA-seq. We also show that *STAG2* gains result in increased expression of *OPHN1*, a known X-chromosome ID gene. Overall we define a novel cohesinopathy with behavioural problems due to copy number gain of Xq25 and *STAG2* in particular.

2981W

Contribution of Copy Number Variation to the Genetic Etiology of Duane Retraction Syndrome. *V. K. Robson, A. Digioia, W. M. Chan, C. Andrews, S. J. Bekele, E. C. Engle.* Neurology, Boston Children's Hospital, Boston, MA.

Duane Retraction Syndrome (DRS) is a congenital eye movement disorder associated with aberrant innervation of the lateral rectus muscle by the motor branches of the third and sixth cranial nerves. We previously reported the only known genetic causes of isolated (*CHN1* [MIM118423,604356]) and syndromic (*HOXA1*[MIM 142955, 126800]and *SALL4* [MIM607343,607323]) DRS. The genetic etiology of most DRS cases, however, remains unclear. Rare case studies of simplex syndromic DRS have implicated de novo chromosomal structural abnormalities that cause damaging copy number variants (CNVs). To investigate these structural variants in detail, we are conducting a systematic analysis of CNVs in a cohort of patients with syndromic DRS. From a cohort of 500 DRS probands without known mutations, we selected 84 simplex probands with syndromic DRS to clinically characterize and code according to phenotype. From among these 84 probands, 32 full and 10 incomplete trios were selected for CNV analysis. Whole genome genotyping data were generated for the 116 samples (42 affected probands, 74 unaffected family members) using a Human Omni2. 5 + Exome SNP Chip bead array from Illumina containing 2. 5 million unique SNPs and supplemental exonic coverage. Our deep phenotypic analysis allowed us to identify probands with known DRS-associated syndromes within our cohort, including Wildervanck (7 patients), Duane's Radial Ray (3), and Goldenhar (1) syndromes. Of the probands that do not fit into a clearly defined syndrome, abnormalities involve the following systems: musculoskeletal (24), neurological (18), hearing (17), ophthalmic (14), developmental motor delay (12), dysmorphia (12), cardiavascular (11), intellectual and social disability (9), gastrointestinal (7), and genitourinary (5). After preliminary investigation, we have identified a variety of de novo CNVs across the cohort, including a deletion on 7q31 and copy number gain on Xq23. Further work will confirm rare or novel CNVs by ddPCR, determine whether they are present among other probands in the cohort, and investigate their functional impact. Thus, this project will systematically define the role of CNVs in syndromic DRS in order to better inform the genetic diagnosis of this disorder, enhance the understanding of CNVs in human disease, and potentially provide new avenues of research for targeted therapeutics of neurodevelopmental disorders.

2982T

***NAA10* missense mutations cause neurodevelopmental delay in eight female patients.** *C. Saunier[1], C. Zweier[2], S. I. Støve[3], B. Gérard[4], A. Piton[4], M. Blenski[3], J. Thevenon[5,6], A. Masurel[5,6], B. Popp[2], M. Wasserstein[7], P. Goldenberg[8], G. M. Mancini[9], K. Tezcan[10], B. Isidor[11], B. Leheup[12], Y. Duffourd[5], JB. Rivière[5,13], L. Faivre[5,6], N. Ah Mew[14], A. Reis[2], T. Arnesen[2,15], C. Thauvin-Robinet[5,6].* 1) Service de Pédiatrie, Hôpital d'Enfants, CHU de Dijon, Dijon, France; 2) Institute of Human Genetics, Friedrich-Alexander-Universitét Erlangen-Nürnberg, Erlangen, Germany; 3) Department of Molecular Biology, University of Bergen, Norway; 4) Laboratoire de Génétique Moléculaire, CHRU Strasbourg, Strasbourg, France; 5) Equipe EA4271 GAD, FHU-TRANSLAD, Université de Bourgogne/CHU Dijon, France; 6) Centre de Référence maladies rares "Anomalies du Développement et syndrome malformatifs" de l'Est et Centre de Génétique, Hôpital d'Enfants, CHU, Dijon, France; 7) Genetics and Genomic Sciences and Pediatrics, Icahn School of Medicine at Mount Sinai, New York City, USA; 8) Division of Pediatrics, Mass General Hospital for Children, Boston, USA; 9) Department of Clinical Genetics, Erasmus MC, Sophia Children's Hospital, Rotterdam, The Netherlands; 10) Medical Genetics, Kaiser Permanente, Sacramento, California, USA; 11) Service de Génétique Médicale, CHU Nantes, France; 12) Service de Génétique Médicale, Hôpital Brabois, CHU Nancy, France; 13) Laboratoire de Génétique Moléculaire, PTB, CHU Dijon, France; 14) Center for Translational Sciences, Children's National Health System, The George Washington University, Washington, DC; Division of Genetics and Metabolism, Children's National Health System, The George Washington University, Washington, DC, USA; 15) Department of Surgery, Haukeland University Hospital, Norway.

The *NAA10* gene encodes for *N*-alpha-acetyltransferase 10, the catalytic subunit of the N-terminal acetyltransferase A complex (NatA). NatA is the major enzyme responsible for N-terminal acetylation, one of the most frequent protein modifications harbored by approximately 80-90% of all human proteins. In mice, the *Naa10* gene is ubiquitously expressed in adults and highly expressed in the brain during embryonic development. It has also been suggested that NAA10 is involved in regulating dendrite growth, cell cycle and apoptosis. *NAA10* mutations have been described in eight males from two different families with X-linked severe lethal Odgen syndrome (OGDNS), linked to the same missense mutation (p. Ser37Pro) and in four related males with Lenz microphthalmia syndrome, linked to a splice donor site mutation (c. 471+2T>A)). Two other *NAA10* missense mutations (p. Val107Phe and p. Arg116Trp) have recently been implicated in a male and a female with non-syndromic developmental delay. N-terminal acetylation assays for the three previously published missense variants demonstrated an impaired acetyltransferase activity. We report on seven new female cases with *de novo NAA10* missense mutations and one with an inherited mutation (six independent cases with a hotspot p. Arg83Cys mutation and two cases with collocated p. Phe128Leu and p. Phe128Ile mutations). Like the previously published female patient, the new cases present severe intellectual disability, truncal hypotonia, microcephaly and minor similar facial dysmorphism. Functional testing of Naa10 Arg83Cys and Phe128Leu revealed distinct, but significant alterations in protein parameters for the two mutants. Both mutants have an impaired ability to carry out protein N-terminal acetylation. In conclusion, this report of eight new female cases strongly supports the involvement of *de novo NAA10* missense mutations in X-linked intellectual disability and particularly in females with severe intellectual disability, truncal hypotonia, borderline microcephaly and minor facial dysmorphism, linked to protein N-terminal acetylation defects. These data also confirm that *de novo* mutations on the X-chromosome can cause developmental disorders in females.

2983F

Characterization of *SCN8A* mutations and phenotypes in comparison to *SCN1A*. *D. Chen[1], A. Ishii[2,3], M. Hammer[4].* 1) Genetics, University of Arizona, Tucson, AZ; 2) University of Arizona Genomics Core, University of Arizona, Tucson, AZ; 3) Department of Pediatrics, Fukuoka University School of Medicine, Fukuoka, Japan; 4) Arizona Research Laboratories, University of Arizona, Tucson, AZ.

SCN8A, a voltage-gated sodium ion channel, is a large protein (260 kDa) with four repeats. Each of these repeats has six transmembrane domains with connecting loops in between them. It is in the same gene family as *SCN1A*, a causal gene of Dravet Syndrome (DS), a severe, infantile epilepsy. *SCN1A* and *SCN8A* are highly conserved and have high homology. Knockouts in *SCN8A* cause ataxic phenotypes in mice and have only recently been shown to cause seizures and major neurodevelopmental delay in humans. Since then, 42 additional pathogenic mutations have been reported. On the other hand, more than a thousand mutations in *SCN1A* have been reported and well studied, with phenotypes characterized by febrile seizures. Using what we know about *SCN1A*, we aim to characterize *SCN8A* mutations and their corresponding phenotypes and ultimately discover patterns that may be used to predict pathogenicity of new mutations. We will be using the 43 *SCN8A* mutations published in the literature as well as a sample of 270 Dravet patients with thorough genetic and phenotypic data.

In *SCN8A*, we have 38 (88%) mutations that are *de novo*, 2 (5%) inherited, and 2 (5%) with unknown inheritance patterns. All mutations are missense except for one nonsense mutation. Interestingly, the patient presented with ataxia, not epilepsy. In comparison, more than a half of the mutations in *SCN1A* are truncation mutations (151/270, 56%) and result in a more severe phenotype. 21 *SCN8A* mutations appear in transmembrane domains while 22 appear in the loops. Despite the loops being nearly three times longer than the transmembrane domains, mutations are significantly more likely to appear in transmembrane domains than the loops in between (Fisher's Exact Test, two-sided p-value=0. 0034). This may be because mutations in loops result in a milder phenotype and may be subclinical. *SCN1A* mutations do not have the same significance (Fisher's Exact Test, two-sided p-value=0. 3829).

We hope to increase our sample size over the next few years as more mutations are reported. We are also establishing a framework to crowdsource patient data by creating an online presence where parents of children with *SCN8A* mutations can contact researchers and each other for information. Although not ready to publish, we have already collected 33 responses. Future plans also include studying mouse models with different *SCN8A* mutations.

2984W

Whole exome sequencing identified a splice site mutation in ARV1 in a consanguineous family with intellectual disability, epilepsy and neurodegeneration. *Y. Guo[1], L. Worgan[2], L. Tian[1], C. Hou[1], F. Collins[3,4,5], H. Hakonarson[1,6], J. Christodoulou[4,5].* 1) Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA; 2) Department of Clinical Genetics, Liverpool Hospital, Liverpool, NSW, Australia; 3) Clinical Genetics, The Children's Hospital at Westmead, Sydney, NSW, Australia; 4) University of Sydney, Sydney, NSW, Australia; 5) Western Sydney Genetics Program, Children's Hospital Westmead, Westmead, NSW, Australia; 6) Department of Pediatrics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

In a large consanguineous family of Lebanese origin, a total of four children are diagnosed with intellectual disability, epilepsy, hypotonia and neurodegeneration. Numerous genetic as well as neurological tests were unrevealing thus we conducted whole exome sequencing (WES) on two of the affected children, each in a separate branch of the family. We identified a homozygous splice site mutation in ARV1, the homolog of S. cerevisiae gene ARV1. The mutation was validated in the family using Sanger sequencing. With only 2 heterozygotes in the ExAC database of more than 60,000 samples, this mutation is at a very conserved position, and could truncate the encoded protein. ARV1 was recently reported as a candidate gene in a family with convulsive disorder/developmental delay/epilepsy, and is involved in sphingolipid metabolism, malfunction of which is seen in other neurodegenerative disorders such as Nieman-Pick disease. At present, we are performing follow-up functional experiments to establish the mutation's deleterious impact on intracellular sterol trafficking as well as sphingolipid metabolism.

2985T

***TUBGCP4,* encoding a member of the γ-tubulin ring complex γTuRC, is a novel gene for autosomal recessive microcephaly with chorioretinopathy (MCMR).** *H. J. Dollfus[1], S. SCHEIDECKER[1], C. ETARD[2], L. HAREN[3], S. HULL[4,5], G. ARNO[4,5], V. PLAGNOL[5], S. DRUNAT[6], S. PASSEMARD[6], A. TOUTAIN[7], C. OBRINGER[1], V. KOOB[8], V. GEOFFROY[1], U. STRAHLE[2], P. OSTERGAARD[9], A. VERLOES[6], C. MERDES[3], M. MOORE[4,5], C. STOETZEL[1].* 1) Medical Genetics Laboratory , CARGO, IGMA, INSERM U1112 , Strasbourg University, Strasbourg , Alsace , France; 2) Institut für Toxikologie und Genetik Campus Nord, Karlsruher Institut für Technologie, Eggenstein-Leopoldshafen, Germany; 3) Centre de Biologie du Deéveloppement, Université Paul Sabatier, Toulouse, France; 4) Inherited Eye Diseases, UCL Institute of Ophthalmology, London, UK; 5) Moorfields Eye Hospital, London EC1V 2PD, UK; 6UCL Genetics Institute, London UK; 6) Unité Fonctionnelle de Génétique Moléculaire, Département de Génétique, Hôpital Robert Debré, Centre Hospitalier Universitaire Paris, Paris, France; 7) Département de Génétique Médicale, Centre Hospitalier Régional et Universitaire de Tours, Tours, France; 8) Service de RadioPédiatrie et Imagerie, Hôpitaux Universitaires de Strasbourg et Laboratoire ICube, Université de Strasbourg, Centre National de la Recherche Scientifique, Strasbourg, France; 9) Human Genetics, Cardiovascular and Cell Sciences Institute, St. George's University of London, London , UK.

In 1966, Pr V McKusick was the first to report on autosomal recessive microcephaly with chorioretinal in dysplasia in a Mennonite family (Arch Ophthalmol, 1966). Recently, mutations in *TUBGCP6* were identified for this unique family (Pfuffenberger, Plos one, 2012) with Autosomal recessive Microcephaly and Chorioretinopathy with or without Mental Retardation, abbreviated as MCMR. Autosomal dominant Microcephaly with or without Chorioretinopathy, Lymphoedema, or Mental Retardation (MCLMR) also known as Alzial syndrome has recently been attributed to mutations in *KIF11*(Oostergaard P, 2012). We have identified *TUBGCP4* variants in individuals with MCLMR with very specific retinal phenotypes. Whole exome sequencing performed on one family with two affected siblings and independently on another family with one affected child revealed compound heterozygous mutations in *TUBGCP4*. Subsequent Sanger sequencing was performed on a panel of individuals with microcephaly and ophthalmic manifestations and one other patient was identified with compound heterozygous mutations in *TUBGCP4*. *TUBGCP4* encodes the γ-tubulin complex protein 4, a component of the γ-tubulin ring complex (γTuRC) known to regulate the nucleation and organization of microtubules. Functional analysis of patient fibroblasts disclosed reduced levels of the γTuRC, altered nucleation and organization of microtubules, abnormal nuclear shape, and aneuploidy. Moreover, zebrafish treated with morpholinos against *tubgcp4* were found to have a reduced head volume and eye developmental anomalies with very specific chorioretinal dysplasia features. The identification of biallelic *TUBGCP4* mutations confirms the existence of autosomal recessive cases with microcephaly and chorioretinopathy that actually present a very specific phenotype in the syndromic microcephaly group.

2986F

A nonsense *C12orf65* mutation in Indian-Jewish monozygotic female twins with Leigh syndrome. *E. Imagawa[1], A. Fattal-Valevski[2], O. Eyal[3], S. Miyatake[1], A. Saada[4], M. Nakashima[1], T. Tsurusaki[1], H. Saitsu[1], N. Miyake[1], N. Matsumoto[1].* 1) Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 2) Paediatric Neurology Unit, Tel Aviv Sourasky Medical Centre, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 3) Paediatric Endocrinology Unit, Tel Aviv Sourasky Medical Centre, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel; 4) Monique and Jacques Roboh Department of Genetic Research and the Department of Genetic and Metabolic Diseases Hadassah-Hebrew University Medical Center, Jerusalem, Israel.

Leigh syndrome (LS: MIM#256000) is an early-onset neurodegenerative disorder leading to mitochondrial dysfunction. LS is characterized by elevated lactate/pyruvate together with bilateral and symmetric high-intensity signals in brainstem, basal ganglia, cerebellum and spinal cord on T2-weighted brain MRI. To date, approximately 50 nuclear genes have been reported in LS. Here, we present Indian-Jewish monozygotic female twins affected with LS. The affected twins were born to consanguineous parents, and showed a failure to thrive, developmental delay, ophthalmoplegia, optic atrophy, spastic paraparesis, muscle weakness of lower limbs, mild intellectual disability and cognitive impairment. They showed elevated serum lactate, symmetrical brain lesions in brainstem on T2-weighted MRI and decreased mitochondrial complex IV activity on muscle biopsy. To identify the genetic cause of them, we performed genetic analysis using whole exome sequencing. Based on the hypothesis of autosomal recessive model (homozygote and compound heterozygote), we identified one homozygous nonsense mutation in *C12orf65* (c. 346delG, p. Val116*; NM_001143905) in both affected twins. Its heterozygous change was identified in their parents and no mutation showed in their healthy sibling. *C12orf65* is a mitochondrial matrix protein belonging to the family of class I peptide chain release factors, associated with the mitochondrial translation termination. The *C12ore65* mutations have been already reported in LS, in addition autosomal recessive spastic paraplegia-55 (MIM #615035), Charcot-Marie Tooth type6 (CMT type 6, MIM #601152) and syndromic autosomal recessive intellectual disability. Interestingly, the previous reported mutations were all truncating, which produce the truncated C12orf65 protein due to escape of nonsense-mediated decay. It is known that clinical severity would be dependent on the residual length of the truncated C12orf65 protein. The identified mutation resulted in mid-length truncation, and the clinical features are moderate. The identical mutation has been previously reported in CMT type 6 patients by Tucci et al. but the presenting LS case displayed more severe phenotypes, thus the phenotypic variability may be altered by unidentified mutations/modifiers.

2987W

Linkage mapping and exome sequencing identifies a novel genetic cause of Charcot-Marie-Tooth neuropathy with pyramidal signs. *M. L. Kennerson[1,2,3], O. Albulym[1,2], A. P. Drew[1], M. Harms[4], M. Auer-Grumbach[5], A. Pestronk[4], A. Connolly[4], R. H. Baloh[6], S. R. Reddel[1,3,7], S. Zuchner[8], G. A. Nicholson[1,2,3].* 1) Northcott Neuroscience, ANZAC Res Inst, Concord, Australia; 2) Sydney Medical School, University of Sydney, NSW 2008, Australia; 3) Molecular Medicine Laboratory, Concord Hospital, NSW 2139, Australia; 4) Department of Neurology; Washington University School of Medicine St. Louis, Missouri USA 63110; 5) Department of Internal Medicine, Medical University of Graz, Austria; 6) Department of Neurology; Cedars Sinai Medical Center, Los Angeles, CA, USA; 7) Brain & Mind Research Institute, Sydney, New South Wales, Australia; 8) John P. Hussman Institute for Human Genomics, Miami University, US.

Introduction: Charcot-Marie-Tooth (CMT) disease is a group of degenerative disorders of human peripheral nerve affecting both the motor and sensory neurons. This genetically and clinically heterogeneous syndrome is the most common inherited neuromuscular disease affecting 1 in 2500 individuals. The CMT phenotype is characterised by progressive weakness and atrophy of distal muscles, high arched feet (pes cavus) and loss of deep tendon reflexes. In this study we have used a combination of genetic linkage and whole exome sequencing (WES) to identify the causative gene in an Australian family (CMT105) diagnosed with CMT2 and pyramidal signs. **Methods:** Genome wide microsatellite markers were genotyped and analysed using standard parametric genetic linkage analysis. WES was performed on selected individuals (3 affected, one normal). Segregation analysis and mutation screening of CMT2 families was performed using high resolution melt analysis (HRM). The GEM. app database was queried to identify additional families harboring mutations in genes of interest. **Results:** Linkage analysis revealed significant LOD scores ≥ +3 on chromosome 22q12. 1-q13. 2 and extended haplotype analysis mapped a new CMT2 locus to a 6. 6-Mb interval flanked by the markers D22S1154 and D22S280. WES identified a novel mutation c. 754C>T (p. R252W) in the *MORC2* gene which mapped within the linkage region and fully segregated with the disease phenotype in family CMT105. Screening additional families and querying the GEM. app database identified the p. R252W mutation in two unrelated early onset CMT2 families and a second mutation c. 707A>G (p. E263G) was identified in two unrelated CMT2 families. Both the mutations occurred at highly conserved amino acid residues and were absent in the normal population and 1000 in-house neurologically normal control chromosomes. **Conclusions:** We have identified a new locus for CMT2 and shown compelling genetic evidence that *MORC2* mutations are the likely pathogenic cause of axonal CMT in these families. *MORC2* encodes the human CW-type zinc finger 2 protein which is a chromatin modifier involved in the regulation of DNA repair as well as gene transcription regulation. Further studies are needed to define the role of MORC2 protein in motor and sensory neurons and to further understand the underlying pathogenic mechanisms in which *MORC2* mutations lead to axonal neuropathy. M. L. K. and O. A. contributed equally to this work.

2988T

Splicing of a 46bp sequence from central inton 9 of human *FMR1* gene results in a truncted FRMP peptide with altered subcellular distribution and different functions when overexpressed. *F. Lan, X. Fu, J. Liao, X. Guo, A. Yan, W. Yang, D. Zhang.* Departmentt of Medical Genetics, Fuzong Clinical College, Fujian Medical University, Fuzhou, Fujian , China.

FMRP is encoded by *FMR1*gene,disease gene of fragile X syndrome (FXS), andhas several isoforms resulted from alternative splicing of *FMR1.* Recent reports demonstrate that the alternative splicing of *FMR1* is far more complex than what has been recoginized. In the analysis of alternatively spliced *FMR1* transcripts in the blood cells from a patient with FXS-like phenotypes, we detected a 46bp sequence lying between sequences of exons 9 and 10, which was spliced from the middle of intron 9. Bioinformatics shows that there are cononical splicing signals around this sequence, indicating the potentiality of this sequence as a cryptic alternative exon. A genetic variation was found in upsteam intron 9, but this could not resulted in any preferred splicing of this exon, as shown by hybrid minigene splice assay. This sequence could also be detected in blood cells of 50 normal controls using semi-nested PCR, and the same results were obtained when commercially available cDNA libraries were used. Given that mRNAcontaining this exon would introduce a premature stop codon, we doubted whether such transcript could produce a protein product. In Western blot analysis of blood cell lysate from normal individuals, a band of 34 kDa,consistent in size with the molecular weight of the predicted truncated protein, was observed. Assays with cycloheximide also revealed that mRNA with this exon could evade NMD. FMRP is mainly localized in the cytoplasm, and can shuttle between the nuclear and cytoplasmic compartments because of the presence of NLS and NES. The truncated FMRP of translated lacks the carboxy-terminal domains of FMRP, including NES, the second KH domain and a RGG. In confocal immunofluorescence,the truncated protein displayed both nuclear and cytoplasmic localization in HEK293T and HeLa cells, while the full-length protein localized mainly in the cytoplasm. Overexpression of the truncated protein regulates FMRP downstream target genes in HEK293T cells, as shown by using RNA microarray analysis and quantitative real-time RT-PCR. Among the 117 regulated genes, the *SLC7A11* gene*,* solute carrier family 7 (anionic amino acid transporter light chain, xc- system), member 11 gene (also known as xCT), was the most prominent. This gene is highlighted for its function as cysteine/glutamate transporter in brain cells, as well as its role in mGluR stimulation. We termed this novel cryptic alternative exon the exon 9a of *FMR1* gene.

2989F

SOX10 Regulates an Alternative Promoter at the Charcot-Marie-Tooth Disease Locus *MTMR2*: Implications for a Nuclear Function in Schwann Cells. *E. A. Fogarty[1], M. H. Brewer[2], N. Steinberg[2], J. F. Rodriguez-Molina[3], W. D. Law[2], J. Svaren[4,5], A. Antonellis[1,2,6].* 1) Neuroscience Graduate Program, University of Michigan, Ann Arbor, MI; 2) Department of Human Genetics, University of Michigan, Ann Arbor, MI; 3) Cellular and Molecular Pathology Program, University of Wisconsin-Madison, Madison, WI; 4) Waisman Center, University of Wisconsin-Madison, Madison, WI; 5) Department of Comparative Biosciences, University of Wisconsin-Madison, Madison, WI; 6) Department of Neurology, University of Michigan, Ann Arbor, MI.

Schwann cells are the myelinating glia of the peripheral nervous system and dysfunction of these cells causes motor and sensory peripheral neuropathy. The transcription factor SOX10 is critical for Schwann cell development and maintenance, and many SOX10 target genes encode proteins required for Schwann cell function. Loss-of-function mutations in the gene encoding myotubularin related protein 2 (*MTMR2*) cause Charcot-Marie-Tooth disease type 4B1 (CMT4B1), a severe demyelinating peripheral neuropathy characterized by myelin outfoldings along peripheral nerves. Previous reports indicate that the expression of *MTMR2* is ubiquitous making it unclear how loss of this gene causes a Schwann cell-specific phenotype. To address this, we performed computational and functional analyses at *MTMR2* to identify transcriptional regulatory elements important for Schwann cell expression. Through these efforts, we identified an alternative, SOX10-responsive promoter at *MTMR2* that displays strong regulatory activity in Schwann cells. This promoter directs transcription of a previously unidentified *MTMR2* transcript that encodes an N-terminally truncated protein isoform, both of which are enriched in immortalized Schwann cells compared to motor neurons. The expression of the endogenous transcript is induced in a heterologous cell line by ectopically expressing SOX10, and is nearly ablated in Schwann cells by impairing SOX10 function. Intriguingly, while overexpressing the two MTMR2 protein isoforms revealed that both localize to puncta in the nucleus, the shorter isoform displays higher nuclear localization compared to the full-length protein. Combined, these findings suggest a nuclear function for MTMR2 that may be particularly important for Schwann cells. Indeed, the loss of this nuclear function may be a critical factor in the etiology of CMT4B1.

2990W

Inheritance of repeat interruptions in spinocerebellar ataxia type 10 (SCA10) expansions. *K. N. McFarland[1], I. Landran[1], J. Liu[1], C. Mulligan[2,3], A. Rasmussen[4], T. Ashizawa[1].* 1) Department of Neurology, McKnight Brain Institute, College of Medicine, University of FLorida Gainesville, FL 32610; 2) Department of Anthropology, College of Liberal Arts and Sciences, University of Florida, Gainesville, FL 32611; 3) Genetics Institute, University of Florida, Gainesville, FL 32611 USA; 4) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104 USA.

Spinocerebellar ataxia type 10 (SCA10), an autosomal dominant cerebellar ataxia with varying presentation of epilepsy, is caused by a non-coding ATTCT microsatellite repeat expansion in *ATXN10.* In a subset of SCA10 families, the expanded ATTCT repeat contains complex interruption motifs—the ATCCT interuption—at the 5' end of the expansion. This interruption region contains heptanucleotide motifs interspersed amongst stretches of pure ATTCT repeats. The presence of this repeat interruption correlates with an increased risk of epileptic seizures and greater overall expansion instability in intergenerational transmission, particularly in paternal lineages. In this study we examined the stability of the motif patterns of the ATCCT-interruption in these alleles across multiple generations of four SCA10 families by sequencing ATCCT-PCR products. Within the interruptions, we found the greatest instability occurred in stretches of pure ATTCT motifs while the interrupting sequences act as insulators against instability. The pure ATTCT stretches within this region are particularly unstable in paternal transmissions despite these the fact that these pure ATTCT stretches contain less than 20 ATTCT units, the number comparable to stable normal alleles. Overall, the change in length of the ATCCT interruption is only by one to three repeat units, which cannot explain the remarkable instability of the entire disease allele. We conclude that the AT-rich interruptions stabilize but do not abolish the instability of short stretches of pure ATTCT repeats, and that greater instability of SCA10 expansion alleles must occur outside of this region.

2991T

Rare Variant Identification in Saudi Children with Mendelian Neurologic Diseases. *W. Charng[1,2], E. Karaca[1,2], Z. Coban Akdemir[1,2], T. Gambin[1,2], M. M. Atik[1,2], S. N. Jhangiani[2,3], R. A. Gibbs[2,3], E. A. Faqeih[4], A. A. Asmari[4], M. A. M. Saleh[4], A. W. El-Hattab[5], J. R. Lupski[1,2,6,7].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston TX 77030, USA; 2) The Baylor-Hopkins Center for Mendelian Genomics, Houston, TX 77030, USA; 3) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA; 4) Section of Medical Genetics, Children's Hospital, King Fahad Medical City, Riyadh, Kingdom of Saudi Arabia; 5) Division of Clinical Genetics and Metabolic Disorders, Department of Pediatrics, Tawam Hospital, Al-Ain, United Arab Emirates; 6) Department of Pediatrics, Baylor College of Medicine, Houston, TX 77030, USA; 7) Texas Children's Hospital, Houston, TX 77030, USA.

In order to identify new disease genes and variants involved in the neurodevelopment, we performed whole exome sequencing (WES) on 35 Saudi families exhibiting neurological features, among which 32 are consanguineous. The majority of samples have been pre-screened with karyotyping and array CGH (aCGH) without significant findings. There are 26 (74. 3%) single affected subjects and 9 (24. 7%) families with multiple affected members. Four were sequenced as trios in families with single probands, and cases with two affected individuals per family were sequenced along with multiple members of the pedigrees. Twenty-two (63%) families exhibited various structural brain malformations while 13 (37%) presented either syndromic or non-syndromic developmental delay/intellectual disability (DD/ID) without any brain abnormalities. In addition to the single nucleotide variations (SNV), we also investigated copy number variants (CNV) predicted from WES data using Convex and HMZDelFinder (an in-house homozygous deletion detection tool at the resolution of single exons). With these tools, we are able to evaluate the genomic regions not covered well in the clinical array and/or to detect variations too small to be identified by the resolution afforded by conventional clinical aCGH. Currently, 9 (25. 7%) families are solved by new variants (single nucleotide variation or indel) in known genes (including *ARSA*, *ACO2*, *SARS2*, *VPS13B*, *PHF6*, and *RNASEH2B*), 5 (14%) may be explained by known genes with phenotypic expansion (including *KIAA0226*, *C12ORF57*, and *ANKRD11*), and 1 (2. 8%) family is solved by homozygous deletion in *GRID2* by clinical array and HMZDelFinder. We identified 17 potential novel candidate genes in 14 (40 %) families. These genes function as histone demethylase, motor proteins, transcription factors, myelin protein, as well as regulators for apoptosis and autophagy. Our genomic approach enabled us to evaluate potential causative SNVs and CNVs more accurately and lead to the identification of numerous interesting findings in genetics and genomics methodology. The genes that we identified function in a wide range of cellular processes involved in cell growth, cell differentiation, cell migration, proliferation, and apoptosis. *In silico* analysis of functional and expression correlation among candidate and known disease genes helped us better understand the genetic and signaling network underlying these neurological traits.

2992F

X-linked intellectual disability related genes disrupted by balanced X-autosome translocations. *M. Moyses-Oliveira[1,5], R. Guilherme[1], V. Meloni[1], A. Di Battista[1], C. Mello[2], S. Bragagnolo[1], D. Moretti-Ferreira[3], K. Popadin[5], N. Kosyakova[4], T. Liehr[4], G. Carvalheira[1], A. Reymond[5], M. Melaragno[1].* 1) Genetics Division, Department of Morphology and Genetics, Universidade Federal de São Paulo, São Paulo, Brazil; 2) Psychobiology Department, Universidade Federal de São Paulo, São Paulo, Brazil; 3) Departament of Genetics, Instituto de BioCiências de Botucatu, Universidade Estadual de São Paulo, São Paulo, Brazil; 4) Jena University Hospital, Friedrich Schiller University, Institute of Human Genetics, Jena, Germany; 5) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland.

Detailed molecular characterization of chromosomal rearrangements involving X-chromosome has been a key strategy in identifying X-linked intellectual disability-causing genes. Patients with balanced chromosomal rearrangements involving an X-chromosome generally present a skewed X inactivation, with the normal X-chromosome being preferentially inactivated. Thus, a gene disruption on the X-chromosome frequently results in the absence of a functional copy of the gene mapping to the breakpoint. We fine-mapped the breakpoints in four women with balanced X-autosome translocations and variable phenotypes to investigate the corresponding genetic contribution to intellectual disability. Three out of four patients presented with cognitive impairment, confirming the association between the disrupted genes (*TSPAN7*, *KIAA2022* and *IL1RAPL1*) and intellectual disability (MRX58, MRX98 and MRX21/34 loci, respectively). RT-qPCR experiments in whole blood showed absence of *TSPAN7* and *KIAA2022* functional copies in patients while gene expression was detected in eight controls. Although *IL1RAPL1* gene is not expressed in peripheral blood under normal conditions (e. g. no expression was observed in eight controls), transcripts from exons 8 and 9 of this gene were detected in the patients' whole blood, suggesting that the der(X) is producing a fusion transcript *ZNF611-IL1RAPL1* under the control of the *ZNF611* promoter, a gene disrupted at the autosomal breakpoint. The X-chromosomal breakpoint definition in the fourth patient, a 42 years old woman with normal intellectual abilities (WAIS III indicated IQ score = 111), revealed disruption of the *ZDHHC15* gene (MRX91). The RT-qPCR in whole blood did not detect *ZDHHC15* gene expression in the patient and detected the gene expression in eight controls, thus exposing the necessity for validation of *ZDHHC15* as an intellectual disability-causing gene. Overall, this study emphasizes the precise breakpoints definition in balanced X-autosome rearrangements as an important tool for searching critical genes and correct diagnosis of the individual patients. The disruption of X-linked genes in patients with cognitive impairment may confirm the involvement of the corresponding genes in neurodevelopment. Furthermore, the breakpoints fine mapping in subjects with normal cognition might also lead to a better characterization of candidate genes for diseases.

2993W

Identification and functional characterization of *de novo FOXP1* variants in cases of autism, intellectual disability and language impairment. *E. Sollis[1], A. Vino[1], C. Gilissen[2], H. Frohlich[3], S. Graham[1], R. Pfundt[2], D. Dimitropoulou[1], H. Brunner[2,4], G. Rappold[2], S. E. Fisher[1,5], P. Deriziotis[1].* 1) Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, the Netherlands; 2) Department of Human Genetics, Radboud Institute for Molecular Life Sciences and Donders Centre for Neuroscience, Radboud University Medical Center, Nijmegen, the Netherlands; 3) Department of Molecular Human Genetics, Ruprecht-Karls-University, Heidelberg, Germany; 4) Department of Clinical Genetics, Maastricht University Medical Centre, Maastricht, the Netherlands; 5) Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, the Netherlands.

Heterozygous variants disrupting the FOXP1 transcription factor (forkhead box protein P1; OMIM 605515) have been implicated in an emerging syndrome of global developmental delay, intellectual disability and autistic features (OMIM 613670). All etiological *FOXP1* variants reported to date have been *de novo*, and include whole gene deletions, translocations, nonsense variants, missense variants and frameshift variants. Speech and language impairments, which range from moderate to severe and affect expressive language to a higher degree than receptive language, are also consistently reported in individuals carrying *FOXP1* variants. The presence of speech and language deficits in this *FOXP1*-deficiency syndrome is of particular interest, because *FOXP1* is the closest paralogous gene to *FOXP2* (OMIM 605317), which is disrupted in a rare form of speech and language disorder (OMIM 602081). In this study, we report three novel heterozygous *de novo FOXP1* variants identified through clinical whole-exome sequencing in individuals diagnosed with intellectual disability and/or autism spectrum disorders. We performed detailed functional characterization of the three novel variants found here alongside three previously detected *de novo* etiological variants. Our assays examined the effects of the six *de novo FOXP1* variants on several aspects of protein function, including subcellular localization and transcriptional repression capability. In addition, because FOXP1 and FOXP2 are both expressed in the striatum where they may interact with the potential to co-regulate downstream targets, we investigated the effects of the six *de novo FOXP1* variants on FOXP1-FOXP2 interactions. Our findings provide further insight into the pathogenic mechanisms of *FOXP1*-deficiency syndrome, and highlight the value of combining clinical exome sequencing with functional analyses in understanding the molecular basis of neurodevelopmental disorders.

2994T

Two patients diagnosed with BPAN in infancy. *K. Takano[1], N. Shiba[2], K. Goto[3], T. Yamaguchi[1], K. Wakui[1], T. Kosho[1], Y. Inaba[2], Y. Fukushima[1].* 1) Department of Medical Genetics, Shinshu University School of Medicine, Matsumoto, Nagano, Japan; 2) Department of Pediatrics, Shinshu University School of Medicine, Matsumoto, Japan; 3) Department of Pediatrics, NHO Nishibeppu National Hospital, Beppu, Japan.

Beta-propeller protein-associated neurodegeneration (BPAN) is newly established as a subtype of neurodegeneration with brain iron accumulation (NBIA). BPAN is caused by mutations in the *WDR45* gene at Xp11. 23, which has an important role in autophagy. It has reported that both heterozygous female and hemizygous male are equally affected by this X-linked NBIA. BPAN is characterized by global developmental delay or intellectual disability until adolescent or early adulthood, followed by severe dystonia, parkinsonism and progressive dementia. Brain magnetic resonance images (MRI) of BPAN show iron deposition in the bilateral globus pallidus (GP) and substantia nigra (SN), hyperintensity of the SN with a central band of hypointensity in T1-weighted images and brain atrophy in adulthood. It is very difficult to diagnose BPAN in early childhood before the development of neurological deterioration. However, next-generation sequencing (NGS) technologies allow us early diagnosis of BPAN and have revealed early manifestations of BPAN, such as Rett-like phenotype and early iron accumulation in the GN and SN. We report clinical features of 2 unrelated patients, 3-year-old girl and 3-year-old boy, who were diagnosed with BPAN by targeted NGS. Both patients presented with severe developmental delay and moderate elevated level of serum aspartate transaminase (AST) and neuron specific enolase (s-NSE). The elevation of serum enzymes may be one of characteristic symptoms of BPAN in infancy.

2995F

Comprehensive molecular diagnosis of 265 individuals addressed for Leber congenital amaurosis or Early Onset Severe Retinal Dystrophy using targeted next generation sequencing. *I. Perrault[1], S. HANEIN[2], C. BOLE[3], P. NITSCHKE[4], N. DELPHIN[5], A. MUNNICH[5], J. KAPLAN[1], JM. ROZET[1].* 1) GENETICS IN OPHTHALMOLOGY, INSTITUT IMAGINE, INSERM UMR 1163, PARIS, FRANCE; 2) NGS MOLECULAR DIAGNOSIS,INSTITUT IMAGINE,INSERM UMR1163, PARIS, FRANCE; 3) GENOMICS PLATFORM,INSTITUT IMAGINE, INSERM UMR1163, PARIS, FRANCE; 4) BIOINFORMATIC PLATFORM,INSTITUT IMAGINE, INSERM UMR1163, PARIS, FRANCE; 5) GENETICS, INSTITUT IMAGINE, INSERM UMR1163, PARIS, FRANCE.

Introduction: Leber congenital amaurosis is the earliest and most severe retinal dystrophy, and a leading cause of blindness in children. The visual outcome of affected infants is variable, ranging from light perception, to low but measurable visual acuity in the first two decades of life. Blind infants with LCA are at risk of developing skeletal, neurologic and renal dysfunctions. Both the visual and extraocular outcomes strongly correlate with the disease-gene. The study presented here aimed at assessing targeted NGS (T-NGS) as a tool to improve patient care by allowing efficient and early molecular diagnosis in infants with severe visual deficiency. Methods: 265 index LCA cases were sequenced using a T-NGS array comprising 45 genes causing isolated or syndromic LCA and 10 genes of differential diagnoses. Mutations were confirmed by Sanger sequencing and familial segregation analysis. Results: We identified causative mutations, including 6 copy number variations, in 150/265 index cases, 119/150 of whom had mutations in genes for isolated LCA/EOSRD and 19/150 in genes for syndromic forms. In addition, we identified convincing mutations in genes for differential diagnoses in 12/265 individuals. Conclusions: T-NGS molecular diagnosis proved powerful to improve the care of infant with severe visual dysfunction by allowing *i)* early identification of children with differential diagnosis which outcome is highly favorable compared to LCA (12/265; <5%) and, *ii)* early discrimination of children at no risk of being affected with syndromic LCA (119/150; 79%) from those, far fewer, at risk of developing extraocular symptoms (19/150; 12%) who require extraocular explorations.

2996W

Molecular analysis and prevalence of Huntington's Disease in northwest of Iran. *M. Shekari khaniani[1], P. Aob[2], M. r. Ranjouri[2], S. Mansouri derakhshan[1].* 1) Medical Genetic Department, Faculty of Medicine, Tabriz University of Medical Sciences, Tabriz, Iran; 2) Medical Genetic Department, Faculty of Medicine, Zanjan University of Medical Sciences, Zanjan, Iran.

Background: Huntington Disease (HD) is a progressive adult-onset neurodegenerative disorder presenting an autosomal dominant inheritance. Onset of symptoms is in middle-age after affected individuals have had children, but the disorder can manifest at any time between infancy and senescence. HD is clinically characterized by involuntary chorea form movements, cognitive impairment and personality disorder tending to depression, anger and temper outbursts. **Objective:** Since there is no information on the frequency of HD disorder in northwest of Iran, the aim of the present study is to determine the frequency of Huntington disease and the number of CAG-trinucleotide repeats in the northwest of Iranian population. **Material and Methods:** Genomic DNA was extracted from the blood samples by salting-out. DNA samples were analyzed to determine the number of CAG trinucleotide repeats of HD gene. An abnormally large number of CAG repeats which is a diagnostic factor for the disease was detected by polymerase chain reaction (PCR) and agarose gel electrophoresis. **Results:** Of 40 cases, we identified 14 unrelative individuals with one expanded CAG allele at the IT15 gene. The frequency of the *HD* mutation in our group of patients was 35% (14/40). Expanded alleles varied from 36 to 70 CAG repeats (mean = 53 CAG units) and normal alleles in HD patients varied from 20 to 26 CAG units (mean = 23 CAGs). There was no significant difference in the size of the expanded alleles between affected males and affected females. **Discussion:** to our knowledge, This is the first study reporting molecular testing of HD patients in the Northwest of Iran. We found a significant correlation between age at onset of the disease and length of the expanded CAG tract. This indicates a tendency for age at onset to decrease as the CAG repeat length increases. Our results showed that not all patients with the "HD" phenotype carried the expansion at the IT15 gene and that autosomal dominant inheritance may not be clearly documented in all HD families.

2997T

Evidence that variants in *PIGG* cause intellectual disability with early onset seizures and hypotonia. *P. Makrythanasis[1,2], M. Kato[3], M. S. Zaki[4], H. Saitsu[5], K. Nakamura[3], F. A. Santoni[1], S. Miyatake[5], M. Nakashima[5], I. Mahmoud[4], M. Guipponi[1,2], H. Hamamy[1], N. Matsumoto[5], T. Kinoshita[6], S. E. Antonarakis[1,2,7], Y. Murakami[6].* 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Geneva, Switzerland; 2) Service of Genetic Medicine, University Hospitals of Geneva, Geneva, Switzerland; 3) Department of Pediatrics, Yamagata University Faculty of Medicine, Yamagata, Japan; 4) Department of Clinical Genetics, National Research Centre, Cairo, Egypt; 5) Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan; 6) Department of Immunoregulation, Research Institute for Microbial Diseases, and World Premier International Immunology Frontier Research Center Osaka University, Osaka, Japan; 7) iGE3, University of Geneva, Geneva, Switzerland.

Glycosylphosphatidylinositol (GPI) is a glycolipid that anchors 150+ proteins to the cell surface. There are at least 27 genes involved in the biosynthesis and transport of GPI anchored proteins. Among them, 12 different mendelian GPI deficiencies were reported up to now; mainly inherited as recessive traits. These deficiencies show intellectual disability, epilepsy, coarse facial features and multiple organ anomalies that vary in severity depending upon the molecular defect of the affected genes. Those symptoms are caused either by the decreased surface expression of GPI anchored proteins or by structural abnormalities of the GPI. Here we present two families, a consanguineous from Egypt and another from Japan in which the patients' main symptoms consisted of intellectual disability, early onset seizures and hypotonia. After exome sequencing, variants in a "new" gene in the GPI pathway, *PIGG* were identified. In the first family, it consisted of the homozygous nonsense NM_001127178. 2:c. 1515G>A: p. (Trp505*), while in the second family the patient was a compound heterozygote of a known rare missense variant, the NM_001127178. 2:c. 2005C>T, p. (Arg669Cys) that was predicted to be pathogenic with a 2. 4Mb deletion that involved PIGG. In both cases the parents were confirmed to be healthy carriers. *PIGG* is the enzyme which attaches ethanolamine phosphate to the second mannose. This ethanolamine phosphate is cleaved after GPI anchor is attached to the protein in the endoplasmic reticulum by PGAP5. In vitro studies using patients' B lymphoblasts, in which PIGG activity was almost completely abolished, GPI-APs were normally expressed on the surface and with normal structure, indicating that pathogenesis of PIGG deficiency is potentialy novel, not yet fully understood. The discovery of pathogenic variants in PIGG expands the spectrum of GPIpathies, and further enhances our understanding of this etiopathogenic class of intellectual disability.

2998F

Contribution of excitatory/inhibitory synaptic imbalance to *MECP2*, *CDKL5* and *FOXG1* related disorders. *E. Landucci[1], S. Amabile[1], T. Patriarchi[1, 2], E. Frullanti[1], A. M. Pinto[1, 3], C. Lo Rizzo[1], F. Ariani[1, 3], F. Mari[1, 3], M. A. Mencarelli[3], J. W. Hell[2], A. Renieri[1, 3], I. Meloni[1].* 1) Medical Genetics University of Siena, Siena, Italy; 2) Department of Pharmacology, University of California, Davis, CA, USA; 3) Genetica Medica, Azienda Ospedaliera Universitaria Senese, Siena, Italy.

MECP2, CDKL5 and *FOXG1* genes are responsible for the Rett spectrum of neurodevelopmental disorders. In spite of their involvement in the same phenotypic spectrum, a functional interaction between the three proteins has not been proven and disease mechanisms remain elusive. We recently established a human neuronal model based on patients-specific iPSCs and we characterized their expression profile to test the hypothesis that common alterations occur in cells mutated in the three genes. An excitation/inhibition imbalance has been suggested in *MECP2*-related Rett syndrome. In line with this finding, we recently reported that expression of the inhibitory presynaptic differentiation inducer GluD1 is increased in iPSCs-derived neurons from patients with mutations in *MECP2* and *CDKL5* (Livide et al, Eur J Hum Genet 2015). Here we describe an over-expression of GluD1 also in *FOXG1*-mutated cells. We also analyzed the expression of a panel of excitatory (VGLUT1, GluA1, GluN1, PSD-95) and inhibitory markers (GAD67 and GABA AR-α1) in iPSCs-derived neurons mutated in the three genes, demonstrating an excitation/inhibition imbalance. A consistent imbalance was also observed in embryonic FOXG1+/- mouse brain. However, a reduction of both excitatory and inhibitory markers is observed in FOXG1+/- post-natal mouse brain. This would suggest that an increase in inhibitory over excitatory synapses/neurons during early brain development might later on trigger counteracting mechanisms resulting in the loss of either inhibitory neurons or synapses. In conclusion, our data underline the essential contribution of an excitation/inhibition imbalance in *MECP2, CDKL5* and *FOXG1* related disorders, opening up the possibility of new therapeutic approaches.

2999W

Functional characterization of a large series of *NKX2-1* variants in Brain-Lung-Thyroid syndrome reveals diverse molecular mechanisms of disorder. *S. A. Graham[1], P. Deriziotis[1], R. Bojoh[1], S. B. Estruch[1], S. E. Fisher[1,2].* 1) Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen 6525XD, The Netherlands; 2) Donders Institute for Brain, Cognition and Behaviour, Nijmegen 6525EN, The Netherlands.

Brain-Lung-Thyroid syndrome (MIM 610978) is an autosomal dominant disorder resulting from disruption of the homeobox transcription factor gene *NKX2-1* (MIM 600635). The disorder exhibits phenotypic variability, both between and within families, but is typically characterized by infant hypotonia that progresses to chorea and other movement abnormalities. Dysarthria (a motor speech disorder) and attentiveness disorder have recently been recognized as common neurological features of the syndrome. Most affected individuals also have thyroid dysfunction and/or lung problems such as neonatal respiratory distress and recurrent infections. Over 100 different lesions affecting the *NKX2-1* gene have been reported in cases of Brain-Lung-Thyroid syndrome, including a number of missense variants. However there has been little functional investigation of the effects of these variants to elucidate the molecular mechanisms of disorder. Here we report the functional characterization of forty *NKX2-1* variants, employing assay methodologies that could be extended for the high-throughput characterization of transcription factor gene variants identified by next-generation sequencing in disorders. By assessing the effects of *NKX2-1* variants on protein expression, subcellular localization, protein-protein interactions and transcriptional regulatory activity, we uncovered diverse molecular-level effects for etiological variants, highlighting the importance of examining multiple aspects of protein function when characterizing putative disorder-related variants. In addition, we confirmed that NKX2-1 interacts with the forkhead transcription factor FOXP2 (MIM 605317) and show that NKX2-1 can also interact with the FOXP2 paralogs FOXP1 (MIM 605515) and FOXP4 (MIM 608924). The interaction between NKX2-1 and FOXP transcription factors may be of significance in brain and lung development. Heterozygous disruption of *FOXP2* results in a severe speech/language disorder (MIM 602081). The presence of motor speech deficits as core features of both the *NKX2-1* and *FOXP2*-related disorders points to roles for both genes in the development of speech-related motor circuitry. Strikingly, we find that disorder-related variants in the DNA-binding domains of both NKX2-1 and FOXP2 abrogate the interaction between these proteins.

3000T

Pontine Tegmental Cap Dysplasia: a rare hindbrain malformation and the search of the genetic etiology. *M. Herlin[1], K. Varvagiannis[2], L. Gammelgaard[3], P. A. Gregersen[1], P. Makrythanasis[2], S. E. Antonarakis[2], M. B. Petersen[1].* 1) Department of Clinical Genetics, Aalborg University Hospital, Aalborg, Denmark; 2) Department of Genetic Medicine & Development, University of Geneva, Geneva, Switzerland; 3) Department of Radiology, Viborg Regional Hospital, Viborg, Denmark.

Pontine Tegmental Cap Dysplasia (PTCD) is a rare congenital malformation of the hindbrain characterized by hypoplasia and flattening of the ventral pons, vaulted pontine tegmentum, and vermian and cerebellar hypoplasia. The typical clinical findings include multiple cranial neuropathies, particularly sensorineural hearing loss, ataxia, and seizures. To date, only 25 cases have been reported in the literature and the cause still remains unknown. We present a case of a 14-year-old girl who, throughout her life, has been a diagnostic challenge displaying symptoms and clinical findings such as delayed developmental milestones, bilateral abducens palsy (Duane syndrome), congenital sensorineural hearing loss, ataxia, and butterfly configuration of the Th3-Th4 vertebrae. Due to inconclusive MRIs at 2 years of age, she was diagnosed with Wildervanck syndrome based on the clinical findings and shortly after she received bilateral cochlear implants, complicating further radiological examinations. However, a recent reevaluation of the MRIs found her to have malformations of the brainstem and cerebellum compatible with PTCD. Currently, we are investigating the genetic cause of PTCD in our case. CGH array revealed no copy number variations and thus, sequencing of all coding regions (Whole Exome Sequencing) has been performed on DNA from the patient and her parents. Variant analysis is still ongoing but it is our hope that data exchange through international collaborations will enable us to delineate the gene(s) associated with PTCD.

3001F

Loss of function variants highlight potential candidate genes in patients with brain malformation and epilepsy. *E. Karaca[1], T. Harel[1], D. Pehlivan[1], Z. Coban Akdemir[1], Y. Bayram[1], V. Topcu[2], G. Yesil[3], S. Tug Bozdogan[4], O. Ozalp Yuregir[5], H. Aslan[6], H. Aydin[7], T. Tos[8], T. Gambin[1], M. M. Atik[1], S. N. Jhangiani[9], D. Muzny[9], E. Boerwinkle[9], R. A. Gibbs[9], J. R. Lupski[1,9,10,11], Baylor-Hopkins Center for Mendelian Genomics.* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston 77030, TX, USA; 2) Department of Medical Genetics, Zekai Tahir Burak Woman Health Training and Research Hospital, Ankara 06230, Turkey; 3) Department of Medical Genetics, Bezmialem University, Istanbul 34093, Turkey; 4) Department of Medical Genetics, Mersin University, Mersin 33343, Turkey; 5) Department of Medical Genetics, Numune Training and Research Hospital, Adana, Turkey; 6) Department of Medical Genetics, Medical Faculty of Eskisehir Osmangazi University, Eskisehir 26480, Turkey; 7) Department of Medical Biology, Namik Kemal University School of Medicine, Tekirdag 59100, Turkey; 8) Department of Medical Genetics, Sami Ulus Children's Hospital, Ankara 06080, Turkey; 9) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; 10) Department of Pediatrics, Baylor College of Medicine, Houston, TX, USA; 11) Texas Children's Hospital, Houston, TX, USA.

Human brain development is a precisely orchestrated process involving multiple genetic and epigenetic interactions and coordination of cellular and molecular mechanisms, perturbation of which leads to a multitude of neurodevelopmental phenotypes depending on the spatial and temporal effect of the disturbance. We used whole exome sequencing in a cohort of 125 mostly consanguineous families with both functional and structural brain malformations as a primary experimental approach to explore the potential genetic basis of brain development and function. Variant alleles were prioritized using a recessive inheritance model, bioinformatic prediction tools of potential pathogenicity, and co-segregation of the allele with the disease. In silico analysis of the gene for brain developmental, expression, interactome and pathway analysis of gene products, further prioritized variants potentially associated with the Mendelizing traits studied. Analysis of the data for both single nucleotide variants (SNVs) and copy number variants (CNVs) identified 40 candidate disease genes, 44 novel variants in known disease genes, and de novo or homozygous deletions in 10 families, with an overall solved rate of >85%. Among these, we observed rare homozygous loss of function variants in UBQLN1, SMARCA1, AGBL2, SNX14 and CPLX1 in patients with various severities of forebrain and hindbrain abnormalities and epilepsy. The respective encoded proteins are highly expressed in the nervous system and involved in crucial pathways such as ubiquitination machinery, SWI/SNF-like chromatin remodeling complex, posttranslational modification of tubulins and synaptic transmission. Thus, our study underscores the utility of small pedigree analysis in genetically heterogeneous disease.

3002W

DYT16/*PRKRA* founder mutation causes childhood-onset generalized dystonia in a family from Southern Italy. *M. Quadri[1], S. Olgiati[1], M. Sensi[2], F. Gualandi[3], E. Groppo[4], V. Rispoli[4], J. Graafland[1], G. J. Breedveld[1], G. Fabbrini[5], V. Bonifati[1].* 1) Departement of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands; 2) Department of Neurology and Rehabilitation, University -Hospital Arcispedale S. Anna, Ferrara, Italy; 3) Department of Reproduction And Growth, UOL of Medical Genetics, Universiy Hospital S. Anna, Ferrara, Italy; 4) Department of Biomedical and Specialistic Surgical Sciences, University-Hospital Arcispedale S. Anna, Ferrara, Italy; 5) Department of Neurology and Psychiatry, Sapienza University of Rome, and IRCSS Neuromed, Pozzilli, Italy.

Dystonias are movement disorders characterized by sustained muscle contractions and abnormal postures. Based on both genetic and clinical evidences, the dystonias are an heterogeneous group of diseases. Twenty-five DYT genetic loci have been described, so far. Clinically, the disorder can occur in isolation or in combination with other neurological signs. Onset age is variable, and the disease can affect different parts of the body. Here we report three brothers originating from a small village in Southern Italy with childhood-onset generalized dystonia and mild parkinsonism. We performed high-density SNP-arrays genotyping in the whole family, and high-coverage whole exome sequencing (WES) in the proband. The linkage analysis under an autosomal recessive mode of inheritance, and assuming distant parental consanguinity, yielded only one locus with a significant LOD of 3. 72. This locus lies on the long arm of chromosome 2 and contains a 4Mb-homozygous region. Within this region, WES uncovered a *PRKRA* c. 665C>T p. Pro222Leu mutation. Sanger sequencing confirmed the mutation in homozygous state in the three affected siblings, while the unaffected parents and the unaffected sister were found to be heterozygous carriers. The same mutation has been initially associated with early-onset dystonia-parkinsonism in two unrelated consanguineous families and one sporadic case from Brazil, and this locus was designated as DYT16. Very recently, the same and additional mutations were reported in three independent pedigrees of Polish, German and Brazilian ancestries, providing conclusive evidence for a disease-causing role of *PRKRA* mutations in DYT16. Our haplotype analysis shows that our Italian patients share a 0. 5 Mb identical homozygous region with all the nine DYT16 cases previously reported with the same mutation. These results support the contention that c. 665C>T is a founder mutation, as already suggested. Inspection of the Exome Aggregate Consortium (ExAC) shows that this mutation is present, although rarely, in several populations. Mutations in this gene might be more common than earlier thought, and *PRKRA* screening is warranted in all patients with dystonia compatible with autosomal-recessive inheritance. Our data contribute to a better delineation of the DYT16 genetic and clinical spectra.

3003T

Expanding the Phenotype of Type XII Collagen Myopathy: Congenital Cataracts, Microcephaly, and Global Developmental Delay? *G. E. Tiller[1], A. N. Filose[1], Y. Hu[2], C. G. Bonnemann[2].* 1) Dept Genetics, Kaiser Permanente, Los Angeles, CA; 2) NIH, National Institute of Neurological Disorders and Stroke, Bethesda, MD.

Hereditary congenital myopathies may be caused by mutations in genes encoding several classes of proteins, including extracellular matrix proteins, contractile proteins, and Golgi-resident modifying enzymes. The archetypal extracellular matrix proteins involved in congenital myopathies are the type VI collagens, which are defective in Bethlem and Ullrich myopathies. Recently, type XII collagen has been recognized as involved in a phenotypically similar congenital myopathy, and both autosomal dominant and recessive forms have been described. Here we report two siblings with a congenital myopathy who are compound heterozygotes for missense mutations in COL12A1, the gene encoding type XII collagen. As neonates they presented with hypotonia, which evolved to spasticity with hyperreflexia, finger flexion contractures, and inability to ambulate. Bilateral congenital cataracts were evident in both sibs, and one underwent unilateral phakectomy. Head circumferences at birth were normal but by age 3 years were below the 3rd centile; brain MRI at age one year in one sib was unremarkable. Both sibs, currently ages 3 and 8 years, are nonverbal. Whole exome sequencing revealed missense mutations in COL12A1 predicted to result in the amino acid substitutions Pro591Arg (maternal) and Ser2042Leu (paternal). Neither parent had a history of motor or cognitive delay or cataracts. Cultured skin fibroblasts from one sib exhibited decreased intracellular immunoreactivity and absent extracellular immunoreactivity against type XII collagen. In the absence of other exomic data to explain cataract formation, brain atrophy, and global developmental delay, we propose that compound heterozygosity for missense mutations in COL12A1 may play a role in the development of these additional clinical findings.

3004F

Somatic Mutations in mTOR and Tau pathways in Focal Cortical Dysplasia. *F. Torres[1], M. G. Mazutti[1], P. A. O. Ribeiro[1], S. H. Avansini[1], R. Secolin[1], M. G. Borges[1], F. Rogério[2], L. S. Queiroz[2], A. C. Coan[3], H. Tedeschi[3], F. Cendes[3], I. Cendes[3].* 1) Department of Medical Genetics, Unicamp, Campinas, Brazil; 2) Department of Pathology, Unicamp, Campinas, Brazil; 3) Department of Neurology, Unicamp, Campinas, Brazil.

Somatic mosaic mutations have been observed in malformations of cortical development (MCD), such as tuberous sclerosis (TS) and hemimegalencephaly (HMG). HMG, TS and focal cortical dysplasia (FCD) share histopathologic characteristics and aberrant expression of genes belonging to the mTOR and Tau pathways. These findings led us to investigate if somatic mosaic mutations in genes related to the mTOR and Tau pathways are present in FCD. Deep-Whole-Exome sequencing was performed in genomic DNA extracted from brain tissue resected by surgery (BTRS) and blood of four patients with FCD. We performed capture and enrichment with Nextera® Expanded Kit (Illumina®), samples were sequenced following a 200bp paired-end protocol in a Hiseq2500 (Illumina®) to achieve at least 200x of average coverage. Realignment around indels and SNP clusters were performed using the Genome Analysis ToolKit (GATK) realigner. Our analysis focused in genes belonging to the mTOR and Tau pathways. Only exons with at least 70x coverage were analyzed. We identified a total of 12 mosaic variants in mTOR genes and four in Tau pathway genes. These mutations are present only in DNA from brain tissue. They are localized in different genomic regions such as 3'UTR, intergenic regions, introns and splicing sites. One of the major genes of interest is *ULK2,* which is involved in axonal development. In conclusion, we have demonstrated that somatic mosaicism in genes related to mTOR and Tau pathways is present in FCD.

3005W

Clinical and Genetic Analysis of 87 Chinese Patients with Pelizaeus-Merzbacher Disease. *J. Wang, H. Ji, D. Li, Y. Wu, Q. Gu, Y. Yang, H. Xiong, Y. Zhang, X. Bao, X. Wu, Y. Jiang.* Pediatrics, Peking University First Hospital, Beijing, Beijing, China.

Pelizaeus–Merzbacher disease (PMD, OMIM 315080) is a rare X-linked recessive genetic disease with diffuse hypomyelination in white matter. Clinical manifestations include nystagmus, development retardation, hypotonia, spastic paraplegia and ataxia. Three different types of PMD have been identified: congenital, transitional, and classical PMD. PMD is caused by *PLP1*(proteolipid protein 1) mutation and it has 3 kinds of mutations: duplication, point mutation and deletion. **Objective** Identify *PLP1* mutations in 87 Chinese patients with PMD, offer genetic counseling and make prenatal molecular diagnoses for 31 families. **Method** Extract genomic DNA from peripheral blood of probands and their families. The fetus' DNA was extracted from Amniotic fluid/chorionic villus samples. We use MLPA (multiplex ligation-dependent probe amplification) to detect *PLP1* duplication/deletion mutations; then detect point mutation by Sanger sequencing for those with negative MLPA result. **Results** We found *PLP1* mutation in 87 cases. Among them, 64 patients are *PLP1* duplications (73. 6%) and 23 are point mutations (26. 4%) including 10 novel mutations. 87(100%) are male and have global development delay, 86(98. 9%) with nystagmus, 53(60. 9%) with hypotonia. All patients present hypomyelination on their brain MRI. 43(49. 4%) patients are classical, 23(26. 4%) are transitional, and 21(24. 1%) are congenital PMD. We preformed 31 prenatal diagnoses in 27 families, 17 cases are *PLP1* wildtype, 9 cases have same mutations as the probands', and 5 are *PLP1* mutation carriers. **Conclusion** Classical type is the most common subtype of PMD, while duplication is the most common kind of *PLP1* mutation. Duplication could be found in classical and transitional PMD while point mutations in congenital and transitional PMD. We have identified *PLP1* mutations in 87 PMD patients, in which 10 are novel mutations; and we also performed 31 prenatal diagnoses for 27 families, which will be useful and helpful. This research was supported by the grants from "973" Project of the Science and Technology Ministry of China (No. 2012CB944602), National Key Research Project "12-5"(No: 2012BAI09B04), National Natural Science Foundation of China (81271257, 81200873), Beijing Natural Science Foundation (7132208), Beijing Key Laboratory.

3006T

Systematic phenotype-based deconvolution of Intellectual Disability disorders into biologically coherent modules. *C. Zweier[1], K. Kochinke[2], B. Nijhof[2], M. Fenckova[2], P. Cizek[3], F. Honti[4], S. Keerthikumar[3], M. A. W. Oortveld[2], T. Kleefstra[2], J. M. Kramer[2], C. Webber[4], M. A. Huynen[3], A. Schenck[2].* 1) Institute of Human Genetics, Friedrich-Alexander-Universitét Erlangen-Nürnberg, Erlangen, Germany; 2) Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud university medical center, Nijmegen, the Netherlands; 3) CMBI, Radboud Institute for Molecular Life Sciences, Radboud university medical centre, Nijmegen, the Netherlands; 4) MRC Functional Genomics Unit, Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom.

Intellectual Disability (ID) disorders, due to their frequency and enormous genetic and phenotypic heterogeneity, represent a major unmet challenge in health care and diagnostics. A comprehensive and systematic understanding of ID disorders and their underlying biology is still limited. We established a curated database of 650 genes, currently known to be mutated in ID, linked them to various functional datasets and classified them according to ID manifestation and severity and 27 associated core clinical features such as organic or neurological anomalies. This catalog also contains genes mutated in autism spectrum disorders that are associated with cognitive impairment. Mutations in 62% of all 650 ID genes follow autosomal recessive inheritance, mutations in 21% are autosomal dominant (mainly de novo), and 16% are X-linked. Using this integrated resource we show that nearly half of all ID proteins physically interact with other ID proteins and that they show a more than 30% increase in connectivity. Furthermore, they are substantially enriched in co-expression, highest in the hippocampus. 86% of ID genes fall into 32 common Gene Ontology based molecular processes: metabolism and nervous system development among the largest, and hedgehog and glutamate signaling among the most enriched groups. Identification of highly enriched functional themes and phenotypes systematically revealed characteristic phenoprofiles of process-defined IDopathies including chromatin- and DNArepairopathies. Strikingly, all these findings are much more pronounced among subgroups of genes that share associated phenotypes, adding significant predictive power for ID gene identification. Furthermore, we utilized custom-made datasets on ID gene function in *Drosophila*. Early onset behavioral and specific morphological wing phenotypes were characteristic for ID genes in general, and several fly phenotypes were particularly representative for specific human clinical phenotype classes. Our study and resource provide systematic insights into the molecular and clinical landscape of ID disorders and prove the utility of systematic human and cross-species phenomic analyses in highly heterogeneous genetic disorders.

3007F

Synergistic activity of the DYT6-associated THAP1 protein and HCFC1 in regulating gene expression. *L. Kötter[1], R. Hollstein[1], B. Reiz[2], S. Schaake[3], K. Lohmann[3], FJ. Kaiser[1].* 1) Section for Functional Genetics at the Institute of Human Genetics, University of Lübeck, Lübeck, Germany; 2) Institute for Integrative and Experimental Genomics, University of Lübeck, Lübeck, Germany; 3) Institute of Neurogenetics, University of Lübeck, Lübeck, Germany.

THAP1, the gene mutated in DYT6 dystonia, encodes a transcription factor protein. While the THAP domain within the N-terminal part of the protein allows specific DNA-binding the functional relevance of the remaining part of THAP1 is largely unknown. Beside a coiled-coil domain and a nuclear localization signal at its C-terminus a stretch of four amino acids (134-137) within the 218 amino acids spanning protein was shown to be relevant for the interaction with the transcription regulator HCFC1. Interestingly, three mutations (Asn136Ser, Asn136Lys, Y137C) have been reported within this motif in DYT6 patients. By protein-protein interaction analyses we demonstrated that these substitutions within this HCFC1-binding domain (HBM) of THAP1 abolish HCFC1-THAP1 complex formation. Notably, HCFC1 co-localization was observed in >90% of the >3,500 chromatin regions loaded with THAP1 in publically available genome-wide ChIP data (ENCODE). To characterize the interaction between HCFC1 and THAP1 in the regulation of gene expression, we performed siRNA-mediated depletion of *HCFC1* or *THAP1*. By this, we detected similar dysregulation of THAP1 target genes in cells with HCFC1- or THAP1 downregulation-indicating synergistic activity of both transcription factors. To examine whether THAP1 mutations that abolish HCFC1-binding do affect THAP1 or HCFC1 recruitment to gene promoters, we used quantitative ChIP-seq on selected promoters. While none of the three THAP1 mutations significantly modified DNA-binding ability of THAP1, HCFC1 was strongly reduced. These findings suggest a THAP1-mediated recruitment of HCFC1 to THAP1 target sites further supporting an interplay of both transcription factors. In a final step, we investigated whether mutations in the THAP1 interacting domains of HCFC1 are a cause of dystonia. For this, we screened 160 dystonia patients but did not detect any clear mutation in *HCFC1*. Of note, mutations in *HCFC1* have been reported in X-linked intellectual disability (XLID) and Cobalamin type X (CblX), a VitaminB12 metabolic disorder with severe neurodevelopmental impairment. In conclusion, we demonstrate that THAP1 recruits HCFC1 to THAP1 target promoters and that both proteins are necessary for the transcriptional regulation. Importantly, DYT6-causing mutations within the 4-amino-acid interaction domain in THAP1 abolish the interaction. Thus, we highlight disturbed HCFC1 interaction as the consequence of three different DYT6-causing mutations.

3008W

***FOXC1* mutations and copy number variants in cases diagnosed with Primary Congenital Glaucoma.** *E. Souzeau[1], O. M. Siggs[1], B. Ridge[1], L. Mauri[3], M. Hayes[2], A. Dubowsky[2], F. Pasutto[4], J. E. Craig[1].* 1) Dept of Ophthalmology, Flinders University, Flinders Medical Centre, Adelaide, Australia; 2) SA Pathology, Adelaide, Australia; 3) Genetica Medica, Niguarda Ca'Granda Hospital, Milan, Italy; 4) Institute of Human Genetics, Friedrich-Alexander-Universitét Erlangen-Nürnberg, Erlangen, Germany.

Purpose: Primary Congenital Glaucoma (PCG) is an important cause of childhood glaucoma blindness. This developmental disorder is characterized by buphthalmos, Haab's striae, increased intraocular pressure, and optic neuropathy. The majority of cases are sporadic. *CYP1B1* mutations are identified in 20% of Caucasian cases, and display an autosomal recessive transmission, accounting for the majority of familial cases. Axenfeld-Rieger syndrome (ARS) is another developmental disorder characterized by anomalies of the anterior segment of the eye including the iris and the cornea, and is associated with a 50% risk of developing glaucoma usually of early onset. ARS is caused by gene defects in *FOXC1* and *PITX2*. *FOXC1* defects have been reported in patients with ARS and congenital glaucoma. Here we investigated the role of the *FOXC1* gene in PCG. **Methods:** Patients diagnosed with PCG were recruited through the Australian and New Zealand Registry of Advanced Glaucoma and the Institute of Human Genetics at the University of Friedrich-Alexander in Germany. Subjects with mutations in the *CYP1B1* gene were excluded. Exome sequencing was performed using the Agilent SureSelect and run on Illumina HiSeq 2500. Mutations and copy number variants were validated by Sanger sequencing and MLPA respectively. **Results:** Exome sequencing identified defects in *FOXC1* in 4% (3/68) of patients with PCG: Two frame-shift variants encoding a premature termination codon (p. Leu240ValfsTer65 and p. Asp117ThrfsTer63), and a deletion of the entire *FOXC1* gene from one allele, were detected in separate individuals. None of the three cases identified with *FOXC1* defects had apparent classical features of ARS. **Conclusions:** PCG has long been considered a sporadic condition, or an autosomal recessive condition due to mutations in *CYP1B1*. Here we report mutations and deletions in the *FOXC1* gene in 4% of cases diagnosed with PCG and without apparent classical features of ARS. These results have important implications for the genetic counseling of these families, including the recurrence risk for siblings and the risk of developing glaucoma in relatives.

3009T

Whole exome sequencing of family trios identifies novel genes in children with Rett syndrome (RTT) phenotype. *S. Rangasamy[1,2], K. M. Ramsey[1,2], S. Szelinger[1,2,3], N. Belnap[1,2], A. Claasen[1,4], A. L. Siniard[1,2], AA. Kurdoglu[1,2], A. L Courtwright[1,2], R. F. Richholt[1,2], M. De Both[1,2], JJ. Corneveaux[1,2], I. Schrauwen[1,2], M. J. Huentelman[1,2], D. W Craig[1,2], V. Narayanan[1,2].* 1) Neurogenomics Division, Translational Genomics Research Institute(TGen), Phoenix, AZ; 2) Center for Rare Childhood Disorders, Translational Genomics Research Institute(TGen), Phoenix AZ; 3) Molecular and Cellular Biology Interdisciplinary Graduate Program, School of Life Sciences, Arizona State University, Tempe AZ; 4) Department of Medical Genetics, University of Antwerp, Antwerp, Belgium.

Rett syndrome (RTT) is an X-linked dominant neurological disorder caused by mutation of the MeCP2 gene. It is characterized by developmental regression, stereotyped hand movements and autistic features. Mutation in MeCp2 gene has been identified in 95% typical RTT cases. Key clinical features have been defined that allow for the recognition of a classical or typical RTT phenotype. An atypical RTT phenotype has also been defined based on clinical criteria. Mutations of genes encoding cyclin-dependent kinase-like 5 (CDKL5) and forkhead box G1 (FOXG1) are found in patients with variant forms of RTT. Although most children with MeCP2 mutations are female, there are male infants with a severe neonatal onset encephalopathy caused by mutations in MeCp2. An X-linked mental retardation phenotype is also reported with specific mutations in MeCP2; thus MeCp2 gene mutation is not synonymous with RTT. At our center, focused on discovering the genetic basis of neurodevelopmental disorders, we have enrolled over 290 families into our research study, and have completed whole-genome or whole-exome sequencing using our in-house bioinformatics pipeline and annotationin 217 families. In this group, we have identified a cohort of patients (n=11) who have phenotype similar to Rett syndrome. All the patients had the features that brought up the diagnosis of Rett syndrome, but neither had mutation in *MECP2*gene (MECP2 negative) nor deletion/duplication. In 7 of these cases, we identified a novel, heterozygous, *de novo* mutation. Four of the genes identified have previously been reported to result in epileptic encephalopathy (GRIN1, GNAO1, KCNQ2, and KCNB1) and one identified has been linked to a syndrome with intellectual disability, speech disorder, and epilepsy (GRIN2A). In the remaining two patients, one had a pathogenic mutation in the GABGR2 gene (associated with generalized epilepsy with febrile seizures), and the other patient had a variant in the ATP1A2 gene (linked to alternating hemiplegia of childhood). Our results suggest thata Rett-like phenotype is caused by mutations in a variety of genes associated with synapse development or function, and previously linked to epilepsy and intellectual disability. It is likely that such studies using exome sequencing will expand the list of genes linked to neurodevelopmental disorders with a phenotype that overlaps with typical or atypical RTT (Rett-like syndrome).

3010F

TMEM38B/TRIC-B deficiency causes an Osteogenesis imperfecta-like bone dysplasia by dysregulation of ER calcium homeostasis and collagen biosynthesis. *W. A. Cabral[1], M. Ishikawa[2], E. Makareeva[3], B. M. Owens[1], A. M. Barnes[1], M. Weis[4], F. Lacbawan[5], D. R. Eyre[4], S. Leikin[3], Y. Yamada[2], J. C. Marini[1].* 1) Bone & Extracellular Matrix Branch, NICHD/NIH, Bethesda, MD; 2) Molecular Biology Section, NIDCR/NIH, Bethesda, MD; 3) Section on Physical Biochemistry, NICHD/NIH, Bethesda, MD; 4) Department of Orthopaedics and Sports Medicine, University of Washington, Seattle, WA; 5) Department of Medical Genetics, Children's National Medical Center, Washington, D. C.

Recessive osteogenesis imperfecta (OI) is caused by defects in proteins involved in post-translational interactions with type I collagen. Recently, a novel form of moderately severe OI was identified in Bedouins and a single Albanian proband due to null mutations of *TMEM38B*. *TMEM38B* encodes the integral ER membrane monovalent cation channel, TRIC-B, proposed to regulate Ca2+ release from intracellular stores. However, the molecular mechanisms by which *TMEM38B* mutations cause OI are unknown. We identified 2 patients with recessive defects in *TMEM38B* causing moderately severe bone dysplasia. Patient 1 is a 20-month old Saudi girl who is homozygous for the Bedouin founder mutation, a deletion of exon 4. Patient 2 is a 27-year old American male with compound heterozygosity for a maternal 1nt duplication in exon 1 and a paternal deletion encompassing exons 1 and 2, identical to that previously reported in the Albanian child. *TMEM38B* transcripts in proband fibroblasts and osteoblasts were significantly reduced to 25-70% of control levels. Absence of TRIC-B protein, confirmed by Western analysis, was associated with decreased intracellular and ER luminal Ca2+ concentrations that are more rapidly depleted upon ATP-induced Ca2+ flux. Lower Ca2+ levels are not due to altered expression or stability of SERCA2b or IP3R, the channels involved in ER Ca2+ loading and release, respectively. As expected, depletion of intracellular Ca2+ is associated with ATF4-mediated ER stress and increased BiP. Absence of TRIC-B dysregulated synthesis of type I collagen, which has increased electrophoretic mobility resulting from a 30% reduction in helical lysine hydroxylation. Collagen telopeptide hydroxylation is normal, despite decreased levels of Ca2+-dependent FKBP65. Although normal PDI protein levels are maintained, there is a 20-minute delay in procollagen chain assembly in proband cells. The resulting misfolded collagen is substantially retained in the cell, with a higher proportion of intracellular collagen observed in proband versus control cells, consistent with a 50-70% reduction in total secreted collagen. A lower-stability secreted collagen species detected by DSC is not deposited into extracellular matrix, which contained only normal stability collagen. These data support a role for TRIC-B in intracellular Ca2+ mobilization. Absence of TRIC-B causes bone dysplasia by dysregulation of multiple Ca2+-regulated collagen-specific chaperones and modifying enzymes in the ER.

3011W

Whole-Exome Sequencing Reveals Novel Gene Implicated in Multiple Epiphyseal Dysplasia and Primary Osteoarthritis. *M. Czarny-Ratajczak[1], V. Dasa[2], J. Eastwood[1], C. Blackstock[1], K. Kozlowski[3], S. M. Jazwinski[1].* 1) Department of Medicine, Center for Aging, Tulane University, School of Medicine, New Orleans, LA; 2) Department of Orthopaedic Surgery, Louisiana State University Health Sciences Center, New Orleans, LA; 3) Department of Medical Imaging, The Children's Hospital at Westmead, Sydney, Australia.

Multiple epiphyseal dysplasia (MED) is a relatively common chondrodysplasia characterized by delayed and irregular ossification of epiphyses and by early-onset osteoarthritis (OA). The first symptoms of MED include waddling gait, restriction of joint mobility, and pain and stiffness in the weight-bearing joints. To date, mutations in six genes have been found as the cause of 60% of MED cases; however, the genetic defects that are responsible for the remaining 40% of MED are unknown, suggesting that mutations in other genes are also involved in its pathogenesis. Furthermore, very little is known about the genetic predisposition to primary osteoarthritis. Currently, the overlapping multiple epiphyseal dysplasia/osteoarthritis phenotype is reinvestigated since it may provide new insights towards understanding the genetic basis for common forms of primary osteoarthritis. To identify novel candidate genes for MED we performed whole-exome sequencing utilizing the Ion Proton next-generation sequencing system (Ion Torrent/Life Technologies) and Ion TargetSeq™ Exome enrichment approach. Whole-exome sequencing was completed among direct relatives within Caucasian families with the autosomal dominant form of MED and early-onset OA. All known MED candidate genes were excluded in these families. Read mapping, determination of exome coverage metrics, and variant calling was completed with the Torrent Suite™ Software. Trio analysis of exome-sequencing data was performed with Ion Reporter ver. 4. 4 software (Ion Torrent/Life Technologies), which utilizes algorithms including PolyPhen, SIFT, PhyloP and Grantham scores to predict effects of the sequence changes. Using whole-exome sequencing we identified a novel candidate gene that plays a causative role in MED and potentially in primary OA. Furthermore, we analyzed co-segregation of identified changes in MED families as well a separate patient population undergoing elective surgery for primary knee osteoarthritis. Overall, our studies indicate a novel candidate gene for MED and provide new insight towards understanding the genetic basis for primary osteoarthritis. This research is supported by: NIH (NIGMS) grant no. U54-GM104940:551192K2 and Bone and Joint Initiative USA Scholarship for Young Investigators to M. C-R. Exomes were analyzed in the Genomics and Biostatistics Core funded by NIH (NIGSM), Mentoring Research Excellence in Aging and Regenerative Medicine grant no. P20GM103629 to S. M. J.

3012T

Bone Robusticity in Two Distinct Skeletal Dysplasias: an Evaluation of the Second Metacarpal, a Surrogate for Bone Strength. *J. MArino[1], K. Jepsen[2], E. Carter[1], C. Raggio[1].* 1) Ctr for Skeletal Dysplasias, Hospital For Special Surgery, New York, NY; 2) Orthopaedic Research Laboratories, University of Michigan, Ann Arbor, MI.

Radiographs of the second metacarpal are used to assess bone strength development in paediatric populations. Children with achondroplasia and osteogenesis imperfecta (OI) have known differences in bone strength. Details of how bone strength develops and compares within these populations to unaffected children are lacking. A data set for patients with achondroplasia and OI was established. A retrospective IRB-approved review of bone-age films(n=67; 1-11 films/patient) from patients(5mos to 16yrs+3mos old) with achondroplasia(6 males; 10 females) or OI(9 males; 11 females) was conducted. A sample of modern controls (diagnosis: leg-length discrepancy) matched historical measurements from the Bolton-Brush collection(6mos-16yrs). Metacarpal length(Le) was measured from the proximal end to the most distal ossified end along the midshaft axis. Outer and inner diameters were measured at 50% and 60% of the length, averaged, and used to calculate total cross-sectional area(Tt. Ar) and cortical area(Ct. Ar) using a circular approximation. To adjust for differences in body size, we compared robustness(Tt. Ar/Le) and relative cortical area(RCA=Ct. Ar/Tt. Ar) among groups. Achondroplasia patients tend to have robustness values above and RCA values below the most robust Bolton-Brush tertile. This robust phenotype was consistent with the reduced longitudinal growth seen in achondroplasia patients because of the constitutive activation of FGFR3. RCA values were decreased compared to but followed the distribution of the control population. OI patients did not follow the established Bolton-Brush pattern of robustness and RCA. OI patients fell into the most slender tertile for robustness and RCA values of OI patients were similar to that of controls. Patients did not separate based on OI subtype. No sexual dimorphism was noted in this study. The lack of sexual dimorphism in the dysplasia populations is in contrast to that reported in the unaffected population. We suggest that the underlying dysplasia overrides the sex-specific effects on bone strength development. The contribution of the specific mutation is unknown and needs to be further studied.

3013F

Variants in *ACTC1* cause distal arthrogryposis: a new skeletal muscle phenotype for a cardiac muscle gene. *C. T. Marvin[1], J. X. Chong[1], K. J. Buckingham[1], K. M. Shively[1], T. M. Harrell[1], A. G. Shankar[1], M. J. McMilliin[1], D. A. Nickerson[2], J. Shendure[2], Y. Alanay[3], A. S. Aylsworth[4], M. I. Roche[4], C. Turnbull[5], M. J. Bamshad[1, 2, 6], University of Washington Center for Mendelian Genomics.* 1) Dept. of Pediatrics, University of Washington, Seattle, WA, USA; 2) Dept. of Genome Sciences, University of Washington, Seattle, WA, USA; 3) Pediatric Genetics, Department of Pediatrics, Acibadem University School of Medicine, Istanbul, Turkey; 4) Departments of Pediatrics and Genetics, Division of Genetics and Metabolism, University of North Carolina, Chapel Hill, NC, USA; 5) Division of Genetics and Epidemiology, Institute of Cancer Research, Sutton, UK; 6) Seattle Children's Hospital, Seattle, WA, USA.

Contraction of the human sarcomere is the result of interactions between myosin cross-bridges and actin filaments. Mutations in genes such as *MYH7*, *TPM1*, and *TNNI3*, that encode parts of the cardiac sarcomere cause muscle diseases that affect the heart, such as dilated cardiomyopathy and hypertrophic cardiomyopathy. In contrast, mutations in homologous genes *MYH2*, *TPM2*, and *TNNI2*, that encode parts of the skeletal muscle sarcomere, cause muscle diseases affecting skeletal muscle, such as the distal arthrogryposis (DA) syndromes. To date, few genes that encode sarcomeric proteins have been found that contain mutations that affect both skeletal and cardiac muscle. Moreover, none of the known genes for DA have been found to contain mutations that also cause a cardiac phenotype. We report two families with dominantly inherited missense variants and one family with a *de novo* missense variant in the gene actin, alpha, cardiac muscle 1 (*ACTC1*). None of these variants were found in either EVS or ExAC. Sanger validation confirmed these variants track with affected individuals. Analysis metrics GERP and CADD indicate these loci are highly conserved and that the variants are predicted to be damaging. In family one, comprised of an affected father and daughter, a variant in exon 4 of *ACTC1* (c. 595G>A, p. Gly199Ser, GERP 5. 08, CADD 22. 3) was identified. This family presented with neck webbing, vertebral fusion, clubfeet, hypoplastic flexion creases, and a ventricular septal defect in the father. The second family, an affected mother and daughter, had a variant in exon 7 of *ACTC1* (c. 1120C>A, p. Arg374Ser, GERP 4. 92, CADD 22). Other features included a history of severe scoliosis, clubfeet, contractures, short stature, mild ptosis, and left eye coloboma in addition to a midline depression in the anterior chest in the daughter. The third case was simplex and presented with clubfeet and camptodactyly. The proband was found to have a variant in exon 3 of *ACTC1* (c. 203C>A, p. Thr68Asn, GERP 5. 49, CADD 18. 46). *ACTC1* is a highly conserved actin gene that binds to myosin and is expressed in both cardiac and skeletal muscle. Mutations in *ACTC1* have previously been found to cause atrial septal defect, dilated cardiomyopathy, hypertrophic cardiomyopathy, and left ventricular noncompaction. Our discovery delineates a new DA phenotype caused by mutations in *ACTC1* and suggests that some functions of actin, alpha, cardiac muscle 1 are shared in cardiac and skeletal muscle.

3014W

Analysis of a broad spectrum of mutations in the *EXT1* and *EXT2* genes in Brazilian patients with multiple osteochondromas. *S. C. L. Santos[1], I. M. P. O. Rizzo[2], R. I. Takata[1], C. E. Speck-Martins[2], C. Sollaci[3].* 1) Molecular Biology Lab, SARAH Network of Hospitals, Brasília, DF, Brazil; 2) Clinical Gentics Dept, SARAH Network of Hospitals, Brasília, DF, Brazil; 3) Orthopedics Dept, SARAH Network of Hospitals, Brasília, DF, Brazil.

Multiple Osteochondroma (MO; OMIM# 133700) is an autosomal dominant disease characterized by the formation of multiple cartilaginous tumors (osteochondromas). The great variability in size and number of osteochondromas reflects the clinical heterogeneity and variable severity, which is characterized by pain, abnormal skeletal growth, deformities and nerve compression. The most serious complication of MO consists in the malignant transformation to a chondrosarcoma (0,5 – 2% of cases). MO is genetically heterogeneous and 70-95% of patients present mutations in one of two tumor suppressor genes *EXT1* (MIM# 608177) or *EXT2* (MIM# 608210). To date, 713 different *EXT1* and 386 *EXT2* mutations have been described. This study is the first Brazilian research in MO, with a broad spectrum of mutations detected in 73 unrelated patients and 25 families affected. We first analyzed 117 probands for *EXT1* mutations by DHPLC analysis and subsequent direct sequencing of all samples with abnormal profile. Negative cases were then screened for *EXT2* mutations using the same approach. Negative cases for both genes were directly sequenced. If no mutation was detected, negative cases were analyzed by MLPA. Mutant allele was found in 79,3% of unrelated MO patients in which *EXT1* mutations were detected in 32,6% of cases and *EXT2* mutations were detected in 46,7%. Nineteen (20,7%) patients had no mutations detected. Mutations were detected in all families wherein 48% presented mutations in *EXT1* and 52% presented mutations in *EXT2* gene. We identified 51 different pathogenic mutations, 33 (14 frame-shift, 10 nonsense, 5 missense, 2 splice-site mutation and 2 large deletions) in *EXT1* and 18 (6 frame-shift, 6 splice-site mutation, 3 nonsense, 2 missense and 1 large deletion) in *EXT2*. Nineteen of the *EXT1* mutations were novel as were 9 of the *EXT2* mutant alleles. Three unrelated patients presented malignant transformation. Two of them had no mutation identified and the other one presented the most common mutation in both unrelated patients and families in the *EXT2* gene suggesting the presence of another site of alteration in the genome causing more severe types of MO. We found a different pattern of mutations in the genes involved in MO from the patterns in literature. Most other studies detected a greater number of mutations in *EXT1* than *EXT2* gene showing that Brazilian population has a unique genetic spectrum that differs from the European populations and other Latin American countries.

3015T

Novel FKBP10 mutation induces osteogenesis imperfecta type XI .
S. Seyedhassani[1,2], M. Yavari[1], F. Harazi[1], F. Hashemi-Gorji[2], V. Yasaee[2]. 1) Dr. Seyedhassani Medical Genetic Center; 2) Genomic Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran.

Introduction: Osteogenesis imperfecta (OI) comprises a group of connective tissue disorders characterized by bone fragility and low bone mass. The disorder is clinically and genetically heterogeneous. OI type XI is an autosomal recessive form of OI . **Clinical Presentation:** The proband was a seven years old boy that born in a third degree consanguineous marriage by normal vaginal delivery. He had knee deformity at the birth time. Since day one he was diagnosed with blue sclera and experienced several bone fractures as OI, and therefore had five surgeries in knee and elbow joints. He was not enabling to standing and walking until now. Radiographic findings at birth time revealed enough bone evidence for OI. There were another two similar cases in familial pedigree that were occurred in fifth degree relatives. **Material and Methods:** Peripheral blood samples were obtained following informed consent adhering to the tenets of the Declaration of Helsinki. DNA was extracted using a Flexigene blood DNA kit (DNA fast, QIAGEN) according to the manufactures protocol and stored in TE buffer. **Results:** All of the exons and exon-intron boundaries of CRTAP and FKBP10 genes were amplified under standard PCR conditions. The other known genes were excluded based on clinical findings and inherritance pattern. PCR products were subjected to Sanger sequencing and output files were aligned to the human genome reference and analyzed by Finch TV 1. 4 software. In addition, Human Gene Mutation Database (HGMD), Exome Variant Server (EVS) and available literature were observed for minor allele frequencies and earlier reports of this novel mutation. Screening of all exons and adjacent introns revealed a homozygous deletion as ATCCCC(^324)GGGaTGGA in exon 6 of the FKBP10 gene resulting in protein alteration. Extended analysis verified that the novel mutation segregated within the family. **Conclusion:** We have identified a novel FKBP10 mutation. This finding can be useful in genetic counseling and detection of priority in molecular analysis of OI in Iranian population. Prenatal diagnosis is suggested to speculate the fate of the fetus in the next pregnancy.

3016F

Mutations spectrum of *COL1A1/COL1A2* in Chinese with osteogenesis imperfecta. *X. Zhao, J. Xiao, Y. Wu, C. Lu, H. Wang, X. Zhang.* McKusick-Zhang Center, Chinese Academy of Medical Sciences - Peking Union Medical College, Beijing.

Osteogenesis imperfecta (OI) is a group of clinically and genetically heterogeneous disorders with an incidence of 1/10000 in live births. The main clinicalmanifestations in OI include multiple fractures, blue sclera, short stature, dentinogenesis imperfecta, hearing loss and so on. Uptonow, fifteen different forms (OI type 1~OI type 15) have been recognized and classified, and at least thirteen genes(*COL1A1*, *COL1A2*, *IFITM5*, *SERPINF1*, *CRTAP*, *LEPRE1*, *PPIB*, *SERPINH1*, *FKBP10*, *SP7*, *BMP1*, *TMEM38B* and *WNT1*)have been reported to relate to OI. Lots of studies have shown that more than 90% affected individuals have mutations in either *COL1A1* or *COL1A2* which encode the chains of type I procollagen. In this study, we collected 202 OI cases (65 familial cases and 137 sporadic cases) and their family members. All the affected individuals showed notable OI phenotypes. We focused on mutation identification in *COL1A1/2* in all the 202 probands. All the coding region and exon/intron boundaries in *COL1A1/2* were amplified using 43 pairs of primers. Next generation sequencing was carried out to detect the candidate mutations in the probands, and all the mutations were verified by PCR-Sanger sequencing. The DNA from the unaffected parents of the sporadic cases was also screened to ascertain the absence of the de novo mutations. Novel mutations were validated in 50 normal unrelated samples by PCR followed by high resolution melting (HRM). In total, we found 127 mutations in 155 probands, including 79 mutations in *COL1A1* and 48 mutations in *COL1A2*. Among them, 63 mutations were novel and haven't been reported in human gene mutation database (HGMD), including 36 mutations in *COL1A* 1 and 27 mutations in *COL1A2*. In summary, we have set up a comprehensive platform including PCR, Sanger/ next generation sequencing and HRM to perform genetic diagnosis for the patients with OI. We extended the mutation spectrum of *COL1A1 /2* in Chinese OI cases.

3017W

Mutations in *MYLPF* cause a novel segmental amyoplasia manifest as distal arthrogryposis. *K. M. Shively[1], J. X. Chong[1], C. T. Marvin[1], A. E. Beck[1,2], T. M. Harrell[1], M. J. McMillin[1], K. J. Buckingham[1], A. S. Aylsworth[3], J. Shendure[4], D. A. Nickerson[4], M. J. Bamshad[1,2,4], University of Washington Center for Mendelian Genomics.* 1) Department of Pediatrics, University of Washington, Seattle, WA, USA; 2) Seattle Children's Hospital, Seattle, WA, USA; 3) Departments of Pediatrics and Genetics, University of North Carolina, Chapel Hill, NC, USA; 4) Department of Genome Sciences, University of Washington, Seattle, WA, USA.

The distal arthrogryposes (DA) are a phenotypically and genetically heterogeneous group of disorders characterized by congenital contractures of the limbs and subcategorized by the severity of the contractures and presence of additional features. Dominantly inherited or *de novo* mutations in *TPM2*, *TNNI2*, *TNNT3*, and *MYH3*, which encode proteins of the contractile apparatus of skeletal muscle, are the most common causes of DA but explain only ~60% of families. During analysis of exome sequencing data from a large cohort of individuals with various forms of DA, we identified two kindreds (Families A and B) with a novel form of DA. Family A was of European ancestry and consisted of two affected siblings whose unaffected parents were second cousins once removed. Each affected individual had severe contractures of the hands, fingers, wrists, elbows, hips, knees, ankles, and neck; pterygia of the elbows and knees; small mouths; and were below the tenth centile for weight and height. Family B consisted of a single, adopted proband of East Indian ancestry whose parents were likely first cousins based on analysis of high-density genotyping data (F=0. 0657). He had a small mouth, micrognanthia, scoliosis, contractures of the hands and wrists, and severe clubfoot that was recalcitrant to treatment, ultimately leading to amputation of the foot. Pathological exam of the foot reveal an absence of skeletal muscle that was confirmed histologically. All affected individuals shared the same homozygous mutation, c. 470G>T, p. (Cys157Phe), in exon 7 of myosin light chain, phosphorylatable, fast skeletal muscle (*MYLPF*). The mutation was observed once in 59,941 control exomes in the ExAC database, not seen in EVS or 1000 Genomes, and is predicted to be deleterious (Polyphen 0. 779, CADD 35. 0). Moreover, *Mylpf* null mice are 30% smaller than their wild type or heterozygous counterparts; they lack skeletal muscle in their whole body and die immediately after birth due to respiratory failure caused by lack of skeletal muscle in the diaphragm. Accordingly, the mouse model appears to recapitulate the human phenotype. These findings suggest that congenital contractures in these families is caused by a novel mechanism that results in hypoplasia or aplasia of skeletal muscle in the distal parts of the limbs.

3018T

Skeletal Ciliopathies – a molecular study of 17 patients with Short-Rib Thoracic Dysplasia (SRTD). *A. Hammarsjö[1, 2], M. Pettersson[1], K. Lagerstedt-Robinson[1, 2], H. Malmgren[1, 2], F. Taylan[1], A. Beleza-Meireles[3], D. Chitayat[4], K. Girisha[5], K. Shimizu[6], E. Horemuzova[7, 8], A. Lin[9], I. Sahai[9], A. Traum[10], H. Jüppner[11], M. Nordenskjöld[1, 2], A. Nordgren[1, 2], G. Nishimura[12], A. Lindstrand[1, 2], G. Grigelioniene[1, 2].* 1) Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden; 2) Department of Clinical Genetics, Karolinska University Hospital, Stockholm, Sweden; 3) Center of Human Genetics, UC Louvain, Cliniques Universitaires Saint-Luc, Brussels, Belgium; 4) The Prenatal Diagnosis and Medical Genetics Program, Mount Sinai Hospital, University of Toronto, Ontario, Canada; 5) Department of Medical Genetics, Kasturba Medical College, Manipal, India; 6) Division of Medical Genetics, Saitama Children's Medical Center, Saitama, Japan; 7) Paediatric Endocrinology Unit, Astrid Lindgrens's children hospital, Karolinska University Hospital, Stockholm, Sweden; 8) Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden; 9) Medical Genetics, MassGeneral Hospital for Children, Boston, MA, USA; 10) Nephrology, MassGeneral Hospital for Children, Boston, MA, USA; 11) Endocrine Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA; 12) Department of Pediatric Imaging, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan.

Short-rib thoracic dysplasia (SRTD [MIM 208500]) is a group of recessive skeletal ciliopathies characterized by extensive genetic and phenotypic overlap, with some differences in malformations and skeletal appearance. SRTD encompasses Ellis-van Creveld syndrome (EVC [MIM 225500]) and the diagnoses previously designated as Jeune syndrome, asphyxiating thoracic dystrophy (ATD), short rib-polydactyly syndrome (SRPS) and Mainzer-Saldino syndrome (MZSDS [MIM 266920]). Mutations in at least 13 genes have been reported to date (*IFT80, EVC, LBN, DYNC2H1, TTC21B, WDR19, NEK1, WDR35, WDR60, IFT140, IFT172, WDR34* and *CEP120*) all encoding proteins important for primary cilia function. There is also a phenotypic and genetic overlap with Sensenbrenner syndrome or cranioectodermal dysplasia (CED [MIM 218330]), which is caused by mutations in *IFT122*, *WDR35*, *WDR19* and *IFT43*. Nearly all genes that have been associated with SRTD are located in genes encoding proteins that are part of the intraflagellar transport (IFT) machinery or have a role in ciliary assembly but the exact mechanisms are not yet fully understood. The aim of our study was to extend the genetic and phenotypic characterization of skeletal ciliopathies. 17 unrelated individuals with a clinical and radiological diagnosis of SRTD (N=12) and CED (N=5) were recruited from six countries (Sweden, Canada, USA, Japan, India and Belgium). The DNA samples were analyzed by whole exome sequencing (WES) and were filtered against a defined clinically relevant gene panel of 347 genes including all known skeletal dysplasia and ciliopathy loci. Identified variants were validated in the probands and their parents using Sanger sequencing. The WES analyses from the first 13 families have identified the causal mutation in 77% of the cases. Specifically we identified 16 novel variants and six previously reported mutations in five genes (i. e. *WDR19, DYNC2H1, NEK1, C2CD3 and IFT140*). Our findings show that WES together with a gene panel is a powerful strategy to establish molecular diagnosis of SRTD, a disease with a heterogenous genetic entity. This study expands current knowledge about genetic and phenotype variability in skeletal ciliopathies leading to improved diagnosis and genetic counseling for the affected families.

3019F

Molecular analysis of 24 Ellis van Creveld syndrome (EVC) individuals using targeted Next Generation Sequencing (NGS) and array CGH technology. *C. Huber[1], B. BAUJAT[1], M. ZARHATE[2], C. BOLE-FREYSOT[2], C. MASSON[3], P. NITSCHKE[3], S. NUSBAUM[4], C. FRANCANNET[5], A. MEGARBANE[6], D. MARTIN-COIGNARD[7], Y. ALEMBIK[8], M-P. ALEX-CORDIER[9], K. DEVRIENDT[10], A. DIEUX[11], L. FAIVRE[12], M. GONZALES[13], N. LAURENT[12], S. MANOUVRIER[11], M. MATHIEU[14], A. PARIS[15], F. PELLUARD[16], L. SPACINI[17], S. SAUNIER[18], V. MALAN[4], V. CORMIER-DAIRE[1].* 1) Service de Génétique, INSERM UMR 1163, Université Paris Descartes- Sorbonne Paris Cité, Institut Imagine, Hôpital Necker Enfants Malades (AP-HP), Paris, 75015, France; 2) Plateforme de Génomique, Institut IMAGINE, Paris, 75015, France; 3) Plateforme de Bioinformatique, Université Paris Descartes, Paris, 75015, France; 4) Service de CytoGénétique, Hôpital Necker-Enfants Malades, Paris, 75015, France; 5) Service de Génétique Médical - CHU Estaing, Clermont - Ferrand, 63001, France; 6) Medical Genetics Unit, Saint Joseph University, Beirut, Lebanon; 7) Service de Génétique, Centre Hospitalier du Mans, Le Mans, 72037, France; 8) Service de Génétique Médicale, Hôpital de Hautepierre, Strasbourg, 67098, France; 9) Service de Génétique, Hôpital femme Mère Enfant GHE, Bron, 69500, France; 10) Kliniekhoofd Centrum Menselijke Erfelijkheid, UZ Leuven, Campus Gasthuisberg, Leuven, B-3000, Belgie; 11) Service de Génétique Clinique, CHRU, Hôpital Jeanne de Flandre, Lille, 59037, France; 12) Génétique et Anomalies du Développement, EA4271, Université de Bourgogne, Dijon, 21034, France; 13) Service de Génétique et d'Embryologie Médicales, Hôpital Armand Trousseau, Paris, 75571 cedex 12, France; 14) Service de Génétique Clinique CHRU d'Amiens- Hôpital Nord, Amiens, 80054, France; 15) Service de Gynécologie Obstétrique, Maison de Santé Protestante Bordeaux-Bagatelle, Talence, 33400, France; 16) Pôle de Biologie, Service de Pathologie, CHU de Bordeaux, Bordeaux, 33076, France; 17) Unita Operativa Complessa di Ostetricia e Ginecologia, Istituti Clinici di Perfezionamento, Milano, Italia; 18) INSERM, UMR 1163, Laboratory of Inherited Kidney Diseases, Université Paris Descartes- Sorbonne Paris Cité, Institut Imagine, Hôpital Necker Enfants Malades (AP-HP), Paris, 75015, France.

The SRP group includes 4 lethal disorders and 2 disorders compatible with life, asphyxiating thoracic dysplasia (ATD) and Ellis-van Creveld (EVC) syndrome. Common characteristic features include constricted thoracic cage, short ribs, shortened tubular bones, and trident aspect of the acetabular roof. EVC is distinct by the presence of post-axial polydactyly, congenital cardiac defects, dysplastic nails and teeth and labio-gingival adhesions. EVC is caused by mutations either in *EVC* or in *EVC2*. Here, we report the molecular screening of 24 EVC individuals by targeted NGS of a customized ciliopathy gene panel, called ciliome, associated with array CGH technology for 6 cases. The series includes 14 fetuses and 11 postnatal cases. Among them, 18/24 (11 fetuses and 6 postnatal cases) were found to harbor variants (10 nonsense, 3 splice site, 1 missense, 4 deletions) either in *EVC* (11), *EVC2* (7) or *EVC*/*EVC2* (1). In 8 cases, the identification of i) an heterozygous variant only (4/8), ii) an "homozygous" variant confirmed only at the heterozygous state in the father (3/8), iii) the absence of any variant (1/8), prompted us to perform microsatellite analysis at the *EVC*/*EVC2* loci. We identified 2 heterozygous deletions, encompassing either *EVC* or *EVC*/*EVC2*. Because of the non informativity of the intragenic *EVC* microsatellite in 6 cases, we performed an array CGH, leading to the detection of 4 deletions encompassing *EVC*, always inherited from the mother. Finally, in 2/24 cases only one heterozygous variant was detected and in 7/24 cases, neither variant nor deletion were detected in *EVC*/*EVC2*, supporting the involvement of another disease gene. By combining ciliome analysis and array CGH, we conclude that *EVC* and *EVC2* account for 75 % of EVC cases. Our study emphasizes the high frequency of maternal deletion (half cases) of variable sizes. Finally, exome sequencing is in progress for the 7 remaining cases.

3020W

***SEC24D* causes non-syndromic autosomal recessive osteogenesis imperfecta.** *S. Moosa[1], B. Chung[2], J. Altmueller[1,3], H. Thiele[3], P. Nuernberg[3], G. Nishimura[4], B. Wollnik[1].* 1) Institute of Human Genetics, University Hospital of Cologne, Cologne, Germany; 2) Department of Paediatrics and Adolescent Medicine, Queen Mary Hospital, Hong Kong, China; 3) Cologne Center for Genomics (CCG), University of Cologne, Cologne, Germany; 4) Department of Pediatric Imaging, Tokyo Metropolitan Children's Medical Center, Tokyo, Japan.

Osteogenesis imperfecta (OI) is a heterogeneous group of disorders characterized by increased bone fragility, decreased bone mass and a propensity to fractures. Recently, biallelic mutations in *SEC24D* were shown to cause a novel form of syndromic OI, resembling Cole-Carpenter syndrome. Here, we report on the first patient with non-syndromic OI, with two novel compound heterozygous mutations in *SEC24D*. Using whole-exome sequencing (WES), we identified the c. 113dupC mutation in exon 2, which leads to a frameshift and premature stop codon (p. T39Nfs*16) and the c. G2496T (p. Q832H) mutation, which affects the last base of exon 19. *In vitro* analysis of constructs harbouring this mutation revealed abnormal splicing. Unlike the original *SEC24D* patients described, our patient does not show the dramatic facial phenotype and skull ossification defects associated with mutated *SEC24D* in patients and corresponding animal models. Here, we present her clinical features and the results of the molecular genetic analysis. Thus, our report expands the *SEC24D*-associated phenotypic and mutational spectrum, and confirms *SEC24D* as an additional gene underlying autosomal recessively inherited non-syndromic OI.

3021T

Tissue specific mosaicism for a lethal *COL1A1* mutation causes mild Ehlers-Danlos syndrome. *S. Symoens[1], W. Steyaert[1], L. Demuynck[1], F. Malfait[1], A. De Paepe[1], K. E. M. Diderich[2], P. J. Coucke[1].* 1) Center for Medical Genetics Ghent, Ghent University Hospital, Ghent, Oost-Vlaanderen, Belgium; 2) Department of Clinical Genetics, Erasmus MC, Rotterdam, The Netherlands.

Type I collagen is the predominant protein of multiple connective tissues such as skin and bone. Mutations leading to structural abnormal type I collagen mainly cause the brittle bone disease Osteogenesis imperfecta (OI), resulting in a wide spectrum of clinical severity ranging from few fractures to intra-uterine lethality with multiple fractures and malformations of the (long) bones. Until now, 1417 unique mutations in the type I collagen genes (*COL1A1* and *COL1A2*) have been reported, of which in-frame deletions account for only 3% and all are almost exclusively associated with a severe-to-lethal OI phenotype. We describe a patient who was referred because of clinical symptoms of Ehlers-Danlos syndrome, including fragile skin, recurrent luxations and easy bruising. Biochemical collagen analysis of the patients' dermal fibroblasts surprisingly showed dedoubling of the type I collagen bands, a finding specific for structural defects in type I collagen. Subsequent molecular analysis by Sanger sequencing detected an in-frame deletion in exon 44 of the *COL1A1* gene (c. 3150_3158del), resulting in the deletion of three amino acids (p. (Ala1053_Gly1055del)) in the triple helical collagen domain. This *COL1A1* mutation has hitherto only been identified in four probands with lethal OI. However, the peaks on the electropherogram corresponding to the mutant allele were decreased in intensity, suggesting the mosaic presence of the *COL1A1* mutation. We applied next generation sequencing in order to determine the possibility of mosaicism in skin and blood. While approximately 9% of the reads originating from DNA extracted from fibroblasts harboured the *COL1A1* deletion, the deletion was not detected in DNA extracted from blood, implying that the degree of mosaicism is different between tissues thereby explaining the mild phenotype of the patient. In conclusion, the biochemical and molecular characterization of the patient described in this study illustrates that the combined use of biochemical collagen analysis and next generation sequencing technologies is a powerful tool in guiding molecular analysis and in pinpointing the clinical diagnosis in those patients in whom the diagnosis is unclear.

3022F

Order of intron splicing is an important determinant of variant splice outcome in *COL1A1*. *J. Schleit[1], P. H. Byers[1, 2].* 1) Department of Pathology, University of Washington, Seattle, WA; 2) Department of Medical Genetics, University of Washington, Seattle.

Variants that alter mRNA splicing account for approximately 20% of inherited pathogenic variants. Often, the phenotype of the individual is dependent on the stability of the mRNA produced from the variant allele. Currently the prediction of splice outcome remains a challenge for molecular diagnostic labs. We sought to understand how variants alter mRNA splicing and why seemingly similar variants can result in different splice outcomes and disease phenotypes. We first analyzed the rate and order of splicing of introns 43-48 of the type I procollagen gene, *COL1A1*, in three wild-type samples using qPCR and capillary electrophoresis. All three samples displayed similar intron splice times and order of intron removal. Introns 45 and 47 were removed early in the order of splicing and introns 44 and 48 were removed last. In several instances we observed multiple splice order combinations consistent with multiple pathways in some regions. The time required to splice half the intron containing transcripts in this region varied from 6 to 45 minutes, the maximum time examined. To examine the effect of inherited variants on mRNA splicing we analyzed splice order in fibroblasts from four individuals with pathogenic splice donor variants in consecutive introns of *COL1A1*. Two variants, IVS44+2T>C and IVS46+1G>T, resulted in use of cryptic splice donor sites, an unstable mRNA and a mild OI phenotype. The remaining two variants, IVS45+2T>C and IVS47+1G>A, resulted in a stable exon skip transcript that resulted in a lethal OI type II phenotype. Both exon skip/OI type II variants were located in introns that were removed earlier in the local order. Variants in the rapidly removed introns resulted in delayed splicing of that intron and were resolved by exon skipping. In the sample containing the IVS47+1G>A the introns adjacent to the variant also showed delayed removal suggesting a "splice paralysis" had occurred on these transcripts. In contrast, variants in introns removed late in the order allowed splicing of the surrounding introns and led to use of cryptic sites, likely due to the absence of surrounding introns to induce skipping. Our findings indicate the order of intron removal is similar among samples from unrelated individuals and is an important determinant of splice outcome. Future work to identify the factors which determine the order of splicing should increase our ability to predict the outcome of variants which alter mRNA splicing.

3023W

Keratoconus (KC) is an eye disorder in which the cornea is swollen, thinned and deformed. . *f. Azadegan Dehkordi[1], a. Rashki[2], n. Bagheri[3], m. Hashemzadeh Chaleshtori[4], m. Hashemzadeh Chaleshtori[1].* 1) Cellular and Molecular Research Center, Shahrekord University of Medical Science, Shahrekord , Iran; 2) Faculty of Veterinary Medicine, Department of Physiopathology, University of Zabol, Zabol,; 3) Department of Microbiology and Immunology, Tehran University of Medical Sciences, Tehran; 4) Isfahan University of Medical Sciences.

Abstract

Objective: Keratoconus (KC) is an eye disorder in which the cornea is swollen, thinned and deformed. Despite extensive studies, the pathophysiological processes and genetic etiol- ogy of KC are unknown. The disease incidence is approxi- mately 1 in 2,000, and it is the most common cause of cor- neal transplantation in the USA. Many genes are involved in the disease, but evidence suggests a major role for *VSX1* in the etiology of KC. This study aimed to determine the fre- quency of mutations in exons 2, 3 and 4 of the *VSX1* gene in

Chaharmahal va Bakhtiari province in the southwest of Iran. *Study Design:* In this experimental study, mutations in 3 ex- ons, namely exons 2, 3 and 4, of *VSX1* were investigated in 50 patients with KC and 50 healthy control subjects. DNA was extracted using a standard phenol-chloroform method. PCR- single-strand conformational polymorphism/heteroduplex analysis was performed, followed by DNA sequencing to confirm the identified motility shifts. *Results:* H244R muta- tions were found in 1 patient and also in 1 healthy control subject. Furthermore, 12 polymorphisms were identified in patients with KC and 7 in healthy control subjects [rs6138482 and c. 546A>G (rs12480307)]. *Conclusion:* Our investigation showed that KC-related *VSX1* mutations were found in a very small proportion of the studied patients from Iran. Further investigations on other genes are needed to clarify their roles in KC pathogenesis.

3024T

Mutation screening and in-depth bioinformatic analyses of the cathepsin C gene in an Indian family with Papillon-Lefevre syndrome. *M. Raveendrababu[1], M. V. Rao[1], T. Y. Mehta[2], U. Ratnamala[3], U. Radhakrishna[4].* 1) Human Cytogenetics, Department of Zoology, Gujarat University, Ahmedabad, Gujarat, India; 2) Samarpan Centre for skin, Sexually transmitted diseases and Research Center, Modasa, India,; 3) Department of Pharmacology, Creighton University, Omaha, NE; 4) Green Cross Pathology & Molecular Laboratory, Ahmedabad, India,.

Papillon-Lefèvre syndrome (PLS), also known as palmoplantar keratoderma with periodontitis (PALS, MIM 245000) is an autosomal recessive disorder characterized by palmoplantar keratoderma and early onset severe periodontitis affecting both deciduous and permanent dentition. Many PALS families and sporadic cases with variable clinical presentation have been reported. Loss-of-function mutations in the gene encoding cathepsin C (CTSC) have been identified in many families, sporadic PALS individuals from various ethnic groups. We describe Papillon-Lefevre syndrome and an autosomal dominant form of palmoplantar keratoderma in the same Indian family and present the mutation analysis of the cathepsin C gene. Sequence analysis of known exons and splice junctions revealed a homozygous nucleotide substitution G to A at nucleotide position 901, resulting in a change from glycine (GGC) to serine (AGC) at amino acid position 301 (G301S) was observed in homozygous condition in an affected individual, and excluded many other affected members with severe palmoplantar keratoderma from the same family. This mutation was found to be at a highly conserved residue in of *CTSC* gene. Structure prediction and energetic analysis of wild-type *CTSC*, comparison with mutant (G301S) revealed that this change in amino acid does not imply any secondary structural change. However, prediction of functional effect(s) of this mutation is possibly damaging the protein structure and/or function. Moreover, based on the energy calculation and the modeled protein structure of the mutant is expected to be energetically unstable.

3025F

Recurrent *de novo* mutations affecting residue p. Arg138 of pyrroline-5-carboxylate synthase cause a novel progeroid form of autosomal dominant cutis laxa. *B. L. Callewaert[1], B. Fischer-Zirnsak[2,3], N. Escande-Beillard[4], J. Ganesh[5], Y. Xuan Tan[4], M. Al Bughaili[2], I. Sahai[6], S. Chadwick[5], A. Loh[8], P. Bahena[7], G. D. Wright[4], J. Liu[4], E. Rahikkala[9], E. Pivnick[10], U. Krüger[2], T. Zemojtel[2,11], C. van ravenswaaij[12], R. Mostafavi[10], I. Stolte-Dijkstra[12], S. Symoens[1], L. Al-Gazali[13], D. Meierhofer[14], P. N. Robinson[2,3,15], S. Mundlos[2,3,15], P. Beyers[16], A. Masri[17], S. P. Robertson[18], U. Schwarze[19], B. Reversade[20], U. Kornak[2,3,15].* 1) Center for medical genetics, Ghent University, Ghent, Select a Country; 2) Institut fuer Medizinische Genetik und Humangenetik, Charité-Universitàtsmedizin Berlin. , Augustenburger Platz 1, 13353 Berlin, Germany; 3) Max-Planck-Institut fuer Molekulare Genetik, FG Development & Disease, Ihnestr. 63-73, 14195 Berlin, Germany; 4) Institute of Medical Biology, A*STAR, Singapore, Singapore; 5) Children's Hospital of Philadelphia, United States; 6) Mass General Hospital for Children, Boston, United States; 7) Departamento de Genética Humana, Instituto Nacional de Pediatría, Mexico City, Mexico; 8) Institute of Molecular and Cellular Biology, A*STAR, Singapore, Singapore; 9) Department of Clinical Genetics, Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland; 10) Division of Medical Genetics, Department of Pediatrics, University of Tennessee; Memphis, United States; 11) Labor-Berlin, Berlin, Germany; 12) Department of Genetics, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 13) United Arab Emirates University, Faculty of Medicine and Health Sciences, Departments of Pediatrics (L. A.), Pathology (B. A. A.) and Radiology (R. L.), Al Ain, United Arab Emirates; 14) Max-Planck-Institut fuer Molekulare Genetik, Mass-Spectrometry facility, Ihnestr. 63-73, 14195 Berlin, Germany; 15) Berlin-Brandenburg Center for Regenerative Therapies, Charité-Universitàtsmedizin Berlin, Germany; 16) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, Washington, United States; 17) Department of Pediatrics, Faculty of Medicine, University of Jordan, Amman, Jordan; 18) Department of Women's and Children's Health, University of Otago, Dunedin, New Zealand; 19) Department of Pathology, University of Washington, Seattle, Washington, United States; 20) Department of Paediatrics, National University of Singapore, Singapore.

Progeroid disorders overlapping with De Barsy syndrome (DBS) are collectively denoted as autosomal recessive cutis laxa type 3 (ARCL3). They are caused by biallelic mutations in *PYCR1* or *ALDH18A1*, encoding pyrroline-5-carboxylate reductase 1 and pyrroline-5-carboxylate synthase (P5CS), respectively, which both operate in the mitochondrial proline cycle. We report here on eight unrelated individuals born to non-consanguineous families clinically diagnosed with DBS or wrinkly skin syndrome. We found three heterozygous mutations in the *ALDH18A1* gene leading to missense mutations of the same highly conserved residue p. Arg138 in P5CS. A *de novo* origin was confirmed in all six probands for whom parental DNA was available. Using patient derived fibroblasts and heterologous overexpression we found that the P5CS-R138W protein was stable, able to interact with wild type P5CS, but showed an altered sub-mitochondrial distribution. A reduced size upon native gel electrophoresis indicated an alteration of the structure or composition of P5CS mutant complex. Furthermore, we found that the mutated cells had a reduced P5CS enzymatic function leading to lower proline synthesis rates, which may account for observed alterations in collagen synthesis. In summary, recurrent *de novo* mutations of the highly conserved p. Arg138 residue of P5CS cause a novel autosomal dominant form of cutis laxa with progeroid features. Our data provide novel insights into the etiology of cutis laxa diseases and will have immediate impact on diagnostics and genetic counseling.

3026T

Single cell allele specific expression (ASE) in T21 and common trisomies: a novel approach to understand Down syndrome and common aneuploidies. *G. Stamoulis[1], P. Makrythanasis[1], F. Santoni[1], A. Letourneau[1], M. Guipponi[2], M. Garieri[1], N. Panousis[1], E. Falconnet[1], P. Ribaux[1], C. Borel[1], S. E. Antonarakis[1,2,3].* 1) University of Geneva, Geneva, Switzerland; 2) Service of Genetic Medicine, Geneva University Hospitals-HUG, Geneva, Switzerland; 3) iGE3 Institute of Genetics and Genomics of Geneva, University of Geneva, Switzerland.

Trisomy 21 is a model disorder of altered gene expression. We have previously used a pair of monozygotic twins discordant for trisomy 21 to study the global dysregulation of gene expression, without the noise due to genetic variation among individuals(*Nature:*508; 345-350;2014). The majority of previous studies on aneuploidies were conducted -n cell populations or tissues. However studies on gene and allelic expression of single cells (SC) may reveal important biological insights regarding the cellular impact of aneuploidy and elucidate the fundamental mechanisms of gene dosage. We calculated the allele specific expression (ASE) fromRNAseq of ~1000 single cells in different aneuploidies. We used 352 SC fibroblasts (172 Normal and 179 T21 cells) from the pair of monozygotic twins discordant for T21, 166 from a mosaic T21, 124 from a mosaic T18, 151 from a mosaic T8, and 146 SC fibroblasts from a mosaic T13. In the monozygotic twins, a considerable number of heterozygous sites throughout the non-chr21 genome were expressed monoallelically (Normal: 73. 2 % monoallelic in 559,134 observations, and T21: 78. 8 % monoallelic in 573,670 observations). There was also considerable monoallelic expression for chr21 sites in Normal and, surprisingly, in T21 cells as well (Normal: 67,2 % monoallelic in 4,985 observations, and T21: 76. 07 % monoallelic in 6,723 observations). We classified the genes on chr21 in three classes according to the level of the aggregate monoallelic expression of their corresponding sites (9 monoallelic, 29 intermediate, 2 biallelic). Our hypothesis is that each class of genes contributes in a specific way to the phenotypic variability of Down Syndrome. Similar results, i. e. extensive monoallelic expression of genes on the supernumerary chromosomes, were also observed in the other aneuploidies. Our analysis demonstrated that, for genes with monoallelic expression, the abnormal gene dosage induced by the aneuploid chromosome is actually rather due to the number of cells expressing the gene and not to the level of expression per gene. This difference in the fraction of expressing cells could contribute to the development and the variability of phenotypes in aneuploidies. This study provides a new fundamental understanding of the allele specific expression in T21 and common aneuploidies.

3027F

Investigation of Variants of Uncertain Significance Identified Through Exome Sequencing in 46,XY Disorders of Sex Development Using the C57BL/6J-Y^{POS} Mouse Model. *H. Barseghyan[1], R. Baxter[1], V. Arboleda[1], A. Eskin[1], S. Nelson[1], E. Delot[1,2], E. Vilain[1,2].* 1) Human Genetics, David Geffen School of Medicine, UCLA, Los Angeles, CA; 2) Pediatrics, David Geffen School of Medicine, UCLA, Los Angeles, CA.

Disorders of Sex Development (DSD) are defined as “congenital conditions in which development of chromosomal, gonadal, or anatomic sex is atypical. ” These conditions have an approximate frequency of 0. 5-1% of live births and encompass a wide variety of urogenital abnormalities ranging from mild hypospadias to sex reversal. We have performed exome sequencing (WES) to identify the underlying genetic cause in patients diagnosed with 46,XY DSD. In the majority of cases a specific genetic diagnosis was not found, nevertheless WES yielded large numbers of variants of uncertain significance (VUS). To investigate the relevance of these VUS in regards to the patient’s phenotype, we utilized a powerful mouse model for studying undervirilization in 46,XY individuals, in which the presence of a Y chromosome originating from a *M. domesticus poschiavinus* strain (YPOS) on C57BL/6J (B6) background results in XY undervirilization and sex reversal. We hypothesized that abnormal gonadal expression of specific genes in B6-YPOS males during gonadal development would correlate with VUS in genes of 46,XY DSD patients identified through WES. We isolated gonadal tissue from wild type (WT) B6 and undervirilized B6-YPOS males at embryonic day 11. 5 and performed RNA sequencing in order to assess differential gene expression. We identified 110 genes whose expression was 1. 5-38 fold lower in B6-YPOS gonads compared to WT gonads, and in which a missense variant with an alternate allele frequency of less than 1% was identified in unexplained 46,XY DSD cases. Among these genes, three (*FBLN2*, *ADAMTS16*, *CYP26B1*) had variants predicted to be damaging by *in silico* tools. This method allowed identification of novel candidate genes, mutations in which could potentially prove to cause 46,XY DSD. Biochemical *in vitro* and knock-down *in vivo* experiments are in progress to validate the deleterious effects of the variants.

3028T

The *APOE* Locus and MicroRNA in Alzheimer’s Disease. *L. Bekris[1,3], M. Shaw[1], J. B. Leverenz[2], Y. Shao[1].* 1) Genomics Medicine Institute, Cleveland Clinic Lerner Research Institute, Cleveland, OH; 2) 2. Cleveland Clinic, Neurological Institute, Cleveland Clinic Lou Ruvo Center for Brain Health, Cleveland, OH; 3) Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, OH.

Background: MicroRNA (miRNA) play an essential role in post-transcriptional gene regulation in the brain. Genetic variants located across the *APOE* locus are the strongest risk factors for late-onset AD described to date. A complex regulatory structure exists at the *APOE* locus that includes putative regulatory sites within the *TOMM40* gene. Little is known about the influence of miRNA post-transcriptional regulation at this locus. **Objective:** The aim of this investigation is to demonstrate that miRNA that are predicted to target the *APOE* locus are expressed differently in brain of AD compared to controls and associated with *APOE* and *TOMM40* brain expression. **Method:** MiRNA array and miRNA qRT-PCR were used to measure miRNA expression in post-mortem brain from AD (n=21) or cognitively normal age-matched control (n=21) hippocampus (HP) and cerebellum (CB). Quantitative RT-PCR was used to measure *APOE* and *TOMM40* mRNA. Western blots were used to measure brain *APOE* and *TOMM40* protein. Linear regression was used to determine if *APOE* and *TOMM40* mRNA expression correlate with protein or mRNA expression in AD (n=8) and cognitively normal age-matched controls (n=8) or in HP and CB. **Results:** MiRNA levels (miR-370, miR-331-3p, miR-744, miR-135a) correlate with brain *TOMM40* or *APOE* expression in AD and controls. **Conclusion:** These results suggest that a complex regulatory structure at the *APOE* locus may be fine-tuned by miRNA post-transcriptional modulation according to brain region or disease status.

3029F

A new method for the functional analysis of regulatory SNPs. *H. Guillen Ahlers[1], D. Perumalla[1], P. Rao[1], A. Jadhav[1], J. M. Proffitt[1], M. J. Montoya[1], M. R. Shortreed[2], L. M. Smith[2], M. Olivier[1].* 1) Department of Genetics, Texas Biomedical Research Institute, San Antonio, TX; 2) Department of Chemistry, University of Wisconsin, Madison, WI.

A large number of disease-associated sequence variants have been identified in the human genome. Genome-wide association studies and, more recently, comprehensive whole-genome sequencing analyses have uncovered variants associated with a wide range of disease-related traits. Most of these associations are located in non-coding regions, indicating a likely regulatory mechanism. Not surprisingly, a number of these sequence variants are strongly associated with both disease-related traits and the expression of individual genes (expression quantitative trait loci, eQTL), reinforcing a functional link between gene expression regulation and disease risk. We recently developed a novel approach, Hybridization Capture of Chromatin-Associated Proteins for Proteomics (HyCCAPP), designed as a tool to identify all of the proteins bound to a specific target chromatin region by mass spectrometry. This approach does not require any prior knowledge of putative binding proteins, and helps discover novel proteins bound to a particular genome region. Here, we present the adaptation of the HyCCAPP approach for the analysis of luciferase reporter plasmid constructs commonly used to assess the impact of sequence variants on promoter activity, and apply the technology to the analysis of selected eQTL variants reported in the literature. As a proof of concept analysis, we analyzed a promoter variant, -105G/A, upstream of the selenoprotein S (*SELS*) gene. *SELS* is a gene involved in stress response and inflammation control. It has been shown before that this mutation leads to decreased levels of *SELS* expression. We tested the two alleles of the -105 variant in luciferase reporter assays introducing promoter-containing plasmids in HepG2 cells. The two variants resulted in differential activity in the reporter assay, suggesting that the two alleles likely result in altered gene expression. ChIP-seq data from ENCODE has not revealed the likely DNA-binding proteins mediating the observed effect. We are analyzing with mass spectrometry the DNA-protein interactions at the SELS promoter using our HyCCAPP technology. The method is being implemented in such a way that experiment conditions will be applicable to any promoter under 1 kb in length. Thus we are developing an application that will lead to the establishment of a new tool for scientists focusing on the functional characterization of putative regulatory variants associated with gene expression and a range of diseases.

3030T

Sequence variation affecting transcription factor occupancy in highly diverged mouse strains. *M. T. Maurano[1,2], J. M. Halow[3], R. Byron[4], M. Groudine[4,5], M. A. Bender[4,6], J. A. Stamatoyannopoulos[3,7].* 1) Institute for Systems Genetics, New York University Langone Medical Center, New York, NY 10016, USA; 2) Department of Pathology, New York University Langone Medical Center, New York, NY 10016, USA; 3) Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA; 4) Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA; 5) Department of Radiation Oncology, University of Washington, Seattle, WA 98109, USA; 6) Department of Pediatrics, University of Washington, Seattle, WA 98195, USA; 7) Division of Oncology, Department of Medicine, University of Washington, Seattle, WA 98195, USA.

Assessment of the functional consequences of sequence variation at non-coding regulatory elements is complicated by their high degree of context sensitivity to both the local chromatin and nuclear environments. Allelic profiling of chromatin accessibility across individuals has shown that a minority of sequence variation affects transcription factor (TF) occupancy, yet the low sequence diversity in human populations means that no experimental data are available for the majority of disease-associated variants. Using deep allelic profiling of chromatin accessibility, we generated high-resolution maps of variants affecting TF activity at their native locus and chromatin environment in liver, kidney, lung and B cells from 5 increasingly diverged strains of F1 hybrid mice. The high density of heterozygous sites in these strains enables a precise quantification of the effect size and cell-type specificity of variants associated with altered chromatin accessibility throughout the genome. We show that functional variation delineates characteristic sensitivity profiles for hundreds of TF motifs, representing nearly all important TF families. Nevertheless, silent variants are found repeatedly at every position within the protein-DNA recognition interface, and the majority of variation is buffered in a site-dependent manner in vivo. We develop a compendium of TF-specific sensitivity profiles to account for genomic context effects and and model cell-type specific perturbations of the protein-DNA interface. Taking advantage of the high conservation of TF coding sequence among mammals, we apply these models to the classification of disease-associated single nucleotide polymorphisms in human. These models enable quantitative prediction of the effect of non-coding variation on TF activity across a variety of cellular contexts, facilitating both fine-mapping and systems-level analyses of common disease-associated variation in human genomes.

3031F

A high-throughput functional screening identifies non-canonical *cis*-regulatory sequences in the *OCT4* Locus. *Y. Shen[1,2], Y. Diao[2], B. Li[2], Z. Meng[3], J. Dixon[3], I. Jung[2], A. Lee[2], K. Guang[3], B. Ren[2,3].* 1) Institute for Human Genetics, San Francisco, CA; 2) Ludwig Institute for Cancer Research; 3) UCSD.

With less than 2% of the human genome coding for proteins, a major challenge confronting researchers today is to interpret the function of the non-coding DNA. Millions of regulatory sequences have been predicted in the human genome through the analyses of DNA methylation, chromatin modification, hypersensitivity to nucleases, and transcription factor binding, but few have been shown to regulate transcription in their native contexts. To begin to address this issue, we have developed a high throughput CRISPR/Cas9-based genome-editing strategy and used it to interrogate 174 candidate regulatory sequences within the 1Mbp *OCT4* locus in the human ES cells. We identified two classical regulatory elements - including a promoter and a proximal enhancer - that are essential for *OCT4* transcription in the ES cells. Unexpectedly, we also discovered three non-canonical *cis* elements that contribute to *OCT4* transcription in an unusual way: disruption of such sequences led to a temporary loss of *OCT4* transcription that is fully restored after a few rounds of cell division. These sequences harbor signatures of *cis*-regulatory elements but do not support reporter gene expression by themselves. These results demonstrate the utility of CRISPR/Cas9 for functional characterization of non-coding DNA, and reveal a previously unrecognized layer of gene regulation in human cells.

3032T

Genetic architecture of gene expression regulation via orthogonal tissue decomposition. *H. E. Wheeler[1,2], N. Knoblauch[3], N. J. Cox[4], D. L. Nicolae[2], H. K. Im[2], GTEx Consortium.* 1) Departments of Biology and Computer Science, Loyola University Chicago, Chicago, IL; 2) Department of Medicine, University of Chicago, Chicago, IL; 3) Committee on Genetics, Genomics, and Systems Biology, University of Chicago, Chicago, IL; 4) Division of Genetic Medicine, Vanderbilt University, Nashville, TN.

Regulatory variation has been shown to play a key role in the genetics of complex traits. While many common diseases have been shown to be polygenic, it is unclear whether gene expression levels are also polygenic or instead have simpler genetic architectures and how much these expression architectures vary across genes. First we sought to quantify the relative proportion of local vs. distal gene regulation by estimating the proportion of gene expression variance explained by each (i. e. local vs. distal heritability, h2). With 922 whole blood RNA-seq samples, the mean local h2 was 0. 13 and 54. 6% of genes had a positive 95% confidence interval (CI), while the mean distal h2 was 0. 076 and just 4. 2% of genes had a positive CI. In order to better understand the genetic architecture of tissue-specific and cross-tissue gene regulation, we developed a model called orthogonal tissue decomposition (OTD) and applied it to the tissue rich GTEx dataset. We found that estimates of the local h2 of cross-tissue gene expression generated with our OTD model have larger magnitude and lower standard errors compared to single tissue estimates due to the borrowing of information across all samples. We also assessed the ability of various models with different underlying assumptions to predict gene expression in order to understand the underlying genetic architecture of gene expression. Gene expression traits with sparse architecture should be better predicted with sparse models such as LASSO (Least Absolute Shrinkage and Selection Operator) while highly polygenic traits should be better predicted with ridge regression. The elastic net model combines both models by mixing these two penalized regression approaches, LASSO and ridge regression. A mixing parameter in the elastic net model controls the degree of sparsity of the model. For each tissue in GTEx, we varied the mixing parameter and assessed the prediction performance. We found that for the majority of genes, ridge regression performs worse than elastic net or LASSO supporting the hypothesis that for most genes the local genetic architecture is sparse rather than polygenic. We also found evidence that genes with large tissue-specific heritability are enriched in common complex disease genes discovered via GWAS.

3033F

Identifying compound heterozygous variants associated with gene expression. *R. Brown[1], G. Kichaev[1], N. Mancuso[2], B. Pasaniuc[1,2,3].* 1) Bioinformatics IDP, UCLA, Los Angeles, CA, USA; 2) Department of Pathology and Laboratory Medicine, Geffen School of Medicine, UCLA, Los Angeles, CA, USA; 3) Department of Human Genetics, Geffen School of Medicine, UCLA, Los Angeles, CA USA.

Compound heterozygotes, common in monogenic traits, are not well characterized by current models for complex and quantitative traits. This decreases the ability to find risk loci and may reduce fine-mapping performance. In this work we propose approaches that model compound heterozygotes to increase power for association and fine-mapping. We use mRNA expression from 373 European individuals in the Geuvadis data (Lappalainen et al. 2013 Nature) to show that our approaches increase performance over existing methods. For example in an exon of the NIPSNAP1 gene we find two variants with frequencies . 19 and . 27 in linkage disequilibrium (r) of -0. 25 that individually attain Bonferroni corrected p-values of 0. 033 and 0. 045 for association with the expression level. However, a compound heterozygote model of these variants yields a p-value of 8. 5e-8 Bonferroni corrected for the additional pairwise tests. Across all chromosome 22, 35% of exons show significant associations between expression level and a variant or compound heterozygote. 19% of the significant associations include a significant compound heterozygous association. In 1% of associated exons only compound heterozygotes reached Bonferroni significance demonstrating that our method uncovers novel associations.

3034T

2p15-p16. 1 microdeletions encompassing and proximal to *BCL11A* are associated with elevated HbF in addition to neurologic impairment. A. P. W. Funnell1, P. Prontera2, V. Ottaviani3, M. Piccione4, A. Giambona5, A. Maggio5, F. Ciaffoni6, S. Stehling-Sun1, M. Marra6, F. Masiello6, L. Varricchio7, A. R. Migliaccio7,8, T. Papayannopoulou9, J. A. Stamatoyannopoulos1,10. 1) Department of Genome Sciences, University of Washington, Seattle, WA; 2) Centro di Riferimento Regionale Genetica Medica, Azienda Ospedaliera-Universitaria di Perugia, Perugia, Italy; 3) Genetica Medica, IRCCS “Ospedale Casa Sollievo della Sofferenza”, San Giovanni Rotondo, Foggia, Italy; 4) Dipartimento Scienze per la promozione della salute e materno-infantile “G. D’Alessandro” Université degli Studi di Palermo, Palermo 90127, Italy; 5) U. O. C. di Ematologia delle Malattie Rare del Sangue e degli Organi Emopoietici. AOOR Villa Sofia Cervello, Palermo 90146, Italy; 6) Department of Biologia Cellulare and Neuroscience, Istituto Superiore Sanita, Viale Regina Elena 299, 00161 Rome, Italy; 7) Department of Medicine, Mount Sinai School of Medicine, New York, NY 10029, USA; 8) Department of Biomedical Sciences, University of Bologna, via Irnerio 48, 40126 Bologna, Italy; 9) Division of Hematology, Department of Medicine, University of Washington, Seattle, WA 98195, USA; 10) Division of Oncology, Department of Medicine, University of Washington, Seattle, WA 98195, USA.

β-hemoglobinopathies, such as sickle cell anemia and β-thalassemia, are amongst the most common genetic disorders worldwide. These diseases arise from mutations that affect the function or expression of adult β-globin (HBB). Treatment typically involves lifelong blood transfusion and chelation therapy, with the only curative option being allogeneic transplantation of hematopoietic stem cells. However, it has long been known that increased fetal hemoglobin (HbF) expression that persists into adulthood considerably ameliorates the clinical severity of these disorders. Reversing the perinatal silencing of HbF has thus become a primary focus of therapeutic efforts. In recent years, the transcription factor BCL11A has emerged as a key, and potent, repressor of HbF. Disrupting BCL11A function in an erythroid-specific fashion is rigorously being sought as a therapeutic option; however, the potential impact of such disruption in humans has been unclear. This is due, in part, to the scarcity of reported mutations affecting the coding region of *BCL11A*. Recently however, a number of patients have been described with shared neurodevelopmental defects arising from chromosome 2p15-p16. 1 microdeletions that variably cover segments of the *BCL11A* locus. Here, we have examined the hematological phenotype of three such patients with distinct *de novo* microdeletions. All three display modestly reduced *BCL11A* expression in erythroblasts, and elevated HbF in peripheral blood (approximately 5-10%). Interestingly, in one patient, the *BCL11A* coding gene is intact, with only a downstream region deleted. We have identified a number of potential regulatory elements within this segment that are co-occupied by multiple principal erythroid transcription factors. We are currently employing high throughput genome editing techniques to dissect the functionality of these candidate elements, with the anticipation that they might proffer attractive targets for therapeutic intervention in the treatment of β-hemoglobinopathies.

3035F

Functional footprinting of regulatory DNA. *J. Vierstra[1], A. Reik[2], K. H. Chang[1,3], S. Stehling-Sun[1], N. Psatha[1,3], C. M. O'Neill[2], A. Mich[2], J. Miller[2], G. Lee[2], S. Tan[2], H. Jiang[2], G. Stamatoyannopoulos[4], T. Papayannopoulou[3], E. J. Rebar[2], P. D. Gregory[2], F. D. Urnov[2], J. A. Stamatoyannopoulos[1,5].* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Sangamo BioSciences, Pt. Richmond, CA; 3) Division of Hematology, Department of Medicine, University of Washington, Seattle, WA; 4) Division of Medical Genetics, Department of Medicine, University of Washington, Seattle. WA; 5) Division of Oncology, Department of Medicine, University of Washington, Seattle, WA.

Transcription regulatory regions harbor the majority of human disease-associated sequence variants. However, our understanding of the impact of regulatory DNA variation is severely limited by the difficulty of precisely assigning the functional contribution of individual nucleotides to phenotypic outcomes. We describe a facile and scalable approach, functional footprinting, to map functional elements within regulatory DNA at nucleotide resolution. We exploit the error-prone nature of double strand break repair triggered during targeted genome editing yielding indels of variable size (1 to >10 nucleotides, typically 2-6 nt), which can be exploited to create a broad spectrum of variant regulatory alleles within a single experimental cycle. Coupling these alleles to a functional readout, such as protein or RNA levels, pinpoints the contribution of specific nucleotides to regulatory activity. We demonstrate the functional footprinting method on the erythroid enhancer of BCL11A, a transcriptional repressor of fetal hemoglobin production in adult erythroid cells. Naturally-occurring variants within this region are associated with reduced *BCL11A* expression, with consequent elevation of fetal globin (γ-globin) to levels that are clinically ameliorative for sickle cell disease and beta thalassemia. Our results reveal individual TF recognition sites within the erythroid enhancer that gate the majority of downstream regulatory function on a fetal globin expression and, critically, these disruptions have the potential to affect a therapeutically meaningful outcome. Functional footprinting encompasses a simple and generalizable strategy to dissect the function of individual *cis*-regulatory elements such that any molecular sensor (i. e. , protein, RNA, etc.) can be linked to the function of individual base pairs within non-coding DNA.

3036T

Deciphering the control of HLA-C expression using the 1000 genomes dataset. *N. Vince[1,2], H. Li[1], S. Anderson[1], M. Carrington[1,2].* 1) Cancer and Inflammation Program, Laboratory of Experimental Immunology, Leidos Biomedical Research Inc. , Frederick National Laboratory for Cancer Research, Frederick, MD; 2) Ragon Institute of MGH, MIT and Harvard, Boston, MA.

The cell surface expression of HLA-C is differentially regulated across alleles, with a continuum of expression level from low (e. g. HLA-C*03) to high (e. g. HLA-C*14). Higher HLA-C expression level associates with better HIV control and with increased risk of developing Crohn's disease. The regulation of HLA-C expression is yet to be fully characterized. One mechanism that accounts in part for differential expression involves an insertion/deletion polymorphism in the 3'UTR of the HLA-C gene, which regulates binding of the microRNA miR-148a. However, this polymorphism does not perfectly mark allele-specific HLA-C expression level, implicating additional mechanisms. In order to further understand the control of HLA-C expression we used the mean expression of HLA-C alleles (as determined by flow cytometry) to impute HLA-C expression levels of individuals from the 1000 genomes dataset (1KG) based on their HLA-C genotypes. Logistic regression was applied to test the association between HLA-C expression and SNPs in the HLA region; HLA-C expression was used as a continuous variable. The most significant association was identified for a SNP located 800 bp upstream of the transcription start site (rs2395471, p=4. 2x10-66). According to online prediction tools, this SNP is located within a potential binding site for transcription factors (TF) Oct1 (*POU2F1*), Oct2 (*POU2F2*) and Oct4 (*POU5F1*). We were able to confirm the presence of a TF binding site by performing an electrophoretic mobility shift assay (EMSA), in which we used Jurkat cell line nuclear protein extract and an oligonucleotide specific for each genotype of rs2395471. Next, using TF specific antibodies, we confirmed that Oct1, but not Oct2 or Oct4, binds this genomic site. Further experiments are planned to test the potential influence of rs2395471 on TF binding and promoter activity. In conclusion, using computational tools and the 1KG dataset, we were able to identify a promoter SNP strongly associated with HLA-C expression, which may provide a novel mechanism of HLA-C gene regulation.

3037F

A LARGE GENOMIC DELETION LEADS TO ENHANCER ADOPTION BY THE LAMIN B1 GENE: A SECOND PATH TO AUTOSOMAL DOMINANT LEUKODYSTROPHY (ADLD). *E. Giorgio[1], D. Robyr[2], M. Spielmann[3], E. Ferrero[1], E. Di Gregorio[4], D. Imperiale[5], G. Vaula[6], G. Stamoulis[2], F. Santoni[2], C. Atzori[5], C. Mancini[1], S. Cavalieri[4], L. Gasparini[7], D. Ferrera[7], C. Canale[8], M. Guipponi[2], L. A. Pennacchio[9], S. E. Antonarakis[2], A. Brussino[1], A. Brusco[1].* 1) University of Torino, Department of Medical Sciences, 10126, Torino, Italy; 2) University of Geneva Medical School, Department of Genetic Medicine and Development, 1211, Geneva, Switzerland; 3) Max Planck Institute for Molecular Genetics, Ihnestr. 63-73, Berlin 14195, Germany; 4) "Città della Salute e della Scienza" University Hospital, Medical Genetics Unit, 10126, Torino, Italy; 5) Centro Regionale Malattie Da Prioni - Domp (ASLTO2), 10100, Torino, Italy; 6) Città della Salute e della Scienza University Hospita, Department of Neurology, 10126, Torino, Italy; 7) Istituto Italiano di Tecnologia, Deartment of Neuroscience and Brain Technologies, 16163 Genoa, Italy; 8) Istituto Italiano di Tecnologia, Nanophysics, 16163 Genoa, Italy; 9) 9 Lawrence Berkeley National Laboratory, Genomics Division, MS 84-171, Berkeley, CA 9472, USA.

Chromosomal rearrangements with duplication of the lamin B1 gene (*LMNB1*) underlie autosomal dominant adult-onset leukodystrophy (ADLD), a rare neurological disorder in which overexpression of *LMNB1* causes progressive CNS demyelination. However, we previously reported an ADLD family (ADLD-1-TO) without evidence of duplication or other mutation in *LMNB1* despite linkage to the *LMNB1* locus and lamin B1 overexpression. By custom array-CGH, we further investigated this family and report here that patients carry a large (~660 kb) heterozygous deletion that begins 66 kb upstream of the *LMNB1* promoter. Lamin B1 overexpression was confirmed in further ADLD-1-TO tissues and in a postmortem brain sample, where lamin B1 was increased in the cortex but not in cerebellum. Through parallel studies, we investigated both loss of genetic material and chromosomal rearrangement as possible causes of *LMNB1* overexpression, and found that ADLD-1-TO plausibly results from an enhancer adoption mechanism. The deletion eliminates a genome topological domain boundary, consenting normally forbidden interactions between at least three forebrain-directed enhancers and the *LMNB1* promoter, in line with the observed mainly cerebral localization of lamin B1 overexpression and myelin degeneration. This second route to *LMNB1* overexpression and ADLD is a new example of the relevance of regulatory landscape modifications in determining Mendelian phenotypes.

3038T

Analysis of transcriptional regulation by Dentatorubral-Pallidoluysian Atrophy protein (DRPLAp) that acts as transcriptional co-regulator. *K. Hatano[1], H. Date[1], H. Ishiura[1], J. Mitsui[1], J. Goto[1,3], J. Yoshimura[2], K. Doi[2], S. Morishita[2], S. Tsuji[1].* 1) Department of Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo; 2) Department of Computational Biology, Graduate School of Frontier Sciences, The University of Tokyo, Chiba; 3) Department of Neurology, International University of Health and Welfare Mita Hospital, Tokyo.

Background: Dentatorubral-Pallidoluysian atrophy (DRPLA [MIM 125370]) is an autosomal dominant neurodegenerative diseases caused by unstable expansion of CAG repeats in the coding sequence of DRPLA gene (*ATN1*). Previous studies have shown that DRPLA protein (DRPLAp) functions as a transcriptional co-regulator. Its target genes, however, remains unknown. **Purpose:** To determine the target genes regulated by wild type DRPLAp based on expression profiling of RNAs in cultured cells expressing full-length wild-type DRPLAp, and to investigate altered transcriptional regulations exerted by mutant DRPLAp based on comparison of expression profiling of RNAs between cultured cells expressing wild-type DRPLAp and those expressing mutant DRPLAp. **Materials:** HEK293 cell lines which express GFP-fused full-length DRPLA (Q19 or Q88) under the control by Tet (tetracycline) -On/Off system were used (Flp-In T-REx 293 cell). **Methods:** RNA was extracted from the cells at 24 hours after induction of expression by Tet. Strand-specific cDNA libraries were generated from ribosomal RNA-depleted RNA with random primers. Expression profiling was achieved by RNAseq analyses using HiSeq2500. Comparison between Q19 (ON) and Q19 (OFF) (Analysis 1) was performed to investigate the physiological functions of DRPLAp. Altered transcription profiling was analyzed by comparison between Q88 (ON) and Q88 (OFF) (Analysis 2). **Result:** The reproducibility of expression data of all genes analyzed from the cDNA libraries in triplicates were confirmed by Spearman's rank correlation coefficient (rho>0. 98). The number of up- or down-regulated were determined for two independent experiments. Twenty one genes were up-regulated and another set of 21 genes were down-regulated (Analysis 1). In contrast, comparison between Q88 (ON) and Q88 (OFF) revealed that 240 genes were up-regulated, whereas 72 genes were down-regulated (Analysis 2). **Conclusion:** The findings support the functions of wild-type DRPLAp as a transcriptional co-regulator. The findings that a much larger number of genes were up-regulated in cells expressing mutant DRPLAp than that in cells not expressing mutant DRPLAp raises a possibility that mutant DRPLAp with expanded CAG repeats may lead to altered transcriptional regulation relevant to neurodegeneration in DRPLA. Further studies using neuronal cells will be needed to clarify the altered transcriptional regulations in DRPLA.

3039F

SMC1B is present in mammalian somatic cells and interacts with mitotic cohesin proteins. *A. Musio[1], L. Mannini[1], V. Quarantotti[1], C. Amato[1], F. Cucco[1,2], M. Tinti[1], L. Tana[3], A. Frattini[4,5], I. Krantz[6], R. Jessberger[7].* 1) Human Genome Dept, Istituto di Ricerca Genetica e Biomedica, Pisa, Italy; 2) Dipartimento di Biologia, Université degli Studi di Pisa, Pisa; 3) Azienda Ospedaliero Universitaria Pisana, U. O. Fisica Sanitaria, Pisa; 4) Istituto di Ricerca Genetica e Biomedica, Consiglio Nazionale delle Ricerche, Milan; 5) Dipartimento di Medicina Clinica e Sperimentale, Université degli Studi dell'Insubria, Varese, Italy; 6) Division of Human Genetics, The Children's Hospital of Philadelphia and the Perelman School of Medicine at the University of Pennsylvania, USA; 7) Institute of Physiological Chemistry, Technische Universitét Dresden, Dresden, Germany.

Cohesin is an evolutionarily conserved protein complex that plays a role in many biological processes: it ensures faithful chromosome segregation upon cell division, regulates gene expression and preserves genome stability. Malfunctioning of cohesin has been linked to human multisystem developmental disorders such as Cornelia de Lange syndrome. In mammalian cells, the mitotic cohesin complex consists of two structural maintenance of chromosome proteins, SMC1A and SMC3, the kleisin protein RAD21 and a fourth subunit either STAG1 or STAG2. Meiotic paralogs in mammals were reported for SMC1A, RAD21 and STAG1/STAG2 and are called SMC1B, REC8 and STAG3 respectively. It is believed that SMC1B is only a meiotic-specific cohesin member, required for sister chromatid pairing and for preventing telomere shortening. Here, we show that SMC1B is also expressed in somatic mammalian cells, in addition to germinal cells. Genome-wide data indicates that cohesin-SMC1B binds to body genes, and in intergenic regions and SMC1B silencing leads to gene expression deregulation. Altogether, our data allows us to hypothesize that SMC1B takes part in transcription regulation.

3040T

Motif/Transcription factor prediction of brain expressed HERV-derived transcripts reveals enrichment of biological signature of transcriptional regulation. *G. Guffanti[1], S. Gaudi[2], J. Fallon[3], F. Macciardi[3].* 1) Department of Psychiatry, Harvard Medical School, Boston, MA; 2) Department of Infectious, Parasitic and Immune-Mediated Diseases, Italian National Institute of Health, Rome, Italy; 3) Department of Psychiatry and Human Behavior, University of California Irvine, Irvine, CA.

Background: Human Endogenous Retroviruses (HERVs) can be found embedded in the sequence of other genes, including introns and untranslated regions at either sides of a coding sequence. The vast majority of gene-embedded transposable elements (TE) appear to be harmless, but some of them are able to influence the transcription of their "host" genes. Although the precise mechanisms are yet to be elucidated, the prevailing theory is that gene-embedded TEs ultimately promote transcript diversity. **Objective:** Our aim is to identify biological processes influenced by dorsal-lateral prefrontal cortex (DLPFC) expressed genes containing HERVs within their sequence in a sample of 10 schizophrenia patients and 10 healthy controls. **Methods:** RNA samples were isolated from DLPFC, prepared for sequencing with Trueseq and sequenced on the Illumina HiSeq 2000. Transcripts were determined using Trinity de novo assembly and aligned to the reference genome for sequence homology using BLASTn and Repeatmasker for HERVs family and type assignment. We used iRegulon Cytoscape plugin to infer motif enrichment and prediction of Transcription Factors targeting enriched motifs of DLPFC expressed genes containing HERVs in their sequence (NES > 3). Then we used Reactome Cytoscape plugin to test biological processes enrichment in in each cluster of genes targeted by common transcription factors (TF) (FDR < 10-3). **Results:** Of the 59,258 transcripts with FDR < 1 and FC>2, 4,177 embedded HERVs in their sequence. Through sequence homology analysis, we were able to identify 13,985 genomic regions overlapping reference HERVs. Of these, 7,263 map to at least one protein-coding or RNA gene. iRegulon identified enrichment of three motifs in the 200 KB region at the TSS of the DLPFC expressed genes with HERV embedded in their sequence: ZNF193, LSM6 and MO1102, targeted by the predicted TFs ZSCAN9, SETDB1, TEAD1 and GATA3. Target genes enriched in the complex motif/TF LSM6/ZSCAN9 revealed enrichment of cell adhesion, synaptic transmission and GTPase signal transduction; those in the complex motif/TF ZNF193/SETDB1 revealed enrichment of DNA dependent regulation of transcription. **Conclusions:** Genes harboring HERVs in their non-coding and untranslated regions share common motifs in their promoter regions, which are in turn targeted by common Transcription Factors. Clustering of DLPFC expressed genes based on shared motif/TF allows identifying putative influence on common biological processes.

3041F

Post-GWAS identification of a *PRSS1* promoter variant protecting against pancreatitis. *A. Boulling[1,2,3], M. Sato[4], E. Masson[1,5], E. Génin[1], J. M. Chen[1,2,3], C. Férec[1,2,3,5].* 1) Institut National de la Santé et de la Recherche Médicale (INSERM), U1078, Brest, France; 2) Faculté de Médecine et des Sciences de la Santé, Université de Bretagne Occidentale (UBO), Brest, France; 3) Etablissement Français du sang (EFS) - Bretagne, Brest, France; 4) Section of Gene Expression Regulation, Frontier Science Research Center, Kagoshima University, Kagoshima, Japan; 5) Laboratoire de Génétique Moléculaire et d'Histocompatibilité, Centre Hospitalier Régional Universitaire (CHRU) Brest, Hôpital Morvan, Brest, France.

Introduction: Over the past two decades, genetic factors have been increasingly appreciated to play an important role in the etiology of chronic pancreatitis. Recently, a genome-wide association study (GWAS) found that the minor T allele of rs10273639, a common single nucleotide polymorphism (SNP) located 408 bp upstream of the translation initiation codon of the *PRSS1* gene (encoding cationic trypsinogen), was associated with protection against chronic pancreatitis. This protective effect is attributed to a decrease of *PRSS1* mRNA expression within pancreas. However, whether this allele is the causal variant or not remains unknown. **Methods:** The *PRSS1* promoter was sequenced in 287 French individuals of European ancestry to search for SNPs in linkage disequilibrium (LD) with rs10273639. Then, several luciferase reporter constructs were used to assess *PRSS1* promoter activity in AR42J cells treated with dexamethasone. The activity of the *PRSS1* promoter was also validated *in vivo* with a novel technic based on the intra-pancreatic parenchymal injection for gene transfer coupled to luciferase assay. **Results:** Here, we first showed that rs10273639 is in perfect LD with rs4726576, a SNP located 204 bp downstream of rs10273639, in French Caucasians. Employing promoter reporter gene assays in both AR42J cells treated with dexamethasone and in a mouse model, we showed that the proximal region of the *PRSS1* promoter is sufficient to drive basal expression. Then, the use of reporter constructs carrying variations corresponding to the different possible haplotypes provided evidence that only the minor A allele of rs4726576 is able to decrease *PRSS1* expression by altering its promoter activity. **Conclusion:** This result highlighted that it is the minor A allele of rs4726576 that was associated with reduced gene expression and therefore represents the causal variant underlying the GWAS-revealed protective effect. This finding also has implications for risk assessment in different populations, wherein the degree of LD between the two SNPs was found to vary significantly.

3042T

Understanding the regulatory network of chromosome 21 transcription factors. *A. Letourneau[1], G. Cobellis[2], E. Falconnet[1], P. Ribaux[1], C. Gehrig[1,3], M. Guipponi[1,3], C. Borel[1], S. E. Antonarakis[1,3,4].* 1) Genetic Medicine & Development, University of Geneva Medical School, Geneva, Switzerland; 2) Dpt of Biophysics, Biochemistry & General Pathology, Seconda Universita di Napoli, Italy; 3) University hospitals of Geneva, Switzerland; 4) iGE3, Institute of Genetics and Genomics of Geneva, Switzerland.

Down syndrome (DS or Trisomy 21) is most likely the result of a gene dosage alteration. However, uncovering the relationship between the DS phenotypes and the increased expression of candidate genes remains a challenge. We hypothesize that some of the DS-associated abnormalities are caused by the dosage imbalance of chromosome 21 (HSA21) transcription factors (TFs). We thus aimed to comprehensively investigate their potential role in the DS manifestations by revealing 1/ their binding sites and target genes, 2/ their underlying transcriptional regulatory networks and 3/ the complex interplay between them. To this end, we studied 12 relevant TFs encoded on HSA21 (Aire, Bach1, Erg, Ets2, Gabpa, Nrip1, Olig1, Olig2, Pknox1, Runx1, Sim2, 1810007M14Rik) as well as the Hmgn1 DNA binding protein. We used a unique cellular model of mouse embryonic stem (mES) cells overexpressing the Flag-tagged protein under the control of an inducible Tet-off system. For each TF, we combined ChIP-sequencing (3 replicates) and RNA-sequencing assays to identify the DNA binding sites and downstream targets. A pilot study on Sim2 (Letourneau *et al*, 2015) already revealed a list of 1229 high-confidence binding sites and confirmed the role of Sim2 in neuronal processes. More specifically, we will use the same system to investigate the role of Hmgn1 in the genome-wide gene expression dysregulation domains (GEDDs) recently described in trisomic cells (Letourneau *et al*, 2014). Given its role in chromatin modification, we hypothesized that the overexpression of the Hmgn1 DNA binding protein might contribute to the GEDDs pattern. Our preliminary RNA-sequencing analysis revealed that overexpression of Hmgn1 in mES cells induces GEDDs similar to those observed in the Ts65Dn mouse model. The analysis of the Hmgn1 binding profile provides the location of its preferential DNA binding sites and offers more insights into the molecular mechanisms contributing to the GEDDs. Altogether, these results will improve our comprehension of the transcriptional regulatory networks of HSA21 TFs and their role on cellular phenotypes associated with DS individuals.

3043F

Developmental expression patterns and alternate splicing of human transcription factors. *K. Siebenthall[1], E. Haugen[1], R. Sandstrom[1], A. Reynolds[1], M. Maurano[1], T. Canfield[1], K. Lee[1], R. Scott Hansen[1], E. Giste[1], S. Vong[1], P. Sabo[1], M. Diegel[1], D. Bates[1], A. Johnson[1], I. Glass[2], R. Kaul[1], J. Stamatoyannopoulos[1].* 1) Genome Sciences, University of Washington, Seattle, WA; 2) Genetics Division, Seattle Children's Hospital, Seattle, WA.

The ~1,600 transcription factors (TFs) encoded in the human genome orchestrate the progression of development, directing gene expression programs that establish the body plan, drive cellular differentiation and organize cellular communities into complex tissues. Most of the human TF repertoire remains functionally uncharacterized. In order to address this deficit, we performed RNA-seq of 15 organs spanning six weeks of human fetal development and examined TF expression patterns across these tissues. We find that 91% of human TFs are utilized during development. While we anticipated uncovering a large cohort of tissue-specific regulators, we instead found the majority of TFs expressed across all organ systems with elevated levels in one or two organs, suggesting that TF stoichiometry plays a role in tissue-specific gene expression programs. Strikingly, each organ uses at least one member of nearly every TF family. A broadly expressed TF may still adopt lineage-restricted roles through the tissue-selective use of alternate splicing. We find widespread alternate splicing among TFs that gives rise to thousands of distinct isoforms, with 14% of TFs expressing a transcript in a tissue-selective manner. Alternate splicing frequently modulates the structural configuration of TFs, including the wholesale removal of DNA-binding and protein-interaction domains. This behavior is especially prominent among the C2H2 zinc-finger family of TFs, which have expanded greatly in the primate lineage. We propose that TF isoforms lacking key structural domains generally serve a dominant-negative function, which can be especially potent in the case of TFs that must dimerize in order to bind DNA. We provide, to our knowledge, the most extensive study to date of TF expression patterns and alternate splicing during human development, resulting in a catalogue of 952 TFs with known or potentially novel roles in organogenesis.

3044T

Genetic and epigenetic signatures of gene regulation specific to type 2 diabetes-relevant tissues. *J. P. Didion[1], S. C. J. Parker[2,3], B. N. Wolford[1], J. R. Hyughe[4], R. Welch[2,3], M. R. Erdos[1], P. S. Chines[1], N. Narisu[1], L. J. Scott[4], M. Stitzel[5], M. Boehnke[4], F. S. Collins[1].* 1) NHGRI, Bethesda, MD. , USA; 2) Department of Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI, USA; 3) Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA; 4) Deptment of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; 5) The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA.

Multiple tissues are involved in glucose metabolism and the etiology of type 2 diabetes (T2D), including skeletal muscle, adipose, and pancreatic islets. Although hundreds of genetic variants have been associated with T2D and related traits, the tissue-specific effects of those variants are not well understood. To identify gene expression signatures in T2D-relevant tissues, the Finland United States Investigation of NIDDM Genetics (FUSION) Study obtained muscle and adipose biopsies from 278 Finnish individuals, and pancreatic islets from 78 unrelated cadaveric donors. We performed dense genotyping and deep mRNA sequencing on all muscle and islet samples; mRNA sequencing of adipose samples and whole-methylome sequencing of a subset of samples is in progress. In addition, we obtained genotype, RNA-seq and methylation data for T2D relevant tissues (including brain, liver, stomach, small intestine) from published studies and the Genotype-Tissue Expression (GTEx) project. We process all samples using a pipeline for imputation, transcriptome assembly, analysis of differential expression (including allele-specific expression) and splicing. We identify transcripts at the exon, gene, isoform and allelic levels that are enriched in each tissue, including hundreds of apparently novel transcripts. We also identify tissue-specific expression and splicing quantitative trait loci (e/sQTL), and we use reference chromatin state maps to quantify the proportion of e/sQTLs that occur in tissue-specific chromatin states. We have already found significant enrichment of SNPs associated with T2D and related traits in enhancer regions of skeletal muscle and islets. In addition, we will identify sites of differential methylation that are associated with e/sQTLs in a tissue-specific manner (mQTLs), which will suggest specific links between genetic and epigenetic variation in gene regulation. Because we have roughly equal numbers of male and female samples, we will also be able to identify sex-specific effects on expression and methylation. This resource will form the basis for further understanding the role of functional and regulatory variation in the contribution of each tissue to the etiology of T2D.

3045F

Polymorphic *Alu* insertions with functional consequences may be the causative variant identified in GWAS. *L. Horvath, J. Steranka, K. Burns.* Pathology, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD.

Transposable elements make up the majority of our genome, with the most abundant form being *Alu* interspersed repeats. Ongoing retrotransposition results in polymorphic insertions. Functional consequences of these have not yet been dissected, but insertions can lead to misregulation of nearby genes and altered phenotypes. As a first step to evaluate functional effects of common insertion variants, we focused on *Alu* insertions occurring near regions identified in genome wide association studies (GWAS) as being associated with a disease (p<10-8). A total of 793 polymorphic *Alu* insertions map to regions that are in linkage disequilibrium (LD) with GWAS signals (r2<0. 8). Given that the responsible functional variant has not been identified at most of these loci, the large (>300 bp) *Alu* insertions are attractive candidates. For a GWAS to have successfully identified a causative *Alu*, the insertion should be in LD with the trait-associated SNP (TAS). To measure LD, we genotyped the presence or absence of the *Alu* in reference individuals with known SNP genotypes. To date, 100 have been assayed and 43 insertions are in LD with nearby TASs. We have prioritized functional evaluation to insertions in LD with a TAS and also those mapping to likely regulatory regions based on ENCODE annotations. The potential for *Alu* insertions to alter transcript levels is being evaluated using enhancer assays with luciferase reporters. Comparing *Alu*-containing alleles to alleles without the insertion for over 100 loci, revealed a continuum of effects with significant outliers that up- or down-regulate luciferase activity. To address the mechanism underlying *Alu* sequence directed changes in transcription levels, the isolated effects of *Alu* subfamily consensus sequences were evaluated in this system. Four (4) types of *Alu* sequences altered reporter expression, 3 by upregulating reporter activity and one by decreasing reporter expression. This demonstrates that some *Alu* sequences have the intrinsic ability to alter transcription regulation. Additional candidates are being evaluated, and the mechanism for misregulation will be dissected for the most significant outliers. *Alu* insertions that are in LD with TASs and have functional effects in the reporter assay are especially promising candidates for further study. Our results indicate that *Alu* sequences have strong potential to alter transcriptional regulation and we have genetic evidence that they could contribute to common disease risk.

3046T

Characterization of a Sjögren's syndrome-associated long non-coding RNA at 2p25. 1. *J. A. Ice[1], I. Adrianto[1], H. Li[1,2], A. Rasmussen[1], G. B. Wiley[1], D. U. Stone[2], B. M. Segal[3], N. L. Rhodus[4], L. Radfar[2], J. A. James[1,2], C. G. Montgomery[1], R. H. Scofield[1,2,5], P. M. Gaffney[1], L. F. Thompson[1], A. D. Farris[1], S. Kovats[1], J. D. Wren[1], K. L. Sivils[1,2], C. J. Lessard[1,2].* 1) Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK; 2) University of Oklahoma Health Sciences Center, Oklahoma City, OK; 3) Hennepin County Medical Center, Minneapolis, MN; 4) University of Minnesota, Minneapolis, MN; 5) US Department of Veterans Affairs, Oklahoma City, OK.

Sjögren's syndrome (SS) [MIM: 270150] is a common autoimmune disorder characterized by immune-mediated exocrine gland destruction and systemic inflammation contributing to clinical heterogeneity. Complex regulatory mechanisms govern these responses, but are poorly understood. In a previous RNA-sequencing (RNA-seq) study, we identified >2,600 differentially expressed (DE) transcripts associated with SS. In this study, we sought to validate, replicate, and functionally characterize one upregulated long non-coding RNA (lncRNA) mapped to chromosome 2p25. 1 to better understand its role in SS pathogenesis. RNA-seq data showed significant upregulation of the 2p25. 1 lncRNA when comparing 27 healthy controls and 57 SS patients [*P*=8. 40x10-7; Fold Change (FC)=2. 54]. Of note, we observed a greater FC and stronger statistically significant DE in anti-Ro(-)/biopsy(+) patients (FC=2. 85, *P*=4. 21x10-6) compared to anti-Ro(+) patients (FC=2. 24, *P*=1. 62x10-4). Technical validation by qPCR confirmed upregulation of the 2p25. 1 lncRNA (*P*=0. 0096), and correlation with RNA-seq results was observed (*r*=0. 869). We next replicated these findings in an independent sample set of 36 SS patients and 21 controls (*P*=0. 0183). Using the bioinformatics tools GAMMA-seq, lncRNA2function, and GeneFriends, we identified co-expression of the 2p25. 1 lncRNA with transcripts in pathways related to T and NK cell signaling and activation. Using only RNA-seq data from anti-Ro(-)/biopsy(+) SS patients, we observed co-expression (*r*>|0. 7|) with >80 transcripts in pathways enriched for IL-12 and IL-2 signaling, immune response and regulation, and NK-T-cell lymphoma, among others. As lncRNAs can be cell-type specific, we assess 2p25. 1 lncRNA expression by qPCR in 9 FACS-isolated immune cell subsets from 2 patients and 2 healthy controls. We identified the strongest expression in CD8+ and CD4+ T cells as well as in CD56dim and CD56bright NK cells. Low expression was detected in CD1c+CD11c+ myeloid DCs, monocytes, and pDCs, but was not detectable in CD141+ myeloid DCs or B cells. We have identified, technically validated, and independently replicated the upregulation of a novel SS lncRNA at 2p25. 1. Furthermore, we have established the high expression of this transcript in CD4+ and CD8+ T cells and NK cells. This study establishes the 2p25. 1 lncRNA as the first associated with SS and lays the groundwork for further functional characterization in the pathogenesis of this complex disorder.

3047F

Using miRNA expression profiling and sparse machine learning methods for diagnostics of inflammatory bowel disease. *M. Huebenthal[1], G. Hemmrich-Stanisak[1], F. Degenhardt[1], S. Szymczak[1,2], Z. G. Du[1], A. Elsharawy[1], A. Keller[3], S. Schreiber[1,4], A. Franke[1].* 1) Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel, Germany; 2) Institute of Medical Informatics and Statistics, Christian-Albrechts-University of Kiel, Kiel, Germany; 3) Chair for Clinical Bioinformatics, Saarland University, Saarbruecken, Germany; 4) Department of Internal Medicine I, University Hospital Schleswig-Holstein, Kiel, Germany.

The diagnosis of inflammatory bowel disease (IBD) still remains a clinical challenge and the most accurate diagnostic procedure is a combination of clinical tests including invasive endoscopy. In this pilot study we evaluated whether systematic miRNA expression profiling, in conjunction with machine learning techniques, is suitable as a non-invasive diagnostic test for the major IBD phenotypes (Crohn's disease (CD) and Ulcerative colitis (UC)).

Based on microarray technology expression levels of 863 miRNAs were determined for whole blood samples from 40 CD and 36 UC patients and compared to data from 38 healthy controls (HC). To discriminate between disease-specific and general inflammation we included miRNA expression data from other inflammatory diseases (inflammation controls (IC): 24 chronic obstructive pulmonary disease, 23 multiple sclerosis, 38 pancreatitis and 45 sarcoidosis cases) as well as another 70 healthy controls from previous studies. Classification problems considering two, three or four groups were solved using different types of penalized support vector machines (SVMs). The resulting models were assessed regarding sparsity and performance and a subset was selected for further investigation. Measured by the area under the receiver operating characteristic (ROC) curve (AUC) the corresponding median holdout-validated accuracy was estimated as ranging from 0. 75 to 1. 00 (including IC) and 0. 89 to 0. 98 (excluding IC), respectively.

In combination the models provide highly accurate tools for solving different diagnostic problem such as the distinction of CD and UC as well as CD, UC and HC by incorporating not more than 16 distinct miRNAs. Experimentally validated target genes of these miRNAs have been previously described as being related to IBD and show significant enrichment for IBD susceptibility loci identified in earlier genome-wide association studies.

3048T

Identification of novel lincRNAs in human adipose and peripheral blood mononuclear cells. *C. Xue[1], X. Zhang[1], H. Zhang[1], J. Lin[1], J. Ferguson[2], M. Li[1], M. Reilly[1].* 1) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Department of Medicine, Vanderbilt University.

LincRNAs are a class of RNAs involved in various molecular functions and processes. Recent studies have identified lincRNAs across multiple tissues using RNA-seq. The power of lincRNA discovery is often limited by sequencing depth due to low expression of lincRNAs. In the Genetics of Evoked-responses to Niacin and Endotoxemia study, we performed deep RNA-seq of human adipose from 25 subjects and peripheral blood mononuclear cells (PBMCs) from 15 subjects, achieving mean depth of ~350 million and ~220 million 101bp paired-end reads per sample in adipose and PBMCs, respectively. Using *de novo* transcriptome assembly and a stringent filtering pipeline, we identified 270 novel lincRNAs from adipose and 109 from PBMC that have not been annotated in Gencode, RefSeq or previous studies. Analyses of these novel lincRNAs suggest that they are much more tissue specific than coding genes. Some of the most abundant novel lincRNAs were replicated in public RNA-seq dataset in relevant tissues. 44% of adipose novel lincRNAs are syntenic in mouse but only 40% of them have annotated lincRNAs from Gencode M4 in syntenic regions. A similar pattern was observed in PBMC novel lincRNAs. 53% of adipose novel lincRNAs and 26% of PBMC novel lincRNAs overlap predicted enhancers from Roadmap Epigenomics datasets. 20% of adipose novel lincRNAs and 14% of PBMC novel lincRNAs overlap predicted TSS. The pattern of enhancer overlap is similar using the stringent set of lincRNAs from Cabili et al. but the extent of TSS overlap is smaller in our study, which can be explained by the overall low expression level of novel lincRNAs. To gain insight into potential clinical importance of novel lincRNAs, we overlapped these lincRNAs with SNPs associated with cardiometabolic traits (waist-hip ratio adjusted BMI, plasma levels of triglycerides, cholesterol, HDL-cholesterol and LDL-cholesterol, as well as type 2 diabetes and fasting glucose) GWAS public data. Seven novel lincRNAs harbor GWAS significant SNPs within +/- 2kb of the lincRNAs. In 3 cases the top signals are in the lincRNA, not in neighboring coding genes and there is no strong LD between the SNPs in lincRNAs and those in coding genes, suggesting potential roles of these lincRNAs in mediating trait associations. In summary, we have identified and characterized a set of novel lincRNAs in human adipose and PBMC that will facilitate the study of lincRNA functions and their role in cardiometabolic disease.

3049F

Differential expression of microRNAs in breast cancer tumors associated to lymph node metastasis. *E. Pérez-Moreno[1], V. Zavala[1], V. Cornejo[2], W. Fernández[2], J. Gamboa[3], M. Carvallo[1].* 1) Department of Cellular and Molecular Biology, Pontificia Universidad Católica de Chile, Santiago, Santiago, Chile; 2) Unidad de Anatomía Patológica, Hospital Clínico San Borja Arriarán, Santiago, Chile; 3) Unidad de Patología Mamaria, Hospital Clínico San Borja Arriarán, Santiago, Chile.

Breast cancer is the most common malignancy in women worldwide and it is responsible for the highest cancer-associated death rates. It is a heterogeneous disease and can be subclassified into several subtypes. Lymph nodes near to the primary breast tumor have an increased chance of developing a secondary tumor, representing one of the first signs of metastasis in breast cancer. The presence of axillary lymph node metastasis is a critical prognostic factor for the application of specific surgical and/or therapeutic strategies, and the amount of positive lymph nodes is known to have an inverse linear correlation with prognosis and survival. MicroRNAs are small non-coding RNAs that are involved in post-transcriptional regulation. Their expression has been frequently described to be altered in different cancer types, including breast cancer. MicroRNAs have emerged as candidate molecular biomarkers and novel therapeutic targets because of their stability, easy detection and ability to regulate a large set of genes that are involved in cancer growth and metastasis. In this work, we analyzed microarray based miRNA expression data to identify microRNAs differentially expressed in primary breast tumors with lymph node metastasis. We extracted total RNA from 31 fresh frozen breast tumors (Invasive Ductal Carcinomas) with different tumor grades (1 to 3). Patients did not receive neoadyuvant chemotherapy. Within the analyzed tumors, 16 were obtained from patients with a detectable lymph node metastasis. Microarray data analyzed using RankProd (R package) revealed 26 microRNAs downregulated in tumors associated with lymph node metastasis ($p<0.05$). Among the identified microRNAs 77% have been previously described to regulate early steps of metastasis in different cancer types, like members of the miR-200 family (miR-200c and miR-141), miR-205 and miR-30a. In addition we found other microRNAs not previously associated with metastasis, like miR-1202 and miR-199a-3p. Our results suggest that downregulation of an important proportion of the identified microRNAs, may lead to an over expression of their target genes in primary breast tumors, inducing a metastatic behavior of tumor cells, promoting invasion and colonization of lymph nodes. In this sense, the loss of expression of these microRNAs may serve as a new biomarker and/or indicator of prognosis in breast cancer patients. Supported by FONDECYT 1120200.

3050T

microRNA expression profiling in breast cancer tumors associated to BRCA1 expression. *V. A Zavala[1], P. Gajardo[1], C. Alvarez[1], W. Fernandez[2], V. Cornejo[2], J. Gamboa[3], P. Carvallo[1].* 1) Pontificia Universidad Catolica de Chile, Santiago, Region metropolitana, Chile; 2) Unidad de Anatomía Patológica, Hospital San Borja Arriarán, Santiago, Chile; 3) Unidad de Patología Mamaria, Hospital San Borja Arriarán, Santiago, Chile.

BRCA1 is a tumor suppressor gene which mutations confer high susceptibility to develop breast cancer. In addition, it has been described that 50% of hereditary breast cancer tumors with no germline mutation loss BRCA1 expression, as well as 30% of sporadic breast tumors. In this relation, different somatic events have been described that silence BRCA1 such as promoter hypermethylation and genomic deletions of BRCA1 locus. These mechanisms do not completely explain BRCA1 loss in breast tumors. miRNAs constitute a relevant mechanism in silencing gene expression, since they have been found deregulated in different cancer types. Our aim was to identify miRNAs differentially expressed in tumors with low/absent BRCA1 protein. We isolated total RNA from 36 fresh frozen breast cancer tumors, not selected for family history. miRNA profiling was performed using 8x15K Human miRNA Microarrays (Agilent Technologies). Raw data was background corrected, normalized using quantiles and log2 converted. We considered for analyses only probes with an intensity 10% over the brightness of negative controls. As reference, for expression levels we used a commercial RNA pool derived from normal mammary glands from 5 donors. Differentially expressed miRNAs among BRCA1 positive and BRCA1 negative tumors were detected using RankProd Package in R. Eighteen miRNAs were found upregulated in BRCA1-deficient tumors ($p<0.05$). Prediction analyses using mirWalk revealed eight miRNAs that could regulate BRCA1 expression. In addition, pathway analysis using mirPath showed that one overexpressed microRNA, miR-575, is implicated in the regulation of different cancer pathways such as MAPK and PI3K-AKT signaling and cell cycle pathways. Among the targets described for miR-575 we found: BRAF, E2F2, PIK3CB, FGF1 and MDM2 genes. Interestingly, it has been described that inhibition of PI3K-AKT pathway, through the silencing of PIK3CA or PIK3CB impairs BRCA1 expression and sensitizes breast cancer cell lines to PARP inhibitors. Our results suggest that miR-575 produce an indirect silencing of BRCA1, through the regulation of PIK3CB. In relation to the other miRNAs, mirPath analysis identified different targets associated to secondary metabolic pathways.

3051F

Genetic variations within microRNAs and microRNA binding sites influence the risk of Parkinson's disease. *M. Ghanbari[1,2], S. Darweesh[1], H. de Looper[3], A. Hofman[1], M. A Ikram[1], O. H Franco[1], S. J Erkeland[3], A. Dehghan[1].* 1) Department of Epidemiology, Erasmus University Medical Center, Rotterdam, Netherlands; 2) Department of Genetics, School of medicine, Mashhad University of Medical Sciences, Mashhad, Iran; 3) Department of Hematology, Erasmus University Medical Center, Rotterdam, Netherlands.

Recently, a large number of genetic variants and genomic loci contributing to Parkinson's disease (PD) have been identified by genome-wide association studies (GWAS). However, the vast majority of the identified variants are thought to be merely proxies of functional variants and their causal mechanisms remain to be elucidated. MicroRNAs (miRNAs) serve as key post-transcriptional regulators of gene expression and are involved in various biological processes. Here, we hypothesized that variants located in miRNAs and their binding sites could constitute a part of the functional variants associated with PD. First, we examined the association of miRNA-related polymorphisms with PD using data from the thus far largest GWAS on PD. We identified two single-nucleotide polymorphisms (SNPs) in miRNAs that are associated with PD, rs897984 in miR-4519 (*p*-value=1. 34×10-5 and OR=0. 93) and rs11651671 in miR-548at-5p (*p*-value=1. 06×10-6 and OR=1. 09). Second, in silico analysis showed that these variants may affect the processing of the pre-miRNAs. In full agreement, in vitro assays demonstrated reduced levels of the mature miRNAs from the mutant alleles. Third, we determined a number of miR-4519 and miR-548at-5p target genes that may mediate their effects on PD. In addition, we found 34 miRNA binding site-SNPs in the 3'UTR of 14 genes that are associated with PD and that have potential to affect miRNA-mediated regulation of their host genes. This includes rs356165 within the miR-658 target site that may disrupt miRNA-dependent regulation of *SNCA*. By implementing this approach, we further identified four new genes that are potentially associated with PD, including *CTSB*, *HSD3B7*, *IGSF9B* and *STX1B*. Taken together, these findings may improve our understanding of the role of miRNAs in the pathophysiology of PD.

3052T

Non-coding RNA dysregulation in schizophrenia patientsobserved in the amygdala region based on RNA-sequencing. *Y. Liu, X. Chang, P. Sleiman, H. Hakonarson.* Center for Applied Genomics CAG, The children's hospital of philadelphia CHOP, Philadelphia, PA.

Schizophrenia (SCZ) is a complex neuronal disease that involves complicated gene networks. The redundancy of the gene networks indicates that many gene combinations have the potentials to cause system dysfunction that manifest as related neuro-behavioral syndromes. Recent studies show that small non-coding microRNA (miRNA) and long non-coding RNA (lncRNA) are important factors shaping those networks that are dynamically regulated by neuronal activation. We selected amygdala brain tissues from 46 human individuals, including 22 SCZ patients and 24 healthy controls, performing RNA sequencing (RNA-seq) for lncRNA detections. 27 individuals (case=13, control=14) were further processed with small RNA-seq platform for microRNA detections. Many SCZ microRNAs have been found to be differentially expressed, including miR-132, miR-212 and miR-34a/miR-34c. A new miR-1307, which is the hot point of SCZ GWAS loci, has been observed to be highly differential expressed in cases and controls. Here we show that two lncRNAs, RP11-677M14. 2 and RP11-724N1. 1, both of which have been associated with SCZ GWAS SNPs and overlapping with SCZ related genes, are also differentially expressed in SCZ compared to control groups.

3053F

MicroRNAs signature in PBMCs from first-treated tuberculosis patients in western China and their potential diagnostic role. *M. Shang, X. Hu, Y. Ye, Y. Zhou, J. Zhou, X. Song, X. Lu, M. Kang, Y. Xie, D. Li, C. Tao, L. Wang, B. Ying.* Department of Laboratory Medicine, West China Hospital, Sichuan University, Chengdu, Sichuan, China.

Background: MicroRNAs (miRNAs) are a class of small non-coding RNAs regulated gene expression at the post-transcriptional level. Many studies have investigated role of miRNAs in the biological processes such as cellular differentiation, development, immunity and apoptosis. Methods: To evaluate role of miRNAs in immune responses during tuberculosis infection, we performedmiRNAprofiling inperipheral blood mononuclear cells isolated from first-treated tuberculosis patients and healthy controls via microarray analysis. Then the remarkably differential expressed miRNAs were chosen for further qRT-PCR validation. The receiver operating characteristic (ROC) curve and 95 % confidence intervals (CI) were calculated to evaluate the diagnostic value of the tuberculosis specific miRNAs. Results: 9 differential expressed miRNAs were chosen for qRT-PCR validation. Among these 9 miRNAs, 3 miRNAs (miR-218, miR-29a, and miR-29c) were down-regulated in the patients. In ROC plots, the area under the curve was 0. 756(95%CI: 0. 672, 0. 841) for miR-218, 0. 705(95%CI: 0. 610. 0. 799) for miR-29a and 0. 657(95%CI: 0. 558, 0. 757) for miR-29c. Conclusion: Our study suggests that a combination of 3 PBMCs miRNAs have great potential to serve as non-invasive biomarkers of tuberculosis.

3054T

Elucidating the role of miR-29a in fibroproliferative diseases. *T. J. Creamer[1,3], A. Lal[2], D. S. Warren[1].* 1) Surgery, Johns Hopkins University, Baltimore, MD; 2) Center for Cancer Research, National Cancer Institute, Bethesda, MD; 3) Institute of Genetic Medicine, Johns Hopkins University, Baltimore, MD.

MicroRNAs (miRNAs) are important regulators of gene expression that have the capacity to potently influence phenotypic outputs. Accordingly, therapeutic approaches that seek to normalize miRNA expression and function are being actively pursued for a number of human diseases. Fibrosis describes an abnormal accumulation of extracellular matrix (ECM) and is a central pathological component of diseases that affect the heart, liver, lungs and kidneys. Despite significant progress in our understanding of fibroproliferative pathways, organ fibrosis continues to account for a significant fraction of the morbidity and mortality in the developed world with few, if any, effective treatments. Members of the miR-29 family (miR-29a/b/c) are down-regulated in all fibrotic tissues examined to date and have been shown to inhibit the expression of numerous ECM proteins including several collagens, elastin and fibrillin. Consistent with a role for miR-29 as a critical regulator of organ fibrosis, we recently demonstrated that systemic administration of a miR-29 expressing adeno-associated virus (AAV) can prevent and even reverse hepatic fibrosis in vivo. In these studies, AAV transgene expression was limited to hepatocytes with no detectable expression in ECM producing hepatic stellate cells. These observations suggest that the therapeutic benefits of enforced miR-29 expression in fibrotic organs involve more than direct repression of ECM protein synthesis. Therefore, to better understand the miR-29 mediated changes in gene expression that occur within specific cellular compartments of fibrotic organs, we used an unbiased approach to identify miR-29 targets in both HuH-7 (hepatocyte) and LX-2 (hepatic stellate) cell lines. Following the method described by Lal et al. (2011), transfection and subsequent pull-down of a biotinylated miR-29a mimic identified 1644 and 2166 potential mRNA targets in HuH-7 and LX-2, respectively. The lists of enriched transcripts include both established and novel miR-29 targets that function across a variety of cellular pathways including apoptosis, histone modification, and signal transduction. In parallel, we also determined the changes in gene expression in HuH-7 and LX-2 cells exposed to the pro-fibrotic cytokine TGF-b. Together our data provide new insight into the normal and disease associated roles of miR-29 in both epithelial and mesenchymal cellular compartments and inform ongoing efforts to developed effective anti-fibrotic therapies.

3055F

Difference in plasma circulating free microRNAs between young and aged mice: A possible contribution of miR-1 on myogenic differentiation under a low temperature condition. *M. Fukuoka, H. Hohjoh.* Department of Molecular Pharmacology, National Institute of Nuroscience, National Center of Neurology and Phychiatry, Kodaira, Tokyo, Japan.

Aging is characterized by a progressive loss of physiological integrity and leads to functional declines in physiology such as a decrease in hearing, visual acuity, muscle strength and memory skills. To see such an aging at a molecular level, we investigated plasma-circulating free microRNAs (miRNAs) between young (6-week old) and aged (2-year old) C57BL6/J mice by means of an expression profile analysis in this study. MiRNAs, which are 21~23-nucleotide-long small non-coding RNAs, function as mediators in gene silencing, and play an essential role in gene regulation by inhibiting translation of messenger RNAs (mRNAs) and by digestion of mRNAs. Recently miRNAs have been found as well in various body fluids including plasma/serum other than inside cells, and such a part of miRNAs are enclosed in extracellular vesicles such as exosomes or associated with RNA-binding proteins, e. g. , Arugonaute2. Such cell-free (extracellular) miRNAs might be implicated in a cell-cell communication and capable of becoming useful biomarkers. From our current expression profiles of miRNAs that were prepared from plasma samples of young and aged mice, the level of miR-1 and miR-133, which are known to be myogenic miRNAs, appeared to be higher in the plasma of young mice than in that of aged mice. In addition, it is noteworthy that young plasma had the potential for inducing myogenic differentiation of mouse C2C12 cells (a myoblast cell line) even under a lower temperature condition (at 30°C). We further investigated the effects of miR-1 and miR-133 on myogenic differentiation of C2C12 cells under a low temperature condition, by examining the expression of *Myogenin* that is a key indicator gene for myogenic differentiation. The results indicated that miR-1 might have the potential for inducing the expression of *Myogenin* even under a low temperature condition. Accordingly, our findings suggested that cell-free miR-1 might contribute to peripheral myogenic differentiation (e. g. , limb muscles) which may occur at a low temperature relative to a core body temperature, and that the reduction of cell-free miR-1 in the plasma of aged mice might be implicated in a decreased muscle mass in aged mice.

3056T

Genomic variation of mouse microRNAs. *I. Hovatta[1,2,3], K. Trontti[1], J. Väänänen[1], K. Icay[3], T. Sipilä[3], D. Greco[4].* 1) Department of Biosciences, University of Helsinki, Helsinki, Finland; 2) Department of Health, National Institute for Health and Welfare, Helsinki, Finland; 3) Research Programs Unit, Molecular Neurology, Biomedicum Helsinki, University of Helsinki, Finland; 4) Unit of Systems Toxicology, Finnish Institute of Occupational Health, Helsinki, Finland.

MicroRNAs (miRNAs) are small non-coding RNAs that function in the post-transcriptional regulation of gene expression. A single miRNA can target hundreds of genes, often within the same biological pathway. Genetic variation within miRNA genes and their putative regulatory regions may affect either the expression level of miRNAs or target gene recognition, thus resulting in phenotypic differences. We investigated systematically genetic variation within the miRNA genes in the mouse genome. We used publicly available dataset of 28 inbred mouse strains, including the most common laboratory and wild-derived strains (http://www. sanger. ac. uk/resources/mouse/genomes/) that are used as models of various human diseases. We performed a genome-wide analysis of SNPs and structural variation within miRNA genes. We observed, as expected, genomic variation occurring less frequently within the miRNA loci compared to the rest of the genome. Furthemore, the seed region, important in the target recognition, harbored less polymorphisms than the rest of the mature miRNA sequence, the pre-miRNA sequence, and the putative promoter region. We next carried out miRNAseq in the hippocampus and frontal cortex tissue of six inbred mouse strains (129S1/SvImJ, A/J, C3H/HeJ, C57BL/6J, DBA/2J, and FVB/NJ). By comparing DNA variation data with miRNAseq data we found consistent RNA editing of three miRNAs. To investigate how RNA editing or SNPs in the miRNA seed region (N=29 seed region SNPs among the six inbred strains) affect their putative mRNA targets, we carried out bioinformatic target predictions using several algorithms with both alternative alleles. As expected, the target sets of the two alternative alleles had minimal overlap. Consequently, the biological processes and pathways potentially regulated by the two miRNA alleles were different asrevealed by theIngenuity Pathways Analysis or DAVID. In conclusion, although mouse miRNA genes are highly conserved, we found a number of seed region SNPs among inbred mouse strainsthat putatively affect which mRNAs the polymorphic miRNAs target. We are currently testing this hypothesis with brain miRNAseq data and in vitro models. This variation likely affect expression levels of numerous genes contributing to the phenotypic differences between the mouse strains. Therefore, inbred mouse strains provide a genetic model system for investigating the effect of genomic variation within the miRNA genes on various phenotypes.

3057F

Mitochondrial miRNA profile in mtDNA-less cells. *R. Li[1,4], Y. Peng[1], N. Dasgupta[1], Z. Tan[1], G. Ciraolo[2], D. Daren Wang[3].* 1) Human Genetics, Cincinnati Children's Hosp, Cincinnati, OH; 2) Division of Pathology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Avenue, Cincinnati, OH 45229; 3) Maxillofacial Pathology and Radiology Department, Ohio State University College of Dentistry,304 W. 12th Avenue, Columbus, OH 43210; 4) Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45229.

In this report, we investigate the novel regulation mechanism in mtDNA-less 206 -o cells. Very low mtDNA copy in 206 -o cells was identified. But no 13 mitochondria-specific proteins were translated in 206 -o cells. Their mitochondrial respiration complex V, III, II were 86. 5%, 29. 4% and 49. 6% of 143B cells respectively. Complex I and IV were completely loss in 206 -o cells. 206 -o cells depend on non-mitochondrial respiration to generate ATP. Some mitochondrial RNAs including 12S rRNA, COX1, COX2, COX3, ND4 and ND5 were low expression. However, ND1, ND3 and Cyto b were not expressed in 206 -o cells. Unequal transcription of mitochondrial RNAs indicated the posttranscriptional cleavage and processing mechanisms in the regulation of mitochondrial gene expression in 206 -o cells. miRNAs may modulate these mitochondrial RNA expression in these cells. RNA-induced silencing complex (RISC) within 206 -o cell mitochondria indicated miRNAs in 206 -o cell mitochondria. miRNA profile in mtDNA-less 206 -o cells was studied by Next-generation sequencing of small RNAs (sRNAs). Several mitochondria-enriched miRNAs such as miR-181c-5p and miR-146a-5p were identified in 206 -o cell mitochondria. miR-181c-5p and miR-146a-5p had 23 and 19 potential targets on mitochondrial RNAs respectively, and these two miRNAs had multiple targets on mitochondria-associated mRNAs encoded by nuclear genes. These data provided the first direct evidence that that miRNAs were imported into mitochondria and regulated mitochondrial RNAs. Thus, our findings may provide new insights for study regulation mechanism of miRNAs on mitochondrial RNAs, particularly in mtDNA-less cells.

3058T

DASHR - Database of small human non-coding RNA. *Y. Y. Leung[1,2], P. P. Kuksa[1,2], A. Amlie-Wolf[3], O. Valladares[1], B. D. Gregory[3,4], L. -S. Wang[1,2,3].* 1) Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Penn Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 3) Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Biology, University of Pennsylvania, Philadelphia, PA.

Small non-coding RNAs (sncRNAs, <200nts) are key regulators of diverse cellular processes. Although thousands of them are known to be present in the human genome, studies continue to uncover new sncRNAs and many of them still have unknown functions. Recent advances in small RNA-sequencing (smRNA-seq) technology and computational tools allow for the unprecedented discovery and characterization of these sncRNAs in a genome-wide manner. However, it is still a great challenge to study the important and versatile roles of sncRNAs due to a lack of completeannotation and resources. Therefore it is invaluable to have a comprehensive human small non-coding RNA database containing the functions, structural features, and expression profile of these complex sncRNAs genome-wide across various human tissues and cell types. We present and analyze DASHR, the most complete and integrated human sncRNA database to date. We first built our own sncRNA annotation on eight different types of known processed sncRNAs (46,306 in total). We then collected smRNA-seq data from over 30 sources, representing 42 human tissues and cell types from 198 individuals. Using a novel smRNA-seq specific pipeline, we uniformly processed all these data to delineate the expression profiles of all sncRNAs. We catalogued expression profiles of the known sncRNAs and captured 70% of mature and primary miRNAs, 35% of piwi-RNAs, 50% of small cytoplasmic RNAs, 87% of small nucleolar RNAs, 21% of small nuclear RNAs, 86% of transfer RNAs derived sncRNAs, and 45% of transfer RNA fragments. We also found 2,912 novel sncRNAs not present in GENCODE or UCSC. 6%, 20%, 37% and 21% of these novel processed sncRNAs resided within promoters, exonic, intronic and intergenic regions respectively. DASHR represents a searchable comprehensive landscape of processed normal human sncRNAs. In addition to the annotation and sequence information of sncRNAs, we also included their expression profiles across all tissues, secondary structures, as well as statistics calculated to infer the specificity of sncRNA processing. This database is thus a unique resource on individual sncRNAs serving as a starting point for any large scale studies. The DASHR database provides a systematic characterization of human normal sncRNA biology, and informs future disease-related studies seeking to characterize the roles of tissue-specific sncRNAs and non-coding variants. All data is available for download and query atwanglab. pcbi. upenn. edu/DASHR. .

3059F

Non-coding transcriptome of the developing mammalian retinal photoreceptors. *L. Zelinger, G. Karakulah, J. W. Kim, H. J. Yang, M. Brooks, V. Chaitankar, A. Swaroop.* Neurobiology, Neurodegeneration & Repair Laborator, NIH-NEI, Bethesda, MD.

Purpose:Only about 2% of the genome encodes for protein-coding transcripts, yet over 80% is actively transcribed. The non-coding transcriptome includes several professional RNA species with wide range of cellular functions. The goals of this study are to identify key long non-coding (lnc) RNAs that are functionally relevant in retinal photoreceptors, the cells responsible for initiating the visual process. Methods:-Photoreceptors were purified from retina of *Nrlp*-GFP and *Nrl*-knock out (KO) *Nrlp*-GFP mice by fluorescence activated cell sorting. We carried out transcriptome analysis of developing and mature photoreceptors using RNA-seq. Data was mapped to GRCm38 and analyzed using the ENSMBL annotation. ChIP-seq experiments were performed using two key photoreceptor transcription factors (TF) NRL and CRX. Weighted correlation network analysis (WGCNA) was performed to predict participation in functioning pathways. High-resolution in-situ imaging was done using the RNAscope kit. Results:We identified 12,448 and 13,333 non-coding transcripts in rod and cone-like photoreceptors, respectively. About 3100 transcripts areprofessional non-coding RNA species, and approximately 40% of these are annotated as lncRNAs, the most abounded group in our data set. In order to identify the key lncRNA associated with photoreceptor function, we used NRL and CRX ChIP-seq data together with differential expression analysis. By implementing this approach, we identified 13 lncRNAs, which are potentially regulated by both NRL and CRX, 13 by CRX alone, and another 14 by NRL. To validate and further investigate these transcripts, we selected 12 lncRNAs for in-situ hybridization experiments. While all lncRNAs showed expression in the photoreceptor layer of the retina, 5 showed especially interesting subcellular localization, (such as near the connecting cilium, or perinuclear). Bioinformatics analysis revealed potential interaction of these lncRNAs with proteins and their participation in functional pathways that included, vision and light perception, chromatin modification and maintenance and establishment of RNA localization. Conclusion:Using a combination of experimental analysis and bioinformatic tools, we have discovered a subset of lncRNAs that appear to play key roles in retinal photoreceptors. Further studies are currently underway to identify the human orthologs of photoreceptor-specific lncRNAs and their relevance, if any, to disease.

3060T

Uncovering the mechanisms and the genetics of enhancer transcription activities via population-scale sequencing of nascent RNAs. *H. Kwak, H. M. Kang.* Department of Biostatistics, University of Michigan, Ann Arbor, Ann Arbor, MI.

Histone modification marks, chromatin accessibility, protein binding sites, and methylation profiles expand our understanding of genetic and epigenetic regulation of gene expression. Recently, nascent RNA sequencing implemented by Precision nuclear Run-On sequencing (PRO-seq/PRO-cap) revealed a precise map of enhancer transcription activities, suggesting strong relationship between promoters and enhancers. However, it is still unclear what the impact of genetic variation is on the transcription activity variations and subsequently on the gene expression levels. In addition, it is unclear how to interpret the raw data to precisely infer transcript start sites (TSSs). For example, recently suggested hypothesis of convergent transcription dramatically changes the algorithm to infer TSSs from PRO-cap sequence reads. Here we present a preliminary analysis of the genetic landscape of transcription activity, by sequencing nascent RNAs of lymphoblast cell lines of ~70 individuals with African ancestry. The main goal for our study is to understand the genetic variation of active enhancer transcription, and their relationship with gene expression levels. To infer active transcript start sites precisely without assuming either divergent or convergent transcription models, we developed a novel algorithm that hierarchically clusters the peaks of PRO-cap sequence reads and deconvolutes pairs of sense and antisense peaks using a bipartite graph matching algorithm. Our method allows to infer convergent and divergent transcription activities in an unbiased manner. By the time of the abstract presentation, we plan to perform 'tssQTL' analysis using the inferred interval of transcription site sites in population scale. If there are individual differences in the predicted transcription binding intervals caused by genetic variation, we will be able to identify association between genetic variations and the presence and absence of predicted transcription binding intervals. We will perform genome-wide tssQTL analysis and compare the difference between tssQTL and the landscape of dsQTLs and eQTLs.

3061F

The identification of Carboxy-terminal frameshift mutations and their role in human disease. *J. White[1], Z. Akdemir[1], S. Jhangiani[2], T. Gambin[3], E. Boerwinkle[2,4], R. A. Gibbs[1,2], J. R. Lupski[1,2,5,6], C. M. B. Carvalho[1], Baylor-Hopkins Center for Mendelian Genomics.* 1) Molecular & Human Genetics, Baylor College of Medicine, Houston, TX; 2) Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX; 3) Institute of Computer Science, Warsaw University of Technology, Warsaw, Poland; 4) Human Genetics Center, University of Texas Health Science Center at Houston, TX; 5) Texas Children's Hospital, Houston, TX; 6) Department of Pediatrics, Baylor College of Medicine, Houston, Texas.

Nonsense-mediated decay (NMD) is an RNA quality control process by which mutant transcripts containing a premature termination codon (PTC) are actively degraded. However, NMD is sometimes unable to recognize a PTC located at the portion of the transcript encoding the C-terminus of the protein, allowing for the production of a mutant protein product with a large C-terminal aberration. The notion that a proportion of variants escape NMD is important clinically, as some genes only display pathogenesic phenotypes due to a dominant-negative or gain-of-function mechanism and have been demonstrated to alter disease inheritance patterns (e. g. recessive loss-of-function disease resulting from a dominant mutation) and explain allelic affinity in genes with a known disease association. To investigate the frequency of these C-terminal mutational events we examined our in-house cohort of ~5000 exomes. We selected all frameshifting variants located in the final exon plus 200bp upstream in the penultimate exon according to the RefSeq transcripts. All candidates were filtered against exome data in publicly available databases, including the NHBLI Exome Sequencing Project and the Atherosclerosis Risk in Communities Study database. Of the resulting ~2000 unique variants we manually curated those predicted to escape NMD. Our list of variants includes genes where C-terminal truncation has previously demonstrated an altered pathogenesis and the disease state is not a result of haploinsufficiency including *AHDC1, FBN1, DMD* and *DVL1.* We have recently reported the variant alleles in *DVL1* escape NMD and all share an identical PTC within the last exon that creates a mutant DVL1 protein with a large atypical C-terminal peptide tail. This rare mutational mechanism was reported as a common cause of Robinow syndrome likely due to a gain-of-function and/or dominant-negative mechanism (PMID:25817016). Similar to the variants identified in *DVL1*, we have identified greater than 100 candidate variants in multiple affected patients with unsolved exomes that present with overlapping phenotypes, including cohorts of osteoporosis, Robinow syndrome, intellectual disability and multiple eye diseases. Our data show that variants that do not lead to haploinsufficiency contribute to human disease and should be carefully analyzed. We suggest that gain-of-function and dominant-negative mutations are highly under-recognized in genomic analyses and likely result in a wide variety of human disease.

3062T

Spatial and temporal dynamics of epitranscriptome in normal neurodevelopment and fragile X syndrome. *F. Zhang[1], Y. Kang[1], Y. Yue[2], C. He[2], P. Jin[1].* 1) Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322; 2) Department of Chemistry, University of Chicago, Chicago, Illinois 60637.

*N*6-methyladenosine (m6A) is the most prevalent internal modification of mammalian messenger RNA (mRNA). This reversible modification can destabilize mRNA, increase translation efficiency, and regulate RNA-protein interactions. However, the role(s) of m6A in neurodevelopment remains unclear. We have found that m6A/A level changed significantly during neurodevelopment, and genes involved in m6A dynamics were differentially expressed during neuronal differentiation, insinuating a potential role for m6A in neurogenesis. To evaluate the spatial and temporal dynamics of m6A during neurodevelopment, we profiled genome-wide m6A sites of four brain regions (cerebellum, cerebrum cortex, hippocampus, and hypothalamus) during neurodevelopment in mouse using anti-m6A immunoprecipitation (MeRIP) followed by high-throughput sequencing. In the combination of four brainregions, 3081 genes (13% of total expressed genes) contain high-confidence (HC) m6A site(s) in their transcripts. In all four regions, the majority of HC m6A sites were near stop codonor in 3'-UTRs, andthemotif[GA]GAC was highly enriched within these sites. Nevertheless, each brain region contained an appreciable number of region-specific HC m6A sites. In cortex,86 transcripts with region-specific HC m6A sites showed 10-fold enrichment for postsynaptic genes ingene ontology analyses(Fisher's Exact test, p=3. 1E-5). Interestingly,one variantof themotif, AGACUG, was prominent only in cerebellum. Transcripts containing this variant displayed a strong correlation with macromolecule metabolic process (p=1. 1E-6). Intriguingly, overlapping analyses revealed that 40% of 842 mRNAs bound by Fragile X Mental Retardation Protein (FMRP) contained HC m6A sites in at least one brain region (3. 0-fold enrichment, p=4. 5E-88), many of which are involved in nervous system development (p=9. 3E-27). Furthermore, transcripts of 27% of 631 autism-related genes (such as NLGN1, NRXN1, BDNF, and UBE3A) contain HC m6A sites in at least one brain region (2. 0-fold enrichment, p=5. 4E-22). Biochemically we have found that FMRP is associated with m6A-binding proteins. Furthermore, using a fly model of fragile X syndrome, we have observed the genetic interaction between *FMR1* and the genes involved inthe m6A pathway. These results together suggest that m6A modification is widespread in the brain transcriptomes, displays spatial and temporal dynamics, and could contribute to the molecular pathogenesis of fragile X syndrome.

3063F

Identification and characterization of genetic determinants of RNA editing in humans. *Y. Z. Kurmangaliyev, S. V. Nuzhdin.* University of Southern California, Los Angeles, CA.

ADAR-mediated adenosine to inosine deamination (A-to-I) is the most common type of RNA editing in metazoans. Inosines form Watson-Crick base pairs with cytosines and can mimic guanosines. Such recoding events may result in alterations of encoded proteins. RNA editing usually affects only a portion of all expressed transcripts and the degree of editing may vary strongly among particular sites. The proper editing profiles are required for function of many human genes, and aberrant editing may result in deleterious phenotypes. Surprisingly, the degree of RNA editing of particular A-to-I sites has been never associated with underlying genetic determinants. In a recent study, we uncovered a novel type of functional genetic variation affecting the editing level of particular A-to-I sites in *Drosophila melanogaster* (Kurmangaliyev et al. , manuscript submitted). We were able to identify putative regulatory variants associated with changes in RNA editing levels (RNA-editing quantitative trait loci or edQTLs).

Here we studied genetic variation in editing of ADAR targets in transcriptomes of 462 individuals generated by the GEUVADIS project. The association analysis allowed us to detect dozens of *cis*-edQTLs associated with altered editing of human A-to-I sites. The analysis was performed separately in the European and African populations, and many of identified associations were replicated in both populations. The overall features of human edQTLs were consistent with observations in Drosophila. Associated variants were enriched near the regulated editing sites. Some of them were associated with editing of several colocalized sites. This is a first study of genetic variation in RNA editing in humans. We characterize a novel type of functional genetic variation affecting human transcriptomes. Similar to studies on expression and splicing QTLs, characterization of edQTLs will greatly expand our understanding of gene regulation. Another intriguing question raised by these findings is the contribution of a novel functional variation to human diseases and other phenotypic traits.

3064T

***MTO1* inhibition impairs mitochondrial and cellular function in human cell lines.** *C. Chen, ZW. Gao, Y. Chen, MX. Guan.* Institute of Genetics, Zhejiang University, Hangzhou, Zhejiang, China.

Mitochondria are essential organelles of cells considering their major function in ATP production through the process of oxidative phosphorylation. An efficiently functional mitochondrion requires active interplay between nuclear and mitochondrial genomes. Accumulating evidence suggests various types of posttranscriptional modifications are indispensable for the survival of cell. The human *MTO1* gene encodes a highly conserved protein which plays a key role in mitochondrial translation and further in mitochondrial respiration. Multiple mutations of *MTO1* have been reported to be associated with hypertrophic cardiomyopathy and lactic acidosis. To investigate the molecular mechanisms underlying the mitochondrial disorders associated gene, we established *MTO1* knocking down cell lines using lentiviral vectors conveying sequence-specific short hairpin RNA. A significant decrease of the mitochondrial protein level was observed in the knocking down cell lines when compared to the negative controls, suggesting impaired mitochondrial functions. In addition, we found that there was anapproximate20% decrease in mitochondrial ATP production relative to total ATP production, and the JC-10 assay revealed a marked decline in mitochondrial membrane potential. Furthermore, using a seahorse extracellular flux analyzer, we showed reduced oxygen consumption rates in the knocking down cells. These data highlight the cellular function of *MTO1* and provide new insights into further understanding of pathogenesis of *MTO1* related diseases.

3065F

Analysis of RNA editing in non-human primates. *M. Bozinoski[1,2], C. E. Mason[2].* 1) Pharmacology Dept , Weill Cornell Graduate School of Medical Sciences, New York, NY; 2) Physiology and Biophysics Dept , Weill Cornell Graduate School of Medical Sciences, New York, NY.

Adenosine deaminases acting on RNA (ADAR) enzyme mediated RNA editing is the prevalent form of post-transcriptional modification in higher organisms. Deamination of Adenosine creates Inosine, which is recognized as Guanosine by both ribosome and spliceosome thereby causing changes in the coding sequence, splicing and miRNA binding. RNA editing is mostly primate specific phenomenon and few studies of limited scope have shown that humans possess more editing sites than other primates. Here, we present the topology and extent of RNA editing throughout the transcriptome of 12 species of non-human primates from our Non-Human Primate Reference Transcriptome Resource (NHPRTR), including great apes, old world monkeys, new world monkeys and prosimians, spanning approximately 70 million years of primate evolution. Using a robust pipeline to resolve the bona fide RNA editing sites, we analyzed a dataset consisting of RNA-Seq data from 21 polled tissues for each of the 12 primate species/subspecies, supplanted by data from genomic DNA sequencing for all of the animals used to obtain theRNA-Seq data. Our results show that primates show less RNA editing than humans, with baboon, gorilla and chimp showing 19. 7% 13. 4% and 12. 1% less editing sites respectfully, perfectly fitting within the phylogenetic tree. Since RNA editing is a tissue-specific phenomenon, we extended our analysis of RNA editing in tissue specific manner, using data from 14 separate sequenced tissues from 10 primate species. By expanding the analysis in tissue-specific manner, we identify most of the RNA editing events to be within the central nervous system. Additionally, we observe distinct conservation of RNA editing sites throughout the different brain regions, with primate species phylogenetically closest to human showing more human-like distributions. Assessment of the topology of identified RNA editing sites in non-human primates, revealed a surprising portion of the edited sites (up to 20%) being in non-coding transcripts, thereby implicating a broad regulatory role for RNA editing. Further analysis of the editing sites in the coding regions of the transcripts show that edited codons are much more likely to be non-synonymous than synonymous (4:1 ratio), further more stressing the importance of RNA editing as a transcriptome (and, subsequently, proteome) diversification mechanism.

3066T

Alternative promoter usage in *STK39* leads to distinct transcripts with human-specific 5'UTRs but identical protein isoforms as in mice. *Y. Chang, C. Mercado, X. Wang.* School of Medicine, University of Maryland, Baltimore, MD 21201, USA.

STK39 encodes a serine threonine kinase, SPAK, that plays a crucial role in maintaining blood pressure through the strict regulation of salt reabsorption along the distal convoluted tubule (DCT) and the thick ascending limb (TAL). SPAK is part of a network of kinases (including WNKs and OSR1) and bumetenide- and thiazide-sensitive ion cotransporters (NKCC2 and NCC, respectively). Environmental stressors such as hypovolemia activate WNK kinases to phosphorylate SPAK and OSR1, which in turn phosphorylate NKCC2 and NCC to allow for the influx of extracellular Na+ and Cl- into renal epithelial cells. In mice, *STK39* undergoes alternative promoter usage and splicing, which results in numerous functionally distinct protein isoforms that have been well characterized. These include the full-length protein (63 kDa) and two shorter isoforms that are both kidney specific; a 50 kDa isoform known as SPAK2 (start codon in exon 2) that lacks part of the kinase domain, and a 38 kDa isoform known as KS-SPAK (start codon in exon 6) that lacks the entire kinase domain. Both isoforms are more abundant within the mouse TAL, whereas full-length SPAK is highly expressed within the DCT, and they both function to inhibit phosphorylation of NCC and NKCC2 by full-length SPAK. However, the existence of orthologous SPAK isoforms (SPAK2 and KS-SPAK) in human kidney, and the role of SPAK in nephron-segment-specific regulation of NCC and NKCC2, still have not been determined. In this study, we performed RNA-seq and 5'RACE to analyze the human and mouse kidney transcriptomes. This approach has led to the discovery of several human-specific *STK39* alternative transcripts, including those with novel alternative start sites and splicing of exon 13, all of which have been confirmed by RT-PCR. Interestingly, while the human and mouse transcript homologs have distinct transcription start sites and 5'UTRs, the resulting human protein isoforms are nearly identical to what was found in mice. In contrast to mouse SPAK homologs, the respective human SPAK isoforms are not kidney specific, yet they exhibit differential expression across tissues. Examining promoter signatures and functional significance of these isoforms will provide a deeper understanding of how *STK39* may have evolved differently between mice and humans, and how human SPAK isoforms may achieve effective blood pressure control.

3067F

Exploration of RNA editing sites in human retina and retinal pigment epithelium-choroid-sclera. *D. Cho[1], K. Kazmierkiewicz[1,2], M. Li[3], D. Stambolian[1].* 1) Department of Ophthalmology, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; 2) Bioinformatics Core, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA; 3) Department of Biostatistics and Epidemiology, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.

RNA editing is the post-transcriptional modification that involves the alteration of a nucleotide. The most common form of RNA editing is adenosine-to-inosine (A-to-I) where the inosine is translated as guanine by ribosomes. Recent studies have shown that diseases are correlated with A-to-I RNA editing such as glaucoma, diabetes, psychiatric disorders, and neurological disorders. In this study, RNAseq was performed on eight human post-mortem eyes to look for known and novel A-to-I RNA editing sites. REDItools was used to computationally predict significant editing sites using Fisher Exact Test and FDR correction ($p < 0.05$). Known single nucleotide polymorphisms from dbSNP, 1000Genomes, and SeattleSeq were removed to lower the possibility of finding a false positive editing site. After a preliminary list of predicted A-to-I editing sites was produced, BLAT was used to confirm the original GSNAP alignment that further filtered the predicted editing sites. After processing, the RNA editing sites were annotated and quantified between locations (macular versus peripheral) and tissues (retina versus RPE-choroid-sclera). 1,327 editing sites were found in the macular retina and 682 editing sites were found in the macular RPE-choroid-sclera while 1,174 editing sites were found in the peripheral retina and 885 editing sites were found in the peripheral RPE-choroid-sclera. For each tissue-location, 6-15% of the predicted editing sites were known from the DARNED database and 20-31% of the editing sites were located within protein coding genes. At least one editing site was found within genes associated with age-related macular degeneration, refractive error, and primary open angle glaucoma. From the analysis, our pipeline was successful in confirming previously known A-I editing sites while discovering possible novel A-I editing sites. Validation of these novel RNA editing sites will also be presented.

3068T

Global binding map of the splicing regulatory factor SRSF5. *G. H. Bruun[1], T. K. Doktor[1], A. Masuda[2], A. R. Krainer[3], K. Ohno[2], B. S. Andresen[1].* 1) Biochemistry and Molecular Biology (BMB), University of Southern Denmark, Odense, Denmark; 2) Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Japan; 3) Cold Spring Harbor Laboratory, PO Box 100, Cold Spring Harbor, NY 11724, USA.

mRNA splicing is a fundamental process required for correct gene expression of most protein coding genes. One prominent group of splicing regulatory proteins, the SR proteins, binds splicing regulatory elements (SREs) in a sequence-specific way to enhance exon inclusion. Thus, SNPs and mutations that affect SREs may disrupt splicing, and an increasing number of human diseases are now known to be caused by aberrant splicing. However, the specific binding motifs are not precisely defined for most splicing regulatory proteins, and maps of their location, as part of SREs in human disease genes, are missing or sparse. To unravel the regulatory landscape of splicing and to predict potential pathogenic consequences of disease-associated sequence variations it is important to precisely identify the binding sites of splicing regulatory factors. We used iCLIP (Individual nucleotide-resolution Crosslinking and Immunoprecipitation) to identify the transcriptome-wide binding sites and derive *in vivo* binding motifs of the SR protein SRSF5. We used HeLa cells expressing T7-tagged SRSF5 and performed UV-crosslinking to induce covalent bonds between SRSF5 and RNA. After immunoprecipitation and removal of the crosslinked protein by proteinase treatment, the RNA was eventually used as a template for cDNA synthesis and PCR. The amplified products were then subjected to next-generation sequencing and analyzed. We identify approximately 20,000 SRSF5 binding sites across the transcriptome, and their distribution correlates with the predicted roles of SRSF5 in mRNA splicing. As an example, we find that a previously reported pathogenic mutation, shown to cause skipping of *IVD* exon 2, disrupts SRSF5 binding, suggesting that this may be the cause of exon skipping. To further elucidate the roles of SRSF5 in mRNA splicing, we performed siRNA-mediated knockdown in HeLa cells followed by RNA-seq analysis. This suggested that SRSF5 regulates splicing of other splicing regulatory proteins, such as *MBNL1* exon 5 and *PTBP2* exon 10, and thus that SRSF5 may both directly and indirectly affect mRNA splicing. Generation of the *in vivo* SRSF5 binding map will enable identification of *in vivo* targets, and in combination with the *in vivo* binding motifs, this will help pinpoint genetic variants that may cause aberrant splicing.

3069F

Sex Chromosome Dosage and the Human Transcriptome: A Study of XO, XX, XXX, XY, XXY, XYY and XXYY Karyotypes. *A. Raznahan[1], N. Parikshak[2], V. Chandran[2], F. Gao[2], J. Blumenthal[1], L. Clasen[1], A. Alexander-Bloch[1], A. Zinn[3], D. Wangsa[4], D. Murphy[5], P. Bolton[5], T. Ried[4], G. Coppola[2], J. Ross[7], J. N Giedd[6], D. H. Geschwind[2].* 1) Developmental Neurogenomics Unit, Child Psychiatry Branch, NIMH Intramural Research Program, NIH, Bethesda, MD; 2) Neurogenetics Program, Department of Neurology and Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA; 3) McDermott Center for Human Growth and Development and Department of Internal Medicine, University of Texas Southwestern Medical School, TX, USA; 4) Genetics Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD, USA; 5) Institute of Psychiatry, Psychology and Neuroscience, King's College London, University of London, UK; 6) Department of Psychiatry, UC San Diego, La Jolla, CA, USA; 7) Department of Pediatrics, Thomas Jefferson University, Philadelphia, PA, USA.

A key question in the basic biology of sex-differences has received little direct empirical study in humans: how does sex chromosome dosage (SCD) influence the expression of sex chromosome and autosomal genes? Here, we address this question using oligonucleotide microarrays to profile gene-expression in lymphoblastoid cell lines from 68 participants across 7 SCD groups: XO, XX, XXX, XY, XXY, XYY, and XXYY. First, we show that agnostic clustering of X- and Y-chromosome genes based on their expression change across groups recovers four gene clusters with differing dosage sensitivity: pseudoautosomal region (PAR) genes, Y-linked genes, X-linked genes that escape X-inactivation (XCIE) and X-linked genes that undergo full inactivation (XCI). However, contrary to expectations, expression of the XCI cluster reduces with increasing X-chromosome count. We validate these findings by bioinformatic comparison with prior evolutionary and epigenetic data, and qtPCR measures of gene expression in an independent sample of ~400 individuals. Second, by surveying expression changes across the genome we show that the total number of differentially-expressed genes is greatest with the lack of an X-chromosome in females (>2000), intermediate with additional copies of an X in males (>200), and smallest with additional copies of an X in females or a Y in males (>20) – reflecting that fact that gonadal sex and/or Y-chromosome status can modify the impact of altered X-chromosome dosage on autosomal gene expression. However, lack of an X in females, and presence of an additional Y in males, induce highly correlated expression changes in autosomal genes – suggesting that X-Y homologous genes and/or epigenetic mechanisms may mediate SCD effects on autosomal gene expression. Finally, Weighted Gene Co-Expression Network Analyses identify two anti-correlated sets of gene modules that show opposing relationships with X-chromosome dosage:(i) DNA replication and cell-cycle control modules enriched for XCIE genes (greater module expression with greater X-chromosome count), (ii) apoptosis and oxidative phosphorylation modules enriched for XCI genes (reduced expression with increasing X count). Strikingly, these two sets of counterpoised SCD-sensitive gene networks both show binding site enrichment for the dosage sensitive X-linked zinc-finger protein transcription factor gene ZFX. Our findings shed new light on basic sex chromosome biology, and clinical aspects of sex chromosome aneuploidy.

3070T

Transcriptional consequences of 2q23. 1 deletion syndrome in iPSC-derived neural progenitor cells provide insight into neurodevelopmental disorders and autism. S. V. Mullegama1, S. R. Williams2, J. T. Alaimo1, L. Chen1,3, F. J. Probst1, C. Haldeman-Englert4, J. W. Innis5, P. Stankiewicz1, S. W. Cheung1, T. Ezashi6, S. H. Elsea1. 1) Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX, 77030, USA; 2) Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22902, USA; 3) Department of Cellular and Genetic Medicine, School of Basic Medical Sciences, Fudan University, Shanghai 200032, China; 4) Fullerton Genetics Center, Asheville, NC 28803; 5) Departments of Human Genetics and Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI 48109, USA; 6) Division of Animal Sciences, University of Missouri, Columbia, Missouri 65211, USA.

2q23. 1 deletion syndrome (2q23. 1 del) is a neurodevelopmental disorder in which haploinsufficiency of *MBD5* is responsible for the core phenotype, which includes intellectual disability (ID), autism spectrum disorder (ASD), seizures, speech impairment, motor delay, and sleep and behavioral problems. To comprehensively understand the molecular role of *MBD5* in early neuronal development, the 2q23. 1 del phenotype, and ASD, we generated neural progenitor cells (NPCs) derived from induced pluripotent stem cells (iPSCs) that originated from 2q23. 1 del and age/sex-matched control fibroblasts and performed transcriptome profiling by RNA-sequencing. Overall, 300 transcripts were differentially expressed as a consequence of reduced *MBD5* dosage using Benjamini-Hochberg FDR (q<0. 05) analysis. Our transcriptome profile through DAVID was enriched for neuronal development (P=0. 0002), neuronal migration (P=0. 0001), and neurogenesis (P=0. 00001). IPA identified a strong statistical enrichment for pathways, including PTEN (P=0. 0003) and WNT signaling (P= 0. 003), which are critical contributors to neural stem cell and nervous system development. We identified genes within our profile that are implicated in phenotypes present in 2q23. 1 del patients including ID (*FMR1*), seizures (*SEZ6*), speech impairment (*FOXP2*), motor delay (*NR4A2*), and behavioral problems (*ARTNL, HTR1B*). Interestingly, 30% of our differentially expressed genes are associated with ASD as defined by the Simons Foundation and AutismKB and are involved in essential cellular processes crucial in ASD pathogenesis including transcriptional regulation (*MEF2C*, *FOXG1, NFIA*), chromatin modification (*CDH8, SIX1*, *MEOX1*), and neuronal function (*NTNG1*, *GRIK2, LAMC3*). In addition, when *MBD5* is haploinsufficient, genes involved in phenotypically similar ASDs such as 5q14. 3 deletion syndrome (*MEF2C*) and congenital variant of Rett syndrome (*FOXG1*) were dysregulated, suggesting that these genes function in a transcriptional cascade of common biological networks that result in overlapping phenotypes. To identify possible novel deletion disorders, we queried our top 50 MBD5-down regulated genes, which were not linked to ASD or OMIM disorders, through genome-wide aCGH databases and identified 17 candidate genes as possible disease causing genes. Overall, we demonstrate the multifactorial role of MBD5 in the regulation of genes and pathways known to contribute to the genetic architecture of neurodevelopment and disease phenotypes.

3071F

Gene expression analysis of chitinase-like protein, YKL-40, with mammalian chitinases using qPCR in normal human tissues. *M. Ohno[1,2], Y. Kida[1], M. Sakaguchi[1], Y. Sugahara[1], F. Oyama[1].* 1) Department of Chemistry and Life Science, Kogakuin Univ, Hachioji, Tokyo, Japan; 2) Research Fellow of Japan Society for the Promotion of Science (DC2), Chiyoda-ku, Tokyo, Japan.

Humans express two active chitinases, chitotriosidase (Chit1) and acidic mammalian chitinases (AMCase) and also express chitinase-like proteins (CLPs), which are structurally homologous to chitinases but lack the ability to degrade chitin. Humans produce primarily YKL-40 (CHI3L1 or human cartilage glycoprotein-39). YKL-40 levels have been reported elevated in patients with asthma, cystic fibrosis, rheumatoid arthritis and malignant tumors. However, expression levels of YKL-40 have not been compared with those of chitinases and housekeeping genes in normal human tissues. Recently, we established a quantitative real-time PCR (qPCR) system using a single standard DNA molecule to quantify multiple genes. Here, we quantified the YKL-40 mRNA levels and compared them with two chitinases, Chit1 and AMCase, and housekeeping genes in normal human tissues using qPCR system with the standard DNA. We found that YKL-40 mRNA was widely expressed in human tissues. The highest levels of YKL-40 mRNA were detected in human liver, followed by the kidney. In the kidney and liver, YKL-40 levels were more than 100-times higher than Chit1, while being lower than the levels of GAPDH and β-actin. The data presented here can be considered as fundamentals for understanding the biological roles of YKL-40 and mammalian chitinases upon pathological human tissues changes.

3072T

Positional coding and noncoding transcriptomes of synovial fibroblasts in joint specific patterns of arthritis. *G. Russo[1], M. Frank-Bertoncelj[2], M. Trenkmann[2], M. Armaka[3], A. Bratus[1], C. Kolling[4], B. A. Michel[2], R. E. Gay[2], C. D. Buckley[5], G. Kollias[3], S. Gay[2], C. Ospelt[2].* 1) Functional Genomics Center Zurich - ETH/UZH, Zurich, Switzerland; 2) Center of Experimental Rheumatology, University Hospital Zurich, Switzerland; 3) Institute of Immunology, Biomedical Sciences Research Center Alexander Fleming, Vari, Greece; 4) Schulthess Clinic Zurich, Switzerland; 5) Center for Translational Inflammation Research, University of Birmingham, UK.

BackgroundMolecular mechanisms underlying differential topographic susceptibility of the joints to rheumatoid arthritis (RA) are unknown. Positional expression of Hox genes along body axes regulates limb development. Adult skin fibroblasts retain the positional embryonic Hox code and exhibit major differences in transcriptome, defining their unique positional identities. **Objectives**We hypothesize that synovial fibroblasts (SF), which locally drive joint destruction in RA, show positional gene expression patterns predisposing joints to certain pathologies. **Methods**SF were isolated from RA and osteoarthritis metacarpophalangeal (MCP), shoulder and knee joints, healthy human knees, and front paws, ankles and knees of wild type and TNFα transgenic arthritic C57BL/6 mice. RNA was extracted from the human samples (n=21)and sequenced. Hierarchical clustering, pathway analysis and qPCR analysis followed. Positional epigenetic regulation and function of HOTAIR were studied in SF treated with TNFα or silenced for HOTAIR by chromatin immunoprecipitation, qPCR and ELISA. **Results**Clustering of the transcript abundances for human SFreflectedthe location rather than disease groups. Small noncoding RNA conferred the strongest positional identity by clustering 21 out of 21 SF into groups accurately reflecting their joints of origin. Conversely, clustering based on the expression of long noncoding RNA and mRNA suggests that, at least in SF, these transcripts are less joint-specific during the course of the arthritis. Several arthritis-relevant pathways, such as chemotaxis, cell adhesion and the Jak-Stat inflammatory pathway were enriched in a joint-specific manner. In particular, matrix-degrading pathways were enriched in MCP compared to knee SF. Hox-residing RNA exhibited strict positional pattern in human and mouse. TNFα significantly repressed HOTAIR levels in SF. Repression of HOTAIR in anterior and TNF treated SF coincided with enrichment of H3K27me3 repressive histone mark at the HOTAIR promoter. Basal and TNFα-induced production of matrix metalloproteinases 1, 3, 13 and 14 was significantly enhanced in HOTAIR-silenced SF. **Conclusion**Large positional differences in the transcriptome of SF suggest that functionally unique SF subsets populate different joints. The existence of an epigenetically imprinted positional “risk” signatures of SF, like positional HOTAIR expression, may confer certain jointsgreater susceptibility to develop destructive arthritis, such as RA.

3073F

Systematic identification of cancer-testis genes in 20 cancer types: a source of epi-driver genes. *Z. Hu, C. Wang, Y. Gu, K. Zhang, K. Xie, M. Zhu.* Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing, Jiangsu, China.

Cancer-testis (CT) genes represent similarity between the process of spermatogenesis and tumorigenesis, and are conventionally considered as targets of anticancer vaccines. We integrated transcriptomics data from several large databases (e. g. , GTEx, TCGA, HPM, FANTOM, ENCODE etc.) and systematically identified 881 novel CT genes in 20 cancer types and explored their relation with testis-specific regulatory elements. We proposed extremely highly expressed CT genes (EECTGs) as potential epi-drivers and found mutually exclusive associations between EECTGs and somatic mutations in significantly mutated genes, such as *PIK3CA* in breast cancer. We also provide evidence that promoter de-methylation and close non-coding RNAs (namely, CT-ncRNAs) may be two reactivation mechanisms of EECTGs. In vitro or In vivo experiments further proved that two newly identified EECTGs (*LIN28B* and *MEIOB*) and their nearby CT-ncRNAs played a crucial role in tumorigenesis. Therefore, our findings provide new perspectives for identifying epi-driver genes of cancer.

3074T

Genome specific transcriptional signatures predict differentiation biases in Human ES/IPS cells. *G. Stein-O'Brien[1,2], A. Jaishankar[1], S. Suel-Kee Kim[1], S. Seo[1], J. Heon Shin[1], D. Hoeppner[1], J. Chenoweth[1], T. Hyde[1], J. Kleinman[1], D. Weinberger[1], E. Fertig[3], C. Colantuoni[1], R. McKay[1].* 1) McKusick-Nathans Inst of Genetic Medicine, JHMI Baltimore, MD; 2) Lieber Inst For Brain Development, Baltimore, MD; 3) Oncology Dept, Biostatistics and Bioinformatics Div, JHMI, Baltimore, MD.

Predicting the effect of an individuals genetic background is key to advancing personalized medicine. To capture the mechanisms by which these effects emerge, time-course data of the transcriptome in human ES/IPS cells during pluripotency and differentiation conditions from numerous backgrounds was collected. Novel whole-genome Coordinated Gene Activity in Pattern Sets (CoGAPS) analysis of this RNA-seq data clearly separated shared developmental trajectories from unique transcriptional signatures for each individual's genome. These signatures were able to identify their respective donors in data from multiple tissues and across technical platforms, including RNA-seq of port-mortem brain, micro arrayed embryoid bodies, and publicly available datasets. Further analysis of these signatures found they were predictive of lineage biases during neuronal differentiation. Individuals whose signatures had high rankings of OTX2 and SOX21 also had enhanced induction of markers of telecephallic precursors (i. e. PAX6, WNT1). Conversely, signatures with enrichment of retinoic acid responsive genes (p<1E-5) corresponded to enhanced induction of hindbrain markers including HOXB1 and HOXB4 as well as OLIG2 positive cells. Further, lineage biases were consistent with early differences in morphogenetic phenotypes within monolayer culture, thus, linking transcriptional genomic signatures to stable quantifiable cellular phenotypes. Interestingly, matched single nucleotide polymorphisms (SNP) and RNA-seq from the Lieber Post-mortem Brain Collection indicate that relatedness of any two individuals transcriptional signature is independent of the race of the individuals. As many complex diseases are often attributed to the 70-90 percent of human variation found within race, the identification of signatures that define the functional rather than physical background of an individual's genome has the potential to profoundly influence understanding of human disease. For example, the relationship between pathway enrichment in these transcriptional signatures and drug response differences between individuals IPS cells may prove powerful for optimal patient stratification in clinical trials. .

3075F

RNAseq based microRNA and transcriptome profiles of the rat hippocampus. *A. Matos[1], A. Vieira[1], A. Canto[1], K. Brumatti[1], C. Rocha[1], B. Carvalho[1], V. Pascoal[2], R. Glioli[1], I. Lopes-Cendes[1].* 1) UNICAMP, Campinas, Brazil; 2) UFF, Nova Friburgo, RJ, Brazil.

Gene expression can be regulated by microRNAs and they are important in many key biological functions. It has been demonstrated that different microRNAs may have different expression pattern in different brain regions. The hippocampus is functionally highly heterogeneous and it is likely to present different pattern of gene expression in different subfields. RNAseq-based transcriptome analyzes offers the possibility of accurate profiling global gene expression. The aim of this study was to analyze and correlate microRNA and transcriptome profiles in different regions of the hippocampus: Dentate Gyrus (DG), CA1, CA2 and CA3 subfields. Five male Wistar rats were euthanized and the brains were processed for laser microdissection using Zeiss PALM LCM. DG, CA1, CA2 and CA3 were collected from each rat, total RNA was extracted, and libraries were prepared for total RNA and small-RNA sequencing in an Illumina Hiseq platform. Sequences were aligned and quantified with the TopHat/DESeq2 pipeline for total RNA and bowtie/DESeq2 for small-RNA. Gene Onthologies, molecular networks and gene interactions were analyzed with the MetaCore® software. The miRNA target was analyzed by miRWalk® software. A total of 63 miRNAs and 4693 genes were found to be differentially expressed (p<0,05) when comparing CA1 with DG. In CA2, 131 miRNAs and 5819 genes were differentially expressed in comparison with DG. In CA3, 96 miRNAs and 6331 genes were found to be differently expressed. Furthermore, when comparing CA2 to CA1, 54 miRNAs and 2645 genes; CA3 to CA1, 48 miRNAs and 4096; and CA3 to CA2, 38 miRNAs and 1744. Enriched gene ontology categories identified miRNAs involved in p53 signaling, immune response IL-10 signaling pathway, microRNA-dependent inhibition of epithelial to mesenchymal transition. For transcriptome data, the networks were involved in receptor-mediated axon growth repulsion, epidermal growth factor receptor signaling pathway, synaptic vesicle fusion and recycling in nerve terminals. Overall, the most activated genes were regulated by miR- 204, 205, 652, 187, 204, 532-3p, 873, 652, 34b and 375. The present data indicates a large spatial heterogeneity in gene and microRNA expression in the rat hippocampus subfields. A better understanding of such variability may improve our interpretation of the molecular mechanisms involved in normal and pathological conditions related to the hippocampus, such as epilepsy and Alzheimer´s disease. Supported by CEPID-BRAINN, FAPESP.

3076T

Integrative analysis of RNA, translation and protein levels reveals distinct regulatory variation across humans. *C. Cenik[1], E. Sarinay Cenik[1], G. W. Byeon[1], F. Grubert[1], S. I Candille[1], D. Spacek[1], H. Tilgner[1], C. L. Araya[1], H. Tang[1], E. Ricci[2], M. P. Snyder[1].* 1) Department of Genetics, Stanford University School of Medicine, Stanford, CA; 2) Université de Lyon, Lyon, France.

Elucidating the consequences of genetic differences between humans is essential for understanding phenotypic diversity and personalized medicine. Although variation in RNA levels, transcription factor binding and chromatin have been explored, little is known about global variation in translation and its genetic determinants. We used ribosome profiling, RNA sequencing, and mass spectrometry to perform an integrated analysis in lymphoblastoid cell lines from a diverse group of individuals. We find significant differences in RNA, translation, and protein levels suggesting diverse mechanisms of personalized gene expression control. Combined analysis of RNA expression and ribosome occupancy improves the identification of individual protein level differences. Finally, we identify genetic differences that specifically modulate ribosome occupancy - many of these differences lie close to start codons and upstream ORFs. Our results reveal a new level of gene expression variation among humans and indicate that genetic variants can cause changes in protein levels through effects on translation.

3077F

Comparison of gene expression biomarker classifiers developed for use as an Alopecia Areata Disease Activity Index (ALADIN). *J. E. Cerise[1], A. Jabbari[1], M. Duvic[3], M. Hordinsky[4], D. Norris[5], V. Price[6], J. Mackay-Wiggan[1], R. Clynes[1], A. M. Christiano[1,2].* 1) Dept of Dermatology, Columbia University, New York, NY; 2) Dept of Genetics and Development, Columbia University, New York, NY; 3) Dept of Dermatology, MD Anderson Cancer Center, Houston, TX; 4) Dept of Dermatology, University of Minnesota, Minneapolis, MN; 5) Dept of Dermatology, University of Colorado, Denver, CO; 6) Dept of Dermatology, UCSF, San Francisco, CA.

Alopecia areata (AA) is a common autoimmune disease manifesting with hair loss ranging from spontaneously resolving patches to lifelong total body involvement. Accessibility of the target organ within skin biopsies has provided the opportunity to develop a novel disease activity score based on quantitative composite gene expression signatures. Using both Affymetrix microarrays and RNA-Seq data, we previously identified three striking gene expression signatures in total skin from both human AA and the C3H-HeJ mouse, namely, an IFN-γ response (IFN) signature, a cytotoxic T cell (CTL) signature, and a hair keratin (KRT) signature. To generate a functional biomarker from the AA transcriptome, we developed the Alopecia Areata Disease Activity Index (ALADIN), a three-dimensional quantitative composite score. ALADIN, derived from expression levels of representative genes from the IFN, CTL and KRT signatures provides a measure of the distance of AA transcriptional levels from a baseline obtained from the scalp skin of healthy individuals. In our earlier studies of mouse AA, we demonstrated that ALADIN scores provide a quantitative measure of pharmacologic prevention and treatment of disease. Here, using Affymetrix microarray data from RNA obtained from scalp biopsies from a multi-center cross-sectional study, we identify single-gene features that differentiate patients with varying levels of disease severity, as well as "meta-genes" comprised of genes whose expression is highly correlated across samples. We develop biomarker classifiers from the single-gene features as well as from composite features, i. e. , ALADIN signatures and co-expression meta-genes using several methods including diagonal linear discriminant analysis, k-nearest neighbors, logistic regression, and support vector machines. We perform both internal cross-validation and external validation of the classifiers' ability to discriminate patients' disease severity in a second cross-sectional study and examine the classifiers' performance using data from a longitudinal study in which we tracked patients enrolled in an interventional clinical study of AA. Our findings demonstrate the potential utility of ALADIN and these classifiers as a dynamic functional biomarker to stratify and longitudinally track patients enrolled in observational and interventional clinical studies of AA.

3078T

Method to Delineate and Confirm Genetically Distinct Cell Populations from Mosaic Individual Shows Differential Expression Related to Disease Pathway. *C. M. Grochowski[1], E. A. Tsai[5], H. C. Lin[2,3], K. M. Loomes[2,3], N. B. Spinner[1,4].* 1) Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Pediatric Gastroenterology, Hepatology, and Nutrition. The Children's Hospital of Philadelphia, Philadelphia, PA; 3) Department of Pediatrics, Perlman School of Medicine, University of Pennsylvania, Philadelphia, PA; 4) Department of Pathology and Laboratory Medicine, Perlman School of Medicine, University of Pennsylvania, Philadelphia, PA; 5) Genomics and Computational Biology Graduate Group, Perlman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Human mosaicism is a unique genetic phenomenon that occurs when two genetically distinct cell populations arise from a single fertilized egg. We studied a proband with Alagille syndrome and a chromosome 20p deletion including *JAG1*, and her clinically unaffected mother, who is mosaic for the deletion. We developed a method to separate the deletion and normal cells from the mosaic mother, and test the resulting lines for purity. Our goal was to then use the purified cells to analyze differential gene expression associated with the *JAG1* deletion, by comparison to the isogenic normal cell line. Normal and *JAG1*-deleted cell lines were created from a fibroblast cell line from the mosaic individual, by clonal expansion. Induced pluripotent stem cells (iPSC's) were generated from each of the lines and genotyped on the HumanHap 610 Quad BeadArray. The B Allele Frequency of heterozygous SNPs contained in the deleted 20p region was used to determine the percent mosaicism of each derived iPSC line. In order to determine the precise breakpoints of the deletion, we used a high-density array followed by long-range polymerase chain reaction (PCR) to obtain a nucleotide-level resolution surrounding the breakpoint region. Subsequent Sanger sequencing revealed the exact breakpoint architecture, which included 7 bases of microhomology. To efficiently identify the presence of the deleted chromosome in our cell lines, we designed a primer set that would only amplify a 600bp PCR product in the presence of the *JAG1*-deleted line. To rule out low-level mosaicism, primers were designed to amplify a SNP (rs2273061) in intron 3 of *JAG1,* which would be heterozygous only in a non-deleted line. Viewing the presence and relative ratio of the SNP on the chromatogram allowed for a rapid and precise assay to determine clone purity. Subsequent microarray analysis on the cloned deleted and normal cell lines showed many genes to be differentially expressed between the two populations. Decreased expression levels of the known deleted gene *JAG1* was used as a positive control within the experimental model. Several differentially expressed genes implicated in the Notch signaling pathway identified, included the following: *EDN1,* an inhibitor of *JAG1*, *HEY2*, a downstream target of Notch genes, and *CNTN1*, a less common Notch ligand all show overexpression in deleted lines. Future analysis will help to delineate how the expression differences in these genes could be related to the disease phenotype.

3079F

Proteomic profile of dorsal and ventral dentate gyrus of a rat epilepsy model induced by electrical stimulation and displaying classical hippocampal sclerosis. *A. Morato do Canto[1, 2], A. Hilario Berenguer Matos[1, 2], A. Schwambach Vieira[1, 2], R. Glioli[3], I. Lopes-Cendes[1, 2].* 1) Medical Genetics, School of Medical Science - UNICAMP, Campinas, Brazil; 2) Brazilian Institute for Neuroscience and Neurotechnology - BRAINN, Campinas, Brazil; 3) Multidisciplinary Center for Biological Investigation of Laboratory Animals (CEMIB), Campinas, Brazil.

PurposeProteomic analysis is a promising tool for the identification of key biological processes leading to epilepsy. However, the power of such "omic" approach is dependent on the preparation of homogeneous cell populations. In this context, laser-capture microdissection presents the ability to select specific cell populations that would give the most informative data in proteomic studies. The aim of this study is to identify differentially expressed proteins in the dorsal and ventral Dentate Gyrus (dDG and vDG) from epileptic rats. In these animals epilepsy was induced by a perforant pathway stimulation protocol which leads to classical hippocampal sclerosis, similar to what is seem in patients with mesial temporal epilepsy. **Methods** Rats were induced as described by Norwood et al. , 2010. Frozen sections were prepared and the dDG and vDG were laser microdissected (Zeiss PALM). Total proteins were obtained from using 8M urea and analyzed by LC-MS/MS using an LQT-Orbitrap (Waters) and the quantitative data were obtained using the software Scaffold 4. 0. **Results**We identified a total of 1271 proteins in samples of dDG and vDG combined. Of these, 42 proteins were found to be differentially expressed in dDG and 50 in vDG. Although there was some overlap between proteins that were differently expressed in dDG and vDG, we found that 76% of proteins differently expressed in dDG and 80% in vDG were unique to these sub-fields. Most of the differentially expressed proteins are involved in neuronal pathways such as GABA-receptor recycling, neurofilaments remodeling and constitutive AMPA receptor pathway, as indicated by gene ontology analysis employing the Metacore® software (Thomson Reuters). **Conclusion**The proteins identified in the present study can indicate new pathways involved in epileptogenesis. Furthermore, we found that additional molecular complexities could be elicited as hippocampal subfields were analyzed separately. We believe that the further integration of proteomic data with other "omics" approaches could generate even more informative data about the neuronal processes involved in epilepsy. Supported by BRAINN-CEPID/FAPESP, BRAZIL.

3080T

RNAseq reveals distinct transcriptomic profiles of diverse cell populations within human skin. *R. Ahn, K. Taravati, K. Lai, K. Lee, J. Nititham, R. Gupta, M. Rosenblum, W. Liao.* Dermatology, University of California, San Francisco, San Francisco, CA.

Human skin consists of multiple cell types, including epithelial, immune, and stromal cells. Transcriptomic analyses have previously been performed from bulk skin samples or from epithelial and immune cells expanded in cell culture. However, transcriptomic analysis of bulk skin tends to drown out expression signals from relatively rare cells while cell culture methods may significantly alter cellular phenotypes and gene expression profiles. To identify distinct transcriptomic profiles of multiple cell populations without substantially altering cell phenotypes, we employed a novel fluorescence activated cell sorting method to isolate keratinocytes, myeloid dendritic cells, CD8+ T effector cells, CD4+ T effector cells, and CD4+ Treg cells from healthy skin samples, followed by RNAseq of each cell population. Principal components analysis revealed distinct clustering of cell types across samples, while gene expression and pathway analyses revealed transcriptional profiles of individual cell populations distinct from bulk skin. Our work provides a high resolution view of cutaneous cellular gene expression and suggests that transcriptomic profiling of bulk skin may inadequately capture the contribution of rarer cell types.

3081F

NGS Library Preparation Method for Transcriptome Profiling With Enhanced Sensitivity of Transcript Detection. *D. Rodriguez, E. Yigit, M. Karaca, D. Munafo, P. Liu, L. Apone, V. Panchapakesa, C. Sumner, C. Chater, L. Mazzola, J. Bybee, D. Rivizzigno, F. Stewart, E. Dimalanta, T. Davis.* New England Biolabs, Ipswich, MA.

Transcriptome profiling by Next Generation Sequencing (RNA-seq) is a powerful technique for large-scale genome-wide expression analysis, including mRNA, non-coding RNA, and miRNA profiling. RNA-seq is beginning to have multiple applications in clinical diagnostics; providing better insights into how altered RNA expression impacts the biological pathways and the molecular mechanisms associated with disease progression. The successful adoption of NGS into the clinical laboratory will depend on the library preparation techniques that can capture the entire molecular repertoire within a sample without any sequence bias. Here, we present a method for directional or strand-specific RNA-seq that retains information about which strand of DNA is expressed. Determining the polarity of RNA transcripts is important for the correct annotation of novel genes, identification of antisense transcripts with potential regulatory roles, and for correct determination of gene expression levels in the presence of antisense transcripts. This method is based on the labeling and excision of the second strand cDNA, and it is compatible with both poly A-tail enriched and ribosome-depleted RNA. We will provide a comprehensive analysis of different eukaryotic and prokaryotic RNA libraries including library performance, library complexity, duplication rate, and continuity of gene coverage and strand specificity. Our results show this improved method generates high library yields that allow use of low amount of input RNA with reduced PCR duplicates, delivering high quality strand-specific RNA-seq data. This streamlined protocol is also amenable to large-scale library construction and automation.

3082T

Longevity and the transcriptome: Identifying gene expression changes in long-lived individuals. *K. A. Mather[1], A. Thalamuthu[1], B. J. Chen[2], M. Janitz[2], N. J. Armstrong[3], P. S. Sachdev[1, 4], Sydney Centenarian Study and the Sydney Memory and Ageing Study teams.* 1) Centre for Healthy Brain Ageing, UNSW Australia, Sydney, NSW, Australia; 2) School of Biotechnology and Biomolecular Sciences, UNSW Australia, Sydney, NSW, Australia; 3) Mathematics and Statistics, Murdoch University, Perth, Australia; 4) Neuropsychiatric Institute, Prince of Wales Hospital, Sydney, Australia.

Background: Our population is ageing and will place a significant burden on the health system. Long-lived individuals are examples of successful ageing, many of whom have avoided age-related disease until very late in their lives or escape them altogether. Age-related changes in gene expression have been reported, however, few studies have examined the transcriptome comprehensively in long-lived individuals. This study aims to assess transcriptomic differences between long-lived and younger individuals using RNA sequencing. **Method:** Total RNA was extracted from whole blood collected in PAXgene tubes from 24 long-lived (mean age = 97. 8 years) non-demented participants from the Sydney Centenarian Study (Sachdev et al. 2010) aged 95 and over and 24 younger cognitively normal controls (mean age = 78. 8 years) from the Sydney Memory and Ageing Study (Sachdev et al. , 2010). The two groups were matched for sex and there were equal numbers of men and women. After ribosomal RNA depletion, poly A+ RNA libraries were prepared. Paired-end (101 bp) sequencing was performed using Illumina HiSeq. Sequenced reads were processed using TopHat and Cufflinks. Transcriptome profiles of the two groups were compared using three popular methods, DESeq, edgeR and baySeq. **Results:** Using the overlap between the results (FDR-corrected p-value<. 05) from three commonly used methods of analysis, a subset of genes that were differentially expressed was identified (N=4,116). These genes included many genes that have been previously associated with longevity such as *ATM*, *WRN*, *SIRT1* and *MTOR*. **Conclusions:** Using RNA sequencing, differentially expressed genes were observed in long-lived individuals compared to younger controls. This work will provide insights into the genes and pathways underlying exceptional longevity and suggest avenues for future research that will inform us about the molecular mechanisms involved in successful ageing.

3083F

RNA-sequencing reveals novel candidate genes for Ring chromosome 20 syndrome. *D. A. McEldrew[1], R. Rajagopalan[1], C. M. Grochowski[1], A. M. Falsey[1], L. K. Conlin[1,2,3], J. A. Mills[1], N. B. Spinner[1,2,3].* 1) Dept. Of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia, Philadelphia, PA; 2) Division of Genomic Diagnostics, The Children's Hospital of Philadelphia, PA; 3) Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA.

Ring chromosomes are a rare chromosome abnormality caused by fusion of the short and long arms of a chromosome. All human chromosomes have been identified with ring structures and the overall frequency of constitutional ring chromosomes isestimated to be 1 in 30,000 to 1 in 60,000 births. Ring chromosomesare often associated with clinical features which can vary depending on the chromosome, but refractory seizures, microcephaly, behavioral abnormalities and mild to severe cognitive impairment are common to a number of rings. Ring chromosome 20 is one of the more common rings, and we previously showed that some rings are present in all cells of a patient and these rings always contain deletions of one or more arms; while other rings are mosaic, with normal and ring containing cells and in all cases studied, these mosaic cases have no demonstrable deletions or duplications. Skin samples were collected were collected from two mosaic R(20) patients, cultured into fibroblast cell lines and the normal and ring cells were cloned out to separate the cell types. We established cell lines containing pure r(20) cells and pure normal from these two patient samples. Cell proliferation assays were performed showing a decreased growth potential in R(20) fibroblast compared to isogenic controls (P < 0. 05). We performed RNA-sequencing to characterize gene expression profiles of these samples and to potentially identify genes associated with decreased growth in R(20) fibroblast cells. We used STAR aligner to align fastq sequences to the human transcriptome and used multiple computational tools such as DESeq2 and the CuffLinks pipeline to estimate differential gene expression. We found a total of 987genes differentially expressed between ring and normal samples. Of the 987 differentially expressed genes, 861 were up-regulatedand 126 were down-regulated in thering sample. PLCB1 was the top hit on chromosome 20 for differential expression, and this gene has been shown to be involved in cell proliferation, differentiation, and cellular maintenance. We also found several other differentially expressed genes (TGM2, SALL4, CASS4, DNMT3B etc.) on chromosome 20 that may be candidates for altered growth potential. To our knowledge, this isthe first report of differential gene expression analysis of ring chromosomes and we believe this will enhance our understanding of the molecularmechanisms of Ring Chromosome 20 syndrome.

3084T

Global Gene Expression Profiling during Pregnancy. *A. Mittal[1], L. Pachter[2], M. Smed[3], H. Kjaergaard[3], V. Zoffmann[3], I. Hallgrimsdottir[2], D. Jawaheer[1].* 1) Children's Hospital Oakland Research Institute, Oakland, CA 94609; 2) University of California, Berkeley, CA 94720; 3) Juliane Marie Center, Rigshospitalet, Copenhagen, Denmark.

Introduction: Little is known about systemic changes in the maternal transcriptome duringnormal human pregnancy since previous studies focused mostly on the maternal-fetal interface. The gene expression profiles specific to different trimesters of pregnancy are also poorly understood. We have examinedglobal transcriptome changes that occur in the bloodstream of healthy women from pre-conception topostpartum to identify genes that are involved in different stages of pregnancy. Methods:Using a longitudinal study design, blood was drawn from 5 healthy women before conception, once each trimester and at 3 intervals postpartum. Total RNA was extracted and used to prepare cDNA libraries which were sequenced on Illumina HiSeq2500 instrument to generate on average 60 million paired-end reads (100bp). The raw RNA sequencing (RNA-seq) data was preprocessed and aligned to the reference Human transcriptome using Bowtie2. Gene expression levels were quantified using eXpress. The edgeR package was used to perform differential gene expression analysis across different time points. The enrichment of specific biological processes and pathways among differentially expressed genes were analyzed using DAVID bioinformatics resource. Results: A total of 790 genes showed at least 2-fold differential expression (p*<0. 05) during pregnancy with respect to the pre-pregnancy baseline, while their post-partum expression profiles reverted back to the pre-pregnancy levels. Among these were genes having common expression patterns across all three trimesters as well as genes with profiles specific to the stage of gestation. Specifically, we identified genes that were a) upregulated in all three trimesters (enriched in processes involved in inflammatory response, translational elongation and ribosomal biogenesis and assembly), b) upregulated in second and third trimesters (enriched in defense response, homeostasis, erythrocyte differentiation and negative regulation of immune system process), c) downregulated in all three trimesters (genes involved in metal ion binding), d) downregulated in first trimester (including ALOX15, IL5RA, NR4A1 (p*<0. 01)), and e) downregulated in third trimester (including ENPP3, HDC, CPA3 (p*<3x10-5)). *FDR correctedConclusions: These gestation-specific gene signatures can lead to a better understanding of what systemic changes occur in the mother during healthy pregnancy, and may provide molecular insights into disorders that are influenced by pregnancy.

3085F

Transcriptome analysis of CD4+ T cells in coeliac disease. *C. Coleman, E. M. Quinn, B. Molloy, P. Dominguez Castro, V. Trimble, N. Mahmud, R. McManus.* Clinical Medicine, Trinity College Dublin, Dublin, Ireland.

Genetic studies have to date identified 43 genome wide significant coeliac diseasesusceptibility (CD) loci comprising over 70 candidate genes. However, how alteredregulation of such disease associated genes contributes to CD pathogenesis remainsto be elucidated. Recently there has been considerable emphasis on characterisingcell type specific and stimulus dependent genetic variants. Therefore in this study weused RNA sequencing to profile over 70 transcriptomes of CD4+ T cells, a cell typecrucial for CD pathogenesis, in both stimulated and resting samples from individualswith CD and unaffected controls. We identified extensive transcriptional changesacross all conditions, with the previously established CD gene IFNy the most stronglyup-regulated gene (log2 fold change 4. 6; Padjusted=2. 40x10-11) in CD4+ T cells fromCD patients compared to controls. We show a significant correlation of differentiallyexpressed genes with genetic studies of the disease to date (Padjusted=0. 002), and 21CD candidate susceptibility genes are differentially expressed under one or more of theconditions used in this study. Pathway analysis revealed significant enrichment ofimmune related processes. Co-expression network analysis identified several modulesof coordinately expressed CD genes. Two modules were particularly highly enrichedfor differentially expressed genes (P<2. 2x10-16) and highlighted IFNy and othergenetically associated transcription factors which showed significantly reducedexpression in coeliac samples as key regulatory genes in CD.

3086T

Transcriptomic analyses reveal sex-dependent expression differences in hundreds of genes and show an upregulation of immune response genes in women. *A. Joensuu[1,2], ML. Nuotio[1,2], E. Raitoharju[3], M. Kähönen[4], T. Lehtimäki[3], V. Salomaa[2], O. Raitakari[5,6], M. Perola[1,2,7], J. Kettunen[2,8,9].* 1) National Institute for Health and Welfare, Helsinki, Finland; 2) Dept of Health, National Institute for Health and Welfare, Helsinki, Finland; 3) Dept of Clinical Chemistry, Fimlab Laboratories and School of Medicine, University of Tampere, Tampere, Finland; 4) Dept of Clinical Physiology, Tampere University Hospital and School of Medicine, University of Tampere; 5) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; 6) Turku University Hospital, Turku, Finland; 7) University of Tartu, Tartu, Estonia; 8) Computational Medicine, Institute of Health Sciences, University of Oulu, Oulu, Finland; 9) NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland.

Background: The expression of genes regulates biological pathways and processes and thus gives rise to differences in observed phenotypes. Identifying common sex-dependent differences in transcription levels in a large cohort with sufficient power is important for better interpretation of smaller studies focusing on specific diseases. **Methods:** We studied differences in autosomal gene expression between males and females in two Finnish population cohorts, the Young Finns Study (575 males and 680 females aged 34-49) and DILGOM (240 males and 278 females aged 25-74). Both cohorts have genome-wide transcriptional coverage (>30,000 transcripts) from whole blood measured with Illumina HumanHT-12 BeadChips. The effect of sex on quantile-normalized expression levels corrected for age, BMI and smoking was analyzed separately in the cohorts with linear regression and combined with random-effects meta-analysis. Women using contraceptive pills or pregnant were excluded. Functional annotations were studied with DAVID using significantly expressed genes as input and all genes in the meta-analysis as the background reference gene set. **Results:** Expression of 512 transcripts in 491 (2. 6%) autosomal genes showed higher expression in males whereas levels of 480 transcripts in 460 (2. 4%) genes were higher in females. Interestingly, in the transcripts more expressed in females a 3. 1-fold enrichment of genes related to Immune response Gene Ontology (GO) processes was seen (FDR=2. 5*10-8). Supporting this, the most strongly elevated transcript in females (b=0. 61; 95% CI:0. 51-0. 70; p-value 7. 5e-37) was seen in the gene *Lymphotoxin beta* (*LTB*), an inducer of the inflammatory response system. Genes more expressed in females also showed enrichment in the GO processes Apoptosis and Programmed cell death as well as in the Cytokine-cytokine receptor interaction KEGG pathway. In contrast no enriched GO processes or KEGG pathways passed the FDR<0. 05 criterion in males. **Conclusions:** We could detect significant sex-dependent enrichment of transcription in 5% of the genes studied. Genes related to immune response are more expressed in females, which could reflect the higher female prevalence of most autoimmune diseases. Next we will use Weighted Gene Co-expression Network Analysis to identify sex-dependent transcription modules which can be tested against further phenotypes. Similar studies may help to gain insight into some of the sex-dependent susceptibility rates observed in various diseases.

3087F

Gene expression of *PTK2B* in peripheral blood is associated with cortical thickness in Alzheimer's disease. *K. Nho[1,2,3], S. Kim[1,2,3], S. L. Risacher[1,3], H. D. Soares[4], P. Singh[4], L. Wang[4], Z. Qi[4], A. He[4], I. Neuhaus[4], V. Patel[4], T. Foroud[1,2,3,5,6], K. M. Faber[6], L. Apostolova[7], M. W. Weiner[8,9], A. J. Saykin[1,2,3,5], Alzheimer's Disease Neuroimaging Initiative.* 1) Radiology and Imaging Sciences, Indiana University School of Medicine, Indianapolis, IN; 2) Indiana Alzheimer Disease Center, Indiana University School of Medicine, Indianapolis, IN; 3) Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN; 4) Bristol-Meyers Squibb, Wallingford, CT; 5) Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN; 6) National Cell Repository for Alzheimer's Disease, IN; 7) Department of Neurology, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA; 8) Departments of Radiology, Medicine, and Psychiatry, University of California-San Francisco, San Francisco, CA; 9) Department of Veterans Affairs Medical Center, San Francisco, CA.

Background: Human brain function depends on the precise regulation of gene expression. Transcriptome profiling has become a major focus of neurodegenerative disease research. Prior studies suggest that abnormal gene expression patterns may contribute to the onset and progression of late-onset Alzheimer's disease (LOAD). We performed transcriptome and whole-brain cortical thickness analyses of 22 LOAD susceptibility genes identified from recent large-scale GWAS using gene expression profiles from the Alzheimer's Disease Neuroimaging Initiative (ADNI). **Methods**: 661 ADNI non-Hispanic Caucasian participants with Affymetrix Human Genome U219 Array data from peripheral blood samples (PaxGene RNA tubes) were included. Raw expression values were pre-processed using the Robust Multi-chip Average normalization followed by standard quality control (QC) procedures on samples and probe sets. Statistical analysis of microarray data was performed using the regularized t-test to evaluate differences in gene expression between LOAD and cognitively normal controls (HC). Data were adjusted for age, gender, plate, and RIN score (RNA integrity number). Multivariate whole-brain analysis was performed to identify highly correlated brain regions with gene expression levels. Statistical maps using SurfStat were thresholded with random field theory correction for multiple testing. **Results**: Of the 19 genes represented on the array after QC, one gene (*ABCA7*) was significantly up-regulated in LOAD (corrected p=0. 034). Whole-brain analysis of cortical thickness identified brain regions, especially entorhinal cortex, significantly associated with blood expression of *ABCA7* and *PTK2B*. Individuals with lower expression showed greater cortical thickness (corrected p<0. 05). *cis*-eQTL mapping analyses of *ABCA7* and *PTK2B* detected 63 and 16 significant associations with $p < 2.5 \times 10^{-12}$, respectively. *cis*-eQTL of rs6987305 (*PTK2B*) in healthy human brains yielded a p=0. 000096 for the temporal cortex tissue (BRAINEAC). **Conclusions**: This is the first study to show that *ABCA7* and *PTK2B* gene expressions in blood are associated with reduced cortical thickness, especially in entorhinal cortex. These data illustrate the potential of combining transcriptome profiles with neuroimaging to understand the genetics and pathobiology of AD, an approach that may be useful for novel therapeutic target identification for AD.

3088T

Dissection of Gene Expression Signatures of Recurrent Acute Pancreatitis Risk in Extreme Hypertriglyceridemia. *K. Tremblay, D. Gaudet.* Université de Montréal Community Genomic Medicine Centre and ECOGENE-21, Chicoutimi, PQ, Canada.

Background: Recurrent acute pancreatitis (RAP) is defined as more than two attacks of acute pancreatitis (AP) without any evidence of underlying chronic pancreatitis. Extreme hypertriglyceridemia (fasting triglycerides >10 mmol/L) is associated with chylomicronemia (CM) and increased risk of RAP. The number of pancreatitis episodes varies significantly between patients with CM. **Aim:** To investigate the gene expression profiles of AP and RAP in patients with extreme hypertriglyceridemia and CM. **Methods:** A total of 62 consenting subjects participated in this study. Prior to the analyses, patients with CM were divided into three groups covering a wide spectrum of recurrence of AP: 0 episode (n=21), 1-3 (n=10) and ≥ 4 (range 4 to 40, n=16) episodes. The gene expression profile of the three CM groups was compared to the profile of 15 healthy controls. AP was defined according to the revised Atlanta classification. RNA samples were extracted from whole blood PAXgene and hybridized on Affymetrix Human Gene 2. 0 ST Array according to manufacturer procedures. Robust Multi-Array normalization has been applied on the probe raw intensities. Differential expression moderated T-tests between studied groups were performed using a linear model of the Bioconductor package Limma. False discovery rate (FDR) estimation was carried out using the Benjamini-Hochberg method. **Results:** The studied groups allowed conducting six differential analyses on a total of 48,226 detected probes. At a p-value < 0. 01, a FDR of 5% and a >2-fold down-expression (down) or over-expression (over) significance levels, a set of 41 probes have been found differentially expressed in CM subjects with no pancreatitis, 106 in the CM group with 1 to 3 pancreatitis, and 90 in the group with ≥ 4 pancreatitis compared to healthy controls. Of the identified probes, 33 are shared by all CM groups; 6 are specific to CM with no pancreatitis [3 down (2 annotated;1 non-annotated) and 3 over (2 annotated;1 non-annotated)]; 41 are specific to CM with 1 to 3 pancreatitis [30 down (15 annotated;15 non-annotated) and 11 over (7 annotated;4 non-annotated)], and finally 27 are specific to CM with ≥ 4 pancreatitis [24 down (11 annotated;13 non-annotated) and 3 over (2 annotated;1 non-annotated)]. Most of the annotated genes are involved in inflammatory, immune or signaling biological pathways. **Conclusion:** Using extreme hypertriglyceridemia as a model, these results reveal gene expression signatures of RAP.

3089F

Impact of Genetic Variation on upstream Open Reading Frame (uORF) Translation and Differential Gene Expression in Human Populations. *Y. Wang[1], T. Chun[2], R. Mills[1,2].* 1) Human Genetics, University of Michigan, Ann Arbor, MI; 2) Computational Medicine & Bioinformatics, University of Michigan, Ann Arbor, MI.

Upstream open reading frames (uORFs) are regions of protein coding potential that begin in the 5' untranslated region (UTR) of an mRNA and end either before or after the start codon of the canonical gene. Previous studies have demonstrated that the existence of uORFs in the 5'-UTR of genes is associated with lower expression level of these genes at both the mRNA and protein levels by multiple possible mechanisms, and recent work has combined DNA and RNA sequencing data with ribosomal profiling and proteomics to investigate the extent to which this occurs. Here, we explore this further by applying a novel method for identifying actively translating regions using ribosomal profiling data to identify differentially translated uORFs due to genetic polymorphisms that create or destroy the open reading frame across human populations. We present an analysis of these regions within the context of local gene expression, translation, protein abundance, and conservation in other primate species. We further examine the effect of such variation on putative ORFs in noncoding regions throughout the genome and compare with previously reported sets of expression quantitative trait loci in order to discover potential regulatory regions.

3090T

PennDiff: Detecting differential alternative splicing from RNA-Seq data. *Y. Hu[1], C. Xue[2], M. Reilly[2], M. Li[1].* 1) Department of Biostatistics and Epidemiology, University of Pennsylvania Perelman School of Medicine Philadelphia, PA; 2) Cardiovascular Institute, University of Pennsylvania Perelman School of Medicine Philadelphia, PA.

Alternative splicing is a regulated process and a major mechanism for generating protein diversity. Detecting differential alternative splicing (DAS) between two groups of samples (e. g. , cases vs. controls, before vs. after drug treatment) could provide an effective way to discover disease susceptibility genes. To detect DAS from RNA-Seq data, we developed a novel statistical method PennDiff, which makes use of information on known gene structures and pre-estimated isoform relative abundances. For each alternatively spliced exon of a gene, PennDiff divides isoforms into two categories depending on whether the exon is included or not. The inclusion level of the alternatively spliced exon is then estimated as the total relative abundances of all isoforms with the exon included. PennDiff has several advantages. First, grouping exons avoids multiple testing for “exons” originated from the same isoform. Second, it utilizes all available sequencing reads in exon-inclusion level estimation, which is in sharp contrast to DEXSeq and rMATS that only use exon+junction reads. Third, collapsing isoforms sharing the same alternatively spliced exons reduces the impact of isoform expression estimation uncertainty and yields more accurate estimate of exon-inclusion level. To detect DAS, we assume the exon-inclusion levels follow a beta distribution, and the correlation between alternatively spliced exons within the same gene is modeled through the use of Gaussian copula. Through this Gaussian copula framework, PennDiff is able to perform both exon-based and gene-based tests, thus offering much more flexibility than existing methods. To evaluate the performance of PennDiff, we analyzed both simulated data and a real RNA-Seq dataset from peripheral blood mononuclear cells (PBMCs) collected before and after experimental human endotoxemia. We compared PennDiff with several existing methods, including Cuffdiff, DEXSeq, rMATS, IUTA and SplicingCompass. Our results indicate that PennDiff has well-controlled type I error rate, and is more powerful than existing methods, especially when sequencing depth is low. In contrast, Cuffdiff and rMATS suffer from substantial loss of sensitivity. We are performing experimental validations of DAS genes that were detected by PennDiff but were missed by all other tests in the PBMC dataset. As the use of RNA-Seq continues to grow exponentially, we expect PennDiff will be useful for studies of alternative splicing in diverse RNA-Seq projects.

3091F

Mesenchymal stem cells of patients with multiple myeloma have different epigenetic characteristics compared to normal bone marrow stem cells. *m. noruzinia[1,2], m. Mashadikhan[1, 2], A. saremi[2, 3], M. ahmadvand[1, 2], A. dehghani[1, 2], O. bashti[1, 2].* 1) tarbiat Modares University of Medical Sciences, Tehran, Tehran , Iran; 2) Sarem Cell Research Center, Tehran, Iran; 3) Sarem Women Hospital, Tehran, Iran.

Epigenetic aberrations are reported to contribute in pathogenesis of multiple myeloma. Histone modifications and DNA methylation are two major epigenetic modifications. Epigenetic differences are detected between patients with multiple myeloma and normal. These differences are reported in gene specific methylation as well as global hypomethylation. Bone marrow Mesenchymal stem cells in patients have been involved in the pathogenesis of multiple myeloma. Osteoblastogenesis is impaired in MM and it has been shown that abnormalities in bone marrow mesenchymal stem cells are involved. We sought to search the role of epigenetics in the difference between normal bone marrow stem cells and multiple myeloma bone marrow stem cells. Materials and methods:BM nucleated cells (BMNC) from 6 normal donors and 6 MM patients were plated, cultured inα-MEM medium. After passage 3 the cell were put through flowcytometry RT PCR and invitro differentiation to confirm the authenticity of mesenchymal stem cells. RT PCR for transcripts of genes involved in epigenetic programing and DNMTs, were performed to compare the two patients and normal. Resuts:Mesenchyal stem cells of MM patients had slower growth. Immunophenotyping confirmed absence of myeloma cells. ALPL expression was significantly lower in MM MSCs compared to normal MSCs. DNMT1 DNM3A and DNM3B are differentially expressed in MM MSCs compared to Normal MSCs. Discussion:Recent investigations show that Mesenchymal stem cells in patients with myeloma might be important in the pathogenesis of this disease. Current treatments do not show satisfying efficacy in the treatment of this malignancy of plasma cells. On the other hand epidrug therapy has shown to be promising in the treatment of many cancers as well as multiple myeloma. We sought to find whether there is a difference between some epigenetic characteristics of MSCs of MM patients and Normal persons. MM MSCs seem to be different from normal MSCs as DNA methyl transferases are differentially expressed. ALPL show a diminished expression too. Taken together these result might show a potential for epidrug therapy of MM Mesenchymal stem cell which would help to treat patients who are resistant to current treatments.

3092T

Transcriptome profiling of hypoxia tolerant heterozygous *EdnrB* mice reveals genes associated with cardiac contractility. *T. Stobdan[1], D. Zhou[1], E. Ao-Ieong[2], D. Ortiz[2], R. Ronen[3], I. Hartley[1], Z. Gan[2], A. D. McCulloch[2], V. Bafna[4], P. Cabrales[2], G. H. Haddad[1,5,6].* 1) Pediatrics, Univ. California San Diego, San Diego; 2) Bioengineering, Univ. California San Diego, San Diego; 3) Bioinformatics & Systems Biology Graduate Program, Univ. California San Diego, San Diego; 4) Computer Science and Engineering, Univ. California San Diego, San Diego; 5) Neurosciences, Univ. California San Diego, San Diego; 6) Rady Children's Hospital, San Diego, California, USA.

The Ethiopians highlanders are considered one of the best adapted to high altitude environment but the genetic basis for the adaptation has never been understood. Candidate genes were identified through unbiased whole genome sequencing of adapted individuals and we focused on one of them for deeper analysis. A heterozygote knockout mouse of one of the candidate genes, *Endothelin Receptor B* (*EdnrB*-/+), could tolerate various levels of hypoxia, even extreme hypoxia (5% O2). To understand the molecular mechanisms, we compared the transcriptomic profile of left ventricles of *EdnrB*-/+ to their littermate controls, both under normoxia and hypoxia and used a bayesian approach, conditioning the criteria for mining specific genes and pathways. For within group comparisons i. e. , *EdnrB*+/+ and *EdnrB*-/+ respectively, of normoxia vs hypoxia, we obtained 3870 differentially expressed genes of which 1547 genes were common. Interestingly there were 3 genes, *Nppa*, *Sln* and *Myl4* (q<0. 05) that were oppositely expressed in *EdnrB*+/+ and *EdnrB*-/+ with an average FPKM of >100. They were subsequently validated by RT-PCR. The composite network analysis of these 3 genes were primarily related to Ca2+ homeostasis, regulation of ATPase activity, and regulation of the cardiac contractility, consistent with a better cardiac performance. We conclude that *EDNRB* plays a key role in hypoxia tolerance and the 3 atria-specific genes *Nppa*, *Sln* and *Myl4*, likely play a role in the improved cardiac performance during hypoxia in the heterozygote mice.

3093F

The microbiota and interleukin-10 modulate the chromatin landscape and transcription in lamina propria macrophages and experimental colitis. *J. M. Simon[1], J. P. Davis[1], M. Weiser[1], S. E. Lee[1], G. R. Gipson[1], J. Whitley[1], V. B. Iyer[1], A. G. Robinson[1], R. B. Sartor[1], H. H. Herfarth[1], R. Rahbar[1], T. S. Sadiq[1], M. J. Koruda[1], D. P. McGovern[2], J. D. Lieb[3], K. L. Mohlke[1], T. S. Furey[1], S. Z. Sheikh[1].* 1) University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Inflammatory Bowel Disease Center, Cedars-Sinai Medical Center, Los Angeles, CA; 3) University of Chicago, Chicago, IL.

Chromatin accessibility is dynamic in response to certain stimuli or through cellular differentiation. However, how the underlying genotype contributes to the cellular response to a given stimulus, potentially by altering the chromatin landscape, is not well understood. In this study, we use Crohn's Disease (CD), a chronic inflammatory disease of the intestine, as a model wherein intestinal microbes, environmental triggers, the immune system, and underlying genotype all contribute to disease onset. Macrophages in particular are one key immune cell type that responds inappropriately to intestinal microbes due to underlying genetic variation, resulting in chronic inflammation. To address how macrophages respond to bacteria at the level of chromatin and gene expression, and how this response is altered based on underlying genetics, we performed Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE-seq) in conjunction with RNA-seq on two types of macrophages in wild-type and colitis prone (*Il10*-/-) mice. We identified ~4,700 regions, mainly near pro-inflammatory cytokine genes, with altered chromatin accessibility associated with genotype that poise intestinal macrophages for an aberrant response to microbes. In order to test whether human CD patients exhibited chromatin alterations consistent with this experimental model, we performed FAIRE-seq on genotyped colonic mucosal biopsies from two CD and two control patients. We identified ~300 regions of accessible chromatin associated with disease status that are syntenic to the genetically driven baseline chromatin alterations seen in murine macrophages. We also tested whether *in vitro* immunotherapy could revert the pro-inflammatory chromatin landscape in macrophages established through genetic susceptibility. We found this therapy affected <5% of the 2,883 accessible chromatin sites, possibly explaining the lack of success of these treatments, and concluded that genetics may irreversibly predispose macrophages during differentiation for aberrant responses to intestinal microbes in CD. Our study provides the first insights into how genetics and the environment may irreversibly shape chromatin and alters the macrophage response to microbes, and extends these findings to human disease. .

3094T

Single-cell RNA-seq on pancreatic islets of Langerhans reveals different heterogeneity between non-diabetic and diabetic individuals. *M. Garieri[1,2], C. Borel[1], T. Stoger[3], F. Santoni[1], C. Howald[1,2], P. Ribaux[1], L. Pelkmans[3], P. A. Halban[1], S. E. Antonarakis[1], E. T. Dermitzakis[1,2].* 1) Department of Genetic Medicine and Development, University of Geneva, Geneva, Geneve, Switzerland; 2) Swiss Institute of Bioinformatics (SIB), Geneva, Switzerland; 3) Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland.

Pancreatic islets of Langerhans consist of several endocrine cell types. We investigate the cellular heterogeneity in non-diabetic (ND) and diabetic (D) donors using single cell (SC) RNA-seq. Dissociated islet cells of 3 ND donors (n=216) and 2 type 2 diabetes (T2D) donors (n=120) were analyzed. We also sequenced 93 sorted-beta SCs from the same individuals (57 ND and 36 D). On average we obtained 36. 6 million 100-bp-reads per SC after quality filtering. Clustering analysis on islet SCs identified a first population expressing high levels of insulin (INS) with a significant (p=4. 58e-12) difference between ND and T2D (62% ND SCs, mean 35924 RPKM vs. 24% T2D SCs 14297 RPKM). A second population expressed high levels of glucagon (GCG) (p=8. 064e-05; 8% ND 119263 RPKM vs. 30% T2D 29983 RPKM). Interestingly the mutually exclusive expression of *INS* and *GCG* in beta and alpha cells was not observed in a third population of SCs co-expressing high levels of *INS* and *GCG*: 7% ND islet SCs (0% ND sorted-beta SCs) vs. 23% T2D islets SCs (20% T2D sorted-beta SCs). Furthermore other bi-hormonal SCs are present: 11 INS-SST (10 ND SCs which represent 3. 6% of all ND SCs, and 1 D SC which correspond to 0. 6% D SCs), 22 INS-PPY SCs (17 - 6. 2% ND, 5 - 3. 2% D), 8 GCG-PPY SCs (0 - 0% ND, 8 - 5% D) and 3 GCG-SST (1 - 0. 4% ND, 2 - 1. 3% D). Differential expression analyses reveals enrichment in RNA splicing in D SCs. Beta ND SCS are enriched in growth factor binding pathway, while D alpha SCs for RNA binding. Correlation analysis with INS allows us identify enrichment in peptide hormone processing in ND SCs, while in D SCs there is an enrichment in cell differentiation and cellular developmental process. Allelic specific expression analysis is compatible with the expression of both INS and GCG in one cell, and convincingly excluded the possibility of erroneous analysis of two cells. We controlled for the absence of transcript contamination by sequencing 11 empty chambers. We are verifying dual hormone gene expressing SCs using another technology (branched-DNA single-molecule FISH) allowing for quantification and cellular localization of transcripts. SC transcriptome analysis provides an unprecedented opportunity to explore the heterogeneity of islet cells and compare this in cells from type2 diabetic vs. non-diabetic subjects. M. G and C. B. contributed equally to the work.

3095F

Ultraconserved elements populate genomic regions disrupted in both cancers and neurodevelopmental disorders. *R. B. McCole[1], C. Y. Fonseka[1], B. Harrison[2], R. L. Collins[2], S. Erdin[2], C. Redin[2], M. E. Talkowski[2], C. -T Wu[1].* 1) Genetics, Harvard Medical School, Boston, MA; 2) Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA.

The ultraconserved elements (UCEs) are regions of the human genome possessing extremely high levels of DNA sequence conservation between related species. The question of why UCEs show such extraordinary conservation and their relationship to disease is widely debated. We have demonstrated that UCE-containing genomic regions are significantly depleted from 43,727 copy number variable (CNV) regions pooled from 8 separate studies in healthy human subjects (P=1. 7 × 10-12, obs/exp ratio=0. 8), and in contrast, that UCEs are enriched within 177 genomic regions demonstrating recurrent copy number alteration specifically in cancerous tissues from 17 studies covering 52 different human cancers (P=1. 2×10-4, obs/exp=1. 1). In a small pilot study, we also found that UCEs were enriched within CNVs found in subjects with neurodevelopmental disorders (NDDs), prompting the hypothesis that UCE enrichment in these datasets could be explained by high occurrence of UCEs within critical genes and pathways implicated in cancer formation and NDDs. From analyses of 557 cancer-associated genes listed in the COSMIC cancer gene census, we discovered that UCEs are almost two-fold over-represented against a background of all genes (P=3. 8×10-7, obs/ex =1. 9). With respect to germline variation contributing to NDDs, as expected UCEs were highly enriched among the 1000 most constrained genes in the genome (P=7. 8×10-5, obs/exp=4. 8), and were significantly over-represented within three different lists of autism and NDD associated genes (168 AutismKB genes, 126 DAWN genes, and 605 SFARI genes, P=5. 1×10-9, 2. 7×10-3, and 0. 004 respectively). When we considered loci disrupted by *de novo* inversions and translocations associated with NDDs, we found the genomic regions flanking 135 balanced rearrangement breakpoints were as enriched for UCEs as the cancer-associated loci (P=8. 0×10-6, obs/exp=2. 5). By contrast, UCEs were not significantly enriched within 418 genes associated with schizophrenia by common variant GWAS. Our findings suggest that UCEs may arise and be maintained specifically in disease-relevant genomic regions where multiple and pervasive deleterious effects occur as a result of mutation, perhaps depending on context, such that NDDs arise from mutations present during development, but cancer results when similar mutations occur somatically. UCEs may also provide a new map to the landscape of deleterious genetic variations that manifest in humans as diseases even beyond cancer and NDD.

3096T

Human umbilical vein endothelial cells Response to shRNA-mediated knock down of ITPA gene in presence of Nitric Oxide. *z. Abedi, a. ahmadi, m. behmanesh.* tarbiat modares, tehran, Iran.

Reactive oxygen species (ROS) can increase the production of non-canonical nucleotides. ITPA gene encodes an enzyme which removes the non-canonical inosine triphosphate (ITP) nucleotide from the nucleotide pool. Since ITP can incorporate into the DNA and increase the rate of mutations, ITPA play critical role in maintenance of genome integrity. ITPA deficiency can result in ITP accumulation and DNA lesions. Deamination of DNA bases can lead to mutagenesis through use of deamination products by DNA polymerase and misrepair, or no repair. Nitric oxide reacts with oxygen, it can react with amines, thiols, and other nucleophiles. N2O3, acts on DNA bases and leading to direct DNA damage via deamination. Deamination of cytosine, adenine, guanine, and 5-methylcytosine forms uracil, hypoxanthine, xanthine, and thymine, respectively. The aim of the present study is to find if ITPA gene deficiency and Nitric Oxide treatment may increase the susceptibility of HUVEC cells to ROS. HUVEC cell were cultured in a medium containing doxorubicin as a ROS inducing agent. shRNAs against the ITPA gene was designed and used to knock down of ITPA gene. The efficiency of designed shRNAs was evaluated by real time polymerase chain reaction (qPCR) and western blot. The viability of shRNA transfected cell was evaluated by MTT assay. Results showed that our shRNA could knock down of ITPA effectively. Furthermore, we found that Nitric Oxide induce cell death in HUVEC cells. All together our shRNA worked properly and it is useful to evaluate if knockdown of ITPA enhances apoptosis and decreases the viability in HUVEC cells. Keywords: Inosine triphosphatase, short hairpin RNA, Reactive oxygen species, Human umbilical vein endothelial cell.

3097F

Stimulating human CD4+ memory T cells uncovers cell state-dependent allelic expression enriched in risk genes for rheumatoid arthritis. *M. Gutierrez-Arcelus[1,2], N. Teslovich[1,2], HJ. Westra[1,2], K. Slowikowski[1,2], DA. Rao[3], J. Ermann[3], MB. Brenner[3], S. Raychaudhuri[1,2,4].* 1) Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Smith Building, 1 Jimmy Fund Way, Boston, MA, USA; 4) 4. Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.

In the last decade, investigators have made important advances in understanding the pathogenesis and genetics of autoimmune diseases. However, many questions remain, particularly in understanding how genetic factors act in the specific cell-types and physiological states most relevant for each disease. Our group and others have previously implicated in rheumatoid arthritis and type I diabetes a role for CD4+ memory T cells, essential players in the adaptive immune system. In order to better understand the genetic regulatory variation active in varying cell states in this cell population, we have subjected CD4+ memory T cells from one individual to a physiologically relevant stimulus and have assayed the transcriptome with RNA-seq at six different time-points following stimulation. We have called heterozygous variants from the RNA-seq data. From this high-resolution time course experiment, we have identified many time and stimulus condition specific (TASC) allelic effects, which reflect the activation or inactivation of genetic regulatory effects upon stimulation. We tested these associations with logistic regression after removing PCR duplicates, but we weighted duplicates to preserve allelic information. In a single individual we have tested 3,815 heterozygous sites for which we have sufficient read depth in all time points. At 10% FDR, we have identified 23 heterozygote SNPs with TASC allelic expression, overlapping 15 protein coding genes (pi1=0. 08; estimate of the proportion of true positives in a P-value distribution). The effect size (beta) of the driving time points in these statistically significant events are concordant with other heterozygous variants within the same exon (r=0. 41, P = 0. 03). Noticeably, we find a 3. 7 fold increase in enrichment of low P-values for genes in RA risk loci associated with super enhancers in human CD4+ T cells (pi1=0. 30, N=57), a 4. 2 fold increase for candidate biological RA risk genes (pi1=0. 34, N=56), and none if we take all genes within risk loci for RA with no prioritization (pi1=0, N=125). We are expanding the experiment to 20 densely genotyped individuals. Additionally, we will evaluate which disease-associated pathways are most affected by genetic regulatory variation during T cell activation. This study will yield novel insights into the genetic variation influencing gene regulation in cell-types and cell-states with important roles in autoimmunity.

3098T

Elucidating the gene regulatory mechanisms that control the acute phase response using comparative genomics. *M. Liang[1,2], A. Alizada[1,2], A. Medina-Rivera[1,3], P. Mazrooei[4,5], M. Lupien[4,5], M. Wilson[1,2].* 1) Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada; 2) Department of Molecular Genetics, The University of Toronto, Toronto, Ontario, Canada; 3) International Laboratory for Research in Human Genomics, Universidad Nacional Autónoma de México, Juriquilla, Querétaro, México; 4) Department of Medical Biophysics, The University of Toronto, Toronto, Ontario, Canada; 5) Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada.

The acute phase response (APR) is a highly conserved response to noxious stimuli that is triggered by inflammatory cytokines, resulting in the rapid reprogramming of liver gene expression. There are over 60 secreted APR proteins in humans, defined clinically as any protein whose serum concentration changes by > 25% following inflammation. These "classic" APR proteins participate in several homeostatic processes including blood coagulation, innate immunity, and lipid metabolism and are often used as clinical markers of health and disease. However, relatively little is known about how transcription factors coordinate this extraordinary transcriptional response. To identify gene regulatory mechanisms that drive APR gene expression, we first defined co-regulated modules using mouse liver gene expression datasets obtained from over 30 TF knockout studies. We found modules of APR genes that were regulated by distinct sets of TFs. These APR modules were enriched for specific biological functions and allowed us to make predictions for additional TFs involved in the APR. For example, genes involved in blood clotting and complement were enriched for motifs of canonical liver transcription factors includingHNF1A and ONECUT1 whereas immune response genes were enriched for STAT2 and IRF motifs. Given that the APR is highly conserved in vertebrates, we performed cross-species (human, macaque, dog, mouse rat) analysis of liver enriched TFs to identify shared and species-specific APR gene regulatory regions. We found common and rare genetic variants within conserved APR regulatory regions that are predicted to affect TF binding. We looked for variants representing cross-species genetic differences whose predicted effects were concordant with observed gain or loss of TF binding in other species. One such variant, rs72842444, is predicted to disrupt ONECUT1 binding and resides in a conserved regulatory region in the 3'UTR of *C2*and upstream of*CFB*, both of which are APR genes associated with reduced risk of age-related macular degeneration (ARMD14 [MIM 615489]). This variant is an ancestral macaque allele and corresponds with a loss of macaque ONECUT1 binding. Overall, our findings show that APR genes can be grouped into biologically meaningful clusters that are controlled by specific sets of liver expressed TFs. We propose that variants within conserved APR regulatory regions are important places to look for determinants of variability in the acute phase response.

3099F

Transcriptional and genetic regulatory response to anti-IL-6 treatment in system lupus erythematosus. *E. E. Davenport[1,2], M. Gutierrez-Arcelus[1,2], K. Slowikowski[1,2], J. S. Beebe[3], B. Zhang[3], M. Vincent[3], S. Raychaudhuri[1,2,4].* 1) Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA; 2) Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA; 3) Pfizer Inc. , Cambridge, MA, USA; 4) Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.

IL-6 is a pleiotropic cytokine and has a range of roles in immune regulation and inflammation. A dose-ranging study conducted in patients with systemic lupus erythematosus (SLE) has provided a unique opportunity to address the questions of response to therapy and identification of genetic variants underlying variation in this response. SLE is a systemic autoimmune disease characterized by the production of autoantibodies, which bind to host tissues. The joints, skin, lungs, blood vessels, kidneys and nervous system are frequently affected but the severity of disease, spectrum of clinical features and response to therapy varies widely between patients. In SLE, IL-6 is thought to play a role in B cell hyperactivity, maintenance of the auto-inflammatory loop and mediating tissue damage. Blocking the effect of IL-6 may improve SLE by reducing the amount of immunoglobulin, produced by hyperactive B cells, and reducing anti-dsDNA antibodies. 180 SLE patients were recruited to evaluate the efficacy and safety of an IL-6 monoclonal antibody, which binds and neutralizes IL-6. The phase II clinical trial randomized patients to placebo or three different doses of study drug (Wallace, D et al. Arth Rheum 66(12) 3529-3540, Dec 2014). The assigned dose of study drug was administered three times during the course of the study. Whole blood was collected at baseline and then at two further time points. Following globin depletion we generated high-depth RNA-seq data (up to 200M mapped reads) for a total of 480 samples. We have aligned data to the reference genome and quantified gene expression using Subread and featureCounts respectively. Following quality control, we will determine the response to drug through differential gene expression analysis using Limma. In addition, genotyping data are available for the individuals, which will allow the mapping of gene expression as a quantitative trait (eQTL). Comparing DNA variants with evidence of an association with gene expression before and after treatment will allow us to identify context specific regulatory variants. Determining the transcriptional response to anti-IL-6 will increase our understanding of the mechanism of action of IL-6 signaling. Integrating these results with the extensive clinical data collected, including outcome measures such as the SLE Disease Activity Index (SLEDAI)-2K score, will yield new insights into the subgroups of patients most likely to benefit from such treatment.

3100T

Aminoglycosides stress together with the mitochondria 12S rRNA 1494C>T mutation leads to mitophagy. *J. Yu[1], J. Zheng[1], P. Jiang[1], M. Guan[1,2].* 1) School of medical, institution of genetics, Zhejiang University, Hangzhou, Zhejiang Province, China; 2) Cincinnati Children's Hospital Medical Center, Human Genetics.

Aminoglycosides as modifying factors, modulated the phenotypic manifestation of mitochondrial rRNA mutations and the incomplete penetrance of hearing loss. In this report, using cybrids harboring the m. 1494C>T mutation as cell model, we showed that gentamycin aggravated mitochondrial dysfunction in a combination of the m. 1494C>T mutation. The m. 1494C>T mutation was responsible for the dramatic reduction in the mtDNA encoded proteins of H-strand, except for the MTND6 protein, accompanied with mild reduction of ATP production and increase in mitochondrial reactive oxygen species, compared with those of control cybrids. After exposure to gentamycin, further reduction of mitochondrial ATP production was observed in mutant cybrids with a marked decrease of the mitochondrial membrane potential. More excessive cellular reactive oxygen species was detected with stimulus of gentamycin than those in mutant cells. With gentamycin and m. 1494C>T stress together, more dysfunctional mitochondria were forced to fuse and exhibited mitophagy via up-regulated LC3-B, as a compensatory protective response to try to optimize mitochondrial function, rather than undergo apoptosis. These findings may provide valuable information to further understand of mechanistic link between mitochondrial rRNA mutation, toxicity of aminoglycosides and hearing loss.

3101F

RNA-sequencing in the DLPFC of schizophrenia patients and controls identifies novel differentially expressed transcripts originated in transposons. *F. Macciardi[1], G. Guffanti[2], S. Gaudi[3], J. Fallon[1].* 1) Psychiatry & Human Behavior, Univ California, Irvine, Irvine, CA; 2) Department of Psychiatry, Harvard Medical School, Boston, MA; 3) Department of Infectious, Parasitic and Immune-Mediated Diseases, Italian National Institute of Health, Rome, Italy.

Background: Genome-wide association analyses identified several genetic variants mapping to non-coding RNAs as risk factors of neuropsychiatric disorders, including schizophrenia (SZ). Functional studies focused on dysregulation of long-intergenic non-coding RNAs in neuro-developmental and neuro-degenerative disorders leading to evidence supporting a critical role of non-coding RNAs in brain transcriptome functional regulation. Recently, evidence is accumulating on the identification of brain-specific non-coding RNAs originated in transposons with potential cis-regulatory functions. **Methods:** We used RNA-sequencing to identify differentially expressed (DE) transcripts in 10 SZ patients and 10 healthy controls. RNA samples were isolated from Dorsal-Lateral Prefrontal Cortex, prepared for sequencing with Trueseq and sequenced on the Illumina HiSeq 2000. Raw FASTQ files were aligned to the human genome (hg19) using TOPHAT. Transcripts were determined using Trinity de novo assembly, aligned to the reference genome for sequence homology using BLASTn and functionally annotated using Trinotate. DE was determined using EdgeR with a false discovery rate (FDR) q-value of 0. 05 and a fold change (FC) of >2. **Results:** More than 790,000 transcripts were identified, of which 59,258 had FC>2 and 10 were also statistically significant after FDR correction. We initially focused on the characterization of transcripts that present sequence homology with mobile elements of type HERV. First, we identified a putative novel transcript that has sequence homology and protein homology with Zinc Finger 259 (ZNF259) as well as sequence homology with MSTA, the LTR of type ERVL-MaLR (Mammalian apparent LTR retrotransposon). The analysis of the sequence of ZNF259 did not reveal sequence homology with MSTA, which led to the hypothesis that our transcript is a novel chimaeric transcript derived from the fusion of HERV and Zinc finger protein. Second, we identified a lincRNA, whose second exon is derived from the sequence of ORF of HERVL-E. Interestingly, this transcript revealed sequence homology with additional 43 LTR copies with different genomic locations. **Conclusions:** We present here preliminary evidence of differential expression of HERV-derived chimaeric transcripts and non-coding RNAs in the postmortem dorsal-lateral-prefrontal cortex of schizophrenia patients and controls. Further analyses on the structure of these transcripts and their functional consequences are warranted.

3102T

Senp1: A Genetic Driver of Hypoxia-Induced Polycythemia. *P. Azad[1], H. ZHAO[1], P. CABRALES[1], R. RONEN[1], D. ZHOU[1], O. POULSEN[1], O. Appenzeller[1], H. Hsiao[1], V. BAFNA[1], G. HADDAD[1,2].* 1) UCSD, La Jolla, CA; 2) Rady's Children Hospital.

Senp1: A Genetic Driver of Hypoxia-Induced PolycythemiaPriti Azad 1, Huiwen W. Zhao 1, Pedro J. Cabrales 2 , Roy Ronen 3, Dan Zhou 1, Orit Poulsen 1, Otto Appenzeller 5, Yu Hsin Hsiao 1, Vineet Bafna 4 and Gabriel G Haddad 1,6,7. 1Division of Respiratory Medicine, Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093, USA2Department of Bioengineering University of California, San Diego La Jolla, CA 92093 USA. 3Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA4Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA 92093, USA5 Department of Neurology, New Mexico Health Enhancement and Marathon Clinics Research Foundation, Albuquerque, NM 87122, USA6Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093, USA7Rady Children's Hospital, San Diego, CA 92123, USACorresponding author:Gabriel G. Haddad: ghaddad@ucsd. edu**Since Polycythemia is a predominant trait in some high altitude dwellers but not others living at the same altitude in the Andes, we took advantage of this human experiment in nature and studied both populations (with and without Chronic Mountain Sickness or CMS). We generated a disease in-the dish-model of polycythemia by re-programming fibroblasts from CMS and non-CMS subjects. As compared to sea level controls and non-CMS subjects who respond to hypoxia by increasing their RBCs modestly or not at all, CMS cells increase theirs remarkably (up to 60 fold) with a dose-dependent response to hypoxia (1. 5, 5, 10% O2). Based on our previous whole genome sequencing of CMS and non-CMS subjects6, we knocked down SENP1 (a desumoylase) in CMS cells using lentiviral constructs and observed a striking reduction (>90%) of the CMS exuberant erythropoietic response to low O2. Furthermore, VEGF, GATA1 and BclXL increase in gene expression in the CMS RBCs, in contrast to the other two populations. By utilizing a SUMO-GATA1 fused construct, we show that GATA1 desumoylation, a target of SENP1, is critical for the CMS phenotype. We conclude that the increased erythropoietic sensitivity in CMS subjects is at least partly due to a gain of function of SENP1 which leads to the activation of GATA1.** .

3103F

Effects of hematopoietic stem cell transplant (HSCT) on the oral microbiome of patients with severe aplastic anemia. *N. Ames[1], J. Barb[2], H. Kim[3], A. Oler[4], N. Chalmers[5], A. Ranucci[1], R. Childs[6], A. Cashion[3], G. Wallen[1].* 1) Clinical Center Nursing Department, National Institutes of Health, Bethesda, MD; 2) Mathematical and Statistical Computing Lab, Center for Information Technology, National Institutes of Health, Bethesda, MD; 3) National Institute of Nursing Research, National Institutes of Health Bethesda, MD; 4) National Institute of Allergy and Infectious Diseases, Bioinformatics and Computational Biosciences Branch, Office of Cyber Infrastructure and Computational Biology, Bethesda, MD; 5) National Institute of Dental and Craniofacial Research, National Institutes of Health Bethesda, MD; 6) National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD.

The human microbiome, especially the oral microbiota, might be the key to understanding infections and other perturbations that assault the human body. Patients with severe aplastic anemia undergoing hematopoietic stem cell transplant (HSCT) experience substantial immune system changes as a result of conditioning for HSCT and reconstituting a new immune system. To investigate the role of the oral microbiome during this process, we collected tongue brushings at baseline prior to HSCT, at engraftment and 100 days post-transplant from 4 severe aplastic anemia patients. We hypothesized that modulation of the immune system, the effects of HSCT with its inherent use of obligate antibiotics, would impact bacterial diversity. Bacterial DNA was subjected to next-generation sequencing using Ion Torrent Personal Genome Machine 400-bp kit and 316 v2 chip and the new Ion 16S Metagenomic Kit ™ (Life Technologies) targeting multiple variable regions of the 16S rRNA gene, specifically variable (V) regions, V2, V3, V4, V67, V8 and V9. We developed a novel analytical pipeline and then applied a published processing pipeline specific to Ion Torrent data to compare results across the 6 different regions, using a combination of tools from Mothur, UPARSE and QIIME. Of the 12 samples, two appeared to be outliers and were excluded from further analysis. Proportions of oral bacteria differed based on the variable region. The tongue microbiome during engraftment demonstrated a striking lack of diversity in 3 out of 4 patients. Indicators of microbial diversity demonstrated a trend where diversity decreased at engraftment from baseline and then increased 100 days post-transplant for all 5 variable regions. The genus *Rothia* was found in large proportions in the engraftment samples in most variable regions but markedly less in V3. In all 10 samples, the number of operational taxonomic units (OTUs) present across the 3 groups differed significantly. The post-transplant 100 day samples had fewer OTUs than at baseline but more than in the engraftment samples. With this novel analytical approach, we were able to separate reads into their respective regions without knowing primer sequences and thus were able to show distinctive changes in the oral microbiome before and after HSCT over 5 different hypervariable regions. Our next goal is to use this approach to correlate interventions during HSCT with the changes identified in the oral microbiome.

3104T

Analysis of mucosal adherent 16S rRNA reveals altered microbial composition and decreased diversity in patients with Crohn's disease. *N. M. Shahir[1], I. M. Carroll[3], D. N. Frank[5], R. D. Newberry[6], R. B. Sartor[2,3,4], H. H. Herfarth[3], R. Rahbar[3], T. S. Sadiq[7], M. J. Koruda[3], S. Z. Sheikh[3,9], T. S. Furey[8,9].* 1) Curriculum in Bioinformatics and Computational Biology, Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; 2) Center for Gastrointestinal Biology and Disease, University of North Carolina at Chapel Hill, Chapel Hill, NC; 3) Department of Medicine, Division of Gastroenterology and Hepatology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 4) Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC; 5) Department of Infectious Diseases, University of Colorado School of Medicine, Aurora, CO; 6) Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO; 7) University of North Carolina School of Medicine, Chapel Hill, North Carolina; 8) Departments of Genetics and Biology, Carolina Center for Genome Sciences, Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC; 9) Equal Contribution.

Crohn's disease (CD) results from an inappropriate immune response to the resident intestinal microbiota. We hypothesize that there exist unique microbial communities in the intestine of patient with CD compared to non-IBD controls. Previous studies have focused on examining the intestinal microbiota using fecal samples, but it is unclear how well these reflect the more physiologically relevant microbial population of mucosally-adherent intestinal bacteria. For our study, we collected mucosal samples from macroscopically non-inflamed and inflamed sections of resected colon and small intestine (ileum) from CD and non-CD patients. For each sample, we performed 16S sequencing of the V1 and V2 regions of the ribosomal RNA of intestinal bacteria, a primary component of the microbiota. Using QIIME, we determined operational taxonomic units (OTU) to classify the bacteria compositions for each sample, and compared the mean bacterial species diversity (α-diversity) between disease phenotypes and between individuals from the same disease phenotype (β-diversity). Overall, we found the α-diversity was significantly decreased in both the ileum and colon in CD patients compared to non-IBD controls. There was no significant change in α-diversity in the ileum due to inflammation within CD patients. We clustered a subset of our CD patients, where we had samples both from inflamed and non-inflamed regions, based on the β-diversity using weighted Unifrac metric. We found that the matched non-inflamed and inflamed samples from the same CD patients clustered with themselves as opposed to by inflammation. This suggests that merely the presence of disease appears to alter microbial compositions across the region as opposed to localizing where the inflammation exists. While the α-diversity exhibited no significant change across inflammation status, we did note changes in the genus of bacteria seen in inflamed versus non-inflamed ileum such as an increase of bacteria in the *Enterococcus* and *Clostridium* genus.

3105F

Role of Environment and Diet on Microbiome Composition in Diverse Rural Africans. *M. A. Rubel[1], M. E. B. Hansen[2], A. Bailey[3], K. Bittinger[3], A. Laughlin[3], W. Beggs[2], A. Ranciaro[2], S. Thompson[2], F. D. Bushman[3], S. A. Tishkoff[2].* 1) Department of Anthropology, University of Pennsylvania, Philadelphia, PA; 2) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 3) Department of Microbiology, University of Pennsylvania, Philadelphia, PA.

Given their antiquity, African populations may harbor unique physiological adaptations to their environments and diets, including diverse gut microbiomes. To address fundamental questions about sources of variation that can influence the microbial composition of understudied Africans, fecal DNA was extracted from sixty Tanzanians practicing a range of subsistence strategies (Hadza hunter-gatherers, Burunge agriculturalists, Maasai pastoralists, and recently settled Sandawe hunter-gatherers). The bacterial 16S rRNA V1-V2 amplicons were sequenced on the Illumina MiSeq platform. Bacterial communities were characterized by presence, absence, and changes in relative abundance of bacterial taxa. This data was paired with nutritional and ethnographic surveys as well as genetic data from an Illumina 5M Omni SNP Array. Bacterial compositional analysis between and within our populations revealed that the Hadza have a lower within-group diversity than the other populations. Principal coordinate analysis of abundant bacterial families showed that Tanzanians had two predominant gradients of bacteria associated with broad global enterotypes: A strong Prevotellaceae to Ruminococcaceae gradient and a separate, weak Bacteroidales to Ruminococcaceae gradient. Bacteroidales is associated with diets high in protein and fats, whereas Prevotellaceae and Ruminococcaceae are associated with diets rich in plants and fiber. This result is consistent with what we would expect from populations with rural, nonwestern diets. Mantel tests indicated that the presence or absence of bacteria in our populations has a stronger geographic correlation than bacterial abundance, consistent with low abundance taxa having strong regional variation. This represents one of the largest microbiome studies to date of ethnically diverse Africans with a wide range of subsistence strategies living indigenous lifestyles and provides novel microbiome data from sparsely characterized African groups.

3106T

ABO antigen and secretor status are not associated with gut microbiota composition. *E. R. Davenport[1], J. K. Goodrich[1,2], J. T. Bell[3], T. D. Spector[3], R. E. Ley[1,2], A. G. Clark[1].* 1) Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA; 2) Department of Microbiology, Cornell University, Ithaca, NY, USA; 3) Department of Twin Research and Genetic Epidemiology, King's College London, UK.

Microbes in the gut interact with their host through mechanisms that may include cell surface and secreted human antigens. One such potential antigen is encoded by the *ABO* gene and is classically known as the major red blood cell histo-incompatibility molecule for blood transfusions. This antigen is not only found on the surface of red blood cells, but is also secreted from epithelial and endothelial cells in individuals with an intact *FUT2* gene ("secretors"). In addition to transfusion medicine, ABO and secretor status play important roles in several diseases, including inflammatory bowel, cardiovascular, and infectious disease risks. It is unclear, however, whether inter-individual variation in the expression of these antigens in the gut affects the microbiota, and in turn, whether the microbiota mediates risk for diseases known to be associated with ABO or secretor status. In this study, we examined the relationship between ABO, secretor status, and the composition of the gut microbiome in a cohort of >900 individuals from the United Kingdom Adult Twin Registry (TwinsUK). 16S rRNA gene sequencing was used to characterize microbial composition in stool for each individual, utilizing Illumina MiSeq technology. First, we sought to identify broad trends in the microbiome with ABO status, secretor status, and the interaction of ABO by secretor status by performing principal components analysis, discriminant analysis of principal components, and principal coordinates analysis of Bray-Curtis and Unifrac indices. These methods applied to both presence/absence data and bacterial abundance measurements of the common operational taxonomic units (OTUs) failed to identify clustering by ABO or secretor status. Next, we applied linear mixed models with terms to account for the relatedness among twins to identify bacterial OTU or taxa abundances that were associated with ABO status or secretor status. After correcting for multiple tests using *q*-values, zero associations were identified even at a lenient $q <= 0.1$ significance threshold. Finally, we applied a random forests classification scheme to attempt to classify ABO or secretor status in individuals using the abundances of common bacterial taxa. The resulting classification error rates for each model were high (>50% error). These results from a large cohort demonstrate that despite previous reports, the taxonomic composition of the microbiota does not appear to be associated with ABO or secretor status.

3107F

Rapid growth between birth and two years is associated with lower diversity in oral and gut microbiomes. *S. J. C. Craig[1,2], D. Blankenberg[3], D. Bouvier[3], M. E. Marini[4], J. S. Beiler[5], N. Verdiglione[5], J. S. Williams[4,6], L. L. Birch[7], F. Chiaromonte[2,8], A. Nekrutenko[3], I. M. Paul[2,9], K. D. Makova[1,2].* 1) Department of Biology, Penn State University, University Park, PA; 2) Center for Medical Genomics, Penn State University, University Park, PA; 3) Department of Biochemistry and Molecular Biology, Penn State University, University Park, PA; 4) Center for Childhood Obesity Research, Penn State University, University Park, PA; 5) Pediatrics Support Services, Penn State University, Milton S. Hershey Medical Center, Hershey, PA; 6) Department of Nutritional Sciences, Penn State University, University Park, PA; 7) Department of Foods and Nutrition, University of Georgia, Athens, GA; 8) Department of Statistics, Penn State University, University Park, PA; 9) Department of Pediatrics, Penn State College of Medicine, Hershey, PA.

Rapid infant growth, commonly defined as upward crossing of two major centile lines on pediatric weight-for-length growth charts, has emerged as one of the strongest risk factors for childhood obesity. The human microbiome, a collection of organisms that form distinct communities on different areas of the human body, is also emerging as a causative and predictive environmental factor to the development of obesity. The initial composition of these microbial communities reflects various perinatal factors and changes drastically over the first years of life. Therefore, the characterization of these communities and their development is essential for a more complete understanding of the etiology of childhood obesity. Changes in the gut microbiome in early life are influenced by diet, mode of delivery, and medication use. However, these differences have not been demonstrated using an unbiased technology in a large and phenotypically well-characterized study cohort. Furthermore, no studies have investigated the oral microbiome in conjunction with the gut microbiome as an additional potential influence on childhood obesity, or the relationship between the gut and oral microbiomes. In this study, we have used 16S sequencing to survey the gut and oral microbiomes of 125 two-year-old children for whom extensive social, behavioral, and phenotypic data were collected. We found that children with rapid weight growth within the first two years after birth have a lower diversity in both their oral and gut microbiomes than those children who do not have rapid weight growth. This indicates that there is a relationship between rapid weight growth rate in the first two years and the establishment of a perturbed microbiome, which is likely influenced by developed lifestyle and behavioral patterns.

3108T

3D-NOME: an integrated **3-D**imensional **N**ucle**O**me **M**odeling **E**ngine for data-driven simulation of spatial genome organization. D. Plewczynski1,2,3, P. Szalaj3,4, Z. Tang2, P. Michalski2, M. Pi-tal1, O. Luo2, X. Li2, Y. Ruan2. 1) Centre of New Technologies, University of Warsaw, 00-927 Warsaw, Poland; 2) The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA; 3) Centre for Innovative Research, Medical University of Bialystok, Bialystok, Poland; 4) I-BioStat, Hasselt University, Belgium.

Background: Chromosomal folding are important features of genome organization, which play critical roles in genome functions, including transcriptional regulation. Using 3C-based mapping technologies to render long-range chromatin interactions has started to reveal some basic principles of spatial genome organization. Among 3D genome mapping technologies, ChIA-PET is unique in its ability to generate multiple datasets (in a single experiment), including binding sites, enriched chromatin interactions (mediated by specific protein factors), as well as non-enriched interactions that reflect topological neighborhoods of higher-order associations. The multifarious nature of ChIA-PET data represents an important advantage in capturing multi-layer structural-functional information, but also imposes new challenges in multi-scale modeling of 3D genome. **Results:** We report the development of 3D-NOME (*3-Dimensional* **N***ucle***O***me* **M***odeling* **E***ngine*), for processing ChIA-PET data, which includes four components: 1) identifying interactions from ChIA-PET reads, 2) multidimensional scaling on proximity graph distances for rapid low-resolution structural inference, 3) multiscale Monte Carlo algorithm for high-resolution simulations, and 4) web-based visualizing tool. Using ChIA-PET data of human B-lymphocytes (GM12878 cell line), we demonstrate the effectiveness of 3D-NOME in building 3D genome models at multiple levels, including the entire nucleome, individual chromosomes, specific segments at megabase (Mb) and kilobase (kb) resolutions. We also consider CTCF-motif orientations and looping patterns in order to achieve more reliable structures. **Conclusions:** 3D-NOME is an integrative computational pipeline that is effective in processing the multifaceted ChIA-PET data for simulation of 3D genome from higher-order chromosome folding to detailed functional interactions mediated by protein factors. Further refinement of 3D-NOME and application to additional ChIA-PET and other types of 3D genome mapping data will help to advance our understanding of genome structures and functions.

3109F

***In vitro* and *in vivo* exploration of regulatory variation in melanocytes.** *M. B. Baker[1], D. U. Gorkin[1], D. Lee[1], M. A. Beer[1,2], A. S. McCallion[1].* 1) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland, United States of America; 2) Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, United States of America.

Recent studies have highlighted the importance of non-coding variation in disease, with conservative estimates suggesting that at least 40% of SNPs implicated in disease by GWAS fall into haplotype blocks devoid of coding sequence. Predicting the impact of genetic variation in the non-coding space of the genome has remained elusive. To that end, we developed a computational classifier to estimate the impact of variation on regulatory sequences (presented by Dongwon Lee). The classifier was trained on a set of putative melanocyte enhancers defined by the intersection of EP300 and H3K4me1 ChIP-seq signals in melanocytes. To validate the classifier, we initially selected two well-characterized melanocyte enhancers for our analyses, using the classifier to predict the functional consequences of all possible nucleotide substitutions at every site along their length. We then selected a subset of variants predicted to have positive, negative, or no effect to test functionally *in vitro* using a luciferase reporter. We observed a high correlation of the reporter expression with our classifier's predictions. We recognize that *in vitro* assays are inherently synthetic, removing candidate sequences from their endogenous context and evaluating them in mini-gene constructs that we introduce into cells. To begin to address this limitation, we have started to employ circularized chromatin conformation capture followed by sequencing (4C-seq), which allows us to infer physical proximity of sequences *in vivo*. Enhancers are thought to be brought into looping structures with the promoters they regulate; with 4C-seq, we capture putative enhancer sequences that physically interact with promoters of interest. We have generated a 4C-seq profile using the promoter of melanocyte-critical *Sox10* and are in the process of validating the contacts identified. To further explore the relationship of putative regulatory sequences to one another and gene expression in hopes of better understanding the role of non-coding variation in disease, we are employing CRISPR/Cas9 to selectively delete candidate sequences and assess the effect of their loss on looping and transcriptional output. .

3110T

Massive parallel sequencing revealed the conformational dynamics of the non-B form DNA at the promoter. *H. Inagaki[1], H. Miyamura[2], M. Tsutsumi[1], T. Kato[1], H. Nishizawa[2], H. Kurahashi[1].* 1) Div Mol Genet, ICMS, Fujita Health University, Toyoake, Aichi, Japan; 2) Dpt Obstet Gynecol, Fujita Health University School of Medicine, Toyoake, Aichi, Japan.

Alternate DNA structures such as non-B form DNAs cause the chromosome structural aberrations and change the gene expression. The connection between the polymorphisms within the annexin A5 gene (*ANXA5*) promoter and the obstetric complications including the recurrent pregnancy loss is highlighted in recent reports, although its mechanism is still unknown. Previously, we have reported that the polymorphisms in the *ANXA5* promoter affected the gene expression possibly through the predisposition for G-quadruplex (G4) structure. *In vivo* G4 formation can be detected by the sodium bisulfite treatment of the genomic DNA, which converts the cytosine to thymine specifically at the single-stranded region. In the present study, we analyzed the pattern of nucleotide changes in each strand to elucidate the G4 formation pattern in a single cell level. The genomic DNA from the placenta was treated with sodium bisulfite in a mild condition, not to dissociate the DNA strand during the treatment. This DNA was used for the PCR template and the *ANXA5* promoter region was amplified with the specific primers adding the adaptor sequences for the massive parallel sequencing at the 5' ends. About one to ten thousands sequences per sample was obtained by Illumina MiSeq sequencer. After the quality control, these sequences were assigned to each strand, i. e. upper strand or lower strand, according to the conversion of the nucleotides (C>T or G>A). The sequences were then clustered into groups according to the conversion pattern by using the MAFFT and other programs. The results indicated that the small amount (5-10%) of the molecules indeed formed the single-stranded form that is shown by the clusters of the converted nucleotides extended across the region ranging from 30 to 100 bp. We found two conversion patterns: one cluster at the complementary position of the G4 consensus region of the promoter indicating the G4 *in vivo*, and the other cluster extending to the downstream from the promoter indicating the signature of mRNA transcription initiation. These patterns were slightly different between normal allele and decreased expression allele (M2), suggesting the G4 formation may contribute the upregulation of the gene expression. The overall patterns of the single-stranded region may represent the temporal change of the DNA structure *in vivo*, and this technique combined with the NGS technology will shed light on the dynamics of DNA conformation in the cell.

3111F

Translational correlation of MeCP2 binding dynamics and clinical presentation of male patients with missense mutations. *T. I. Sheikh[1,2,7], J. Ausio[3], H. Faghfoury[4], J. Silver[4], A. Percy[5], J. B. Lane[5], P. MacLeod[6], J. B. Vincent[1,2,7].* 1) Molecular Neuropsychiatry & Development (MiND) Lab, Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada; 2) Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada; 3) Department of Biochemistry and Microbiology, University of Victoria, BC, Canada; 4) The Fred A. Litwin Family Centre in Genetic Medicine, University Health Network and Mount Sinai Hospital, Toronto, ON, Canada; 5) Civitan International Research Center, University of Alabama at Birmingham, AL, USA; 6) Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada; 7) Department of Psychiatry, University of Toronto, Toronto, ON, Canada.

The Methyl CpG Binding Protein 2 (MeCP2) binds to methylated DNA through its methyl binding domain (MBD) to regulate gene expression. Mutations in the*MECP2*gene are the main cause of Rett syndrome (RTT). In heterozygous females the variable severity of phenotypic conditions is modulated by non-random X inactivation, thus making phenotype/genotype comparisons unreliable. However phenotype/genotype correlations in males with hemizygous*MECP2*mutations may provide more accurate insights into the true biological effect of specific mutations. The functional effects of different MeCP2 mutants have been studied extensively, but reports of direct correlation of functional effects of missense mutations in the hemizygous males are extremely rare. Here, we present a rational comparison of 1) the effect on binding properties of various common and rare MeCP2 MBD mutants and their consequences on overall chromatin organization; 2) the distinct mobility dynamics and binding patterns of the mutant MeCP2 proteins; and 3) phenotypic comparison for MeCP2-variant hemizygous males. We found a complete absence of MeCP2-DNA interaction for mutations on or close to the MBD-methylated DNA binding interface such as Arg111Gly, Arg133Leu, Thr158Met, and Arg106Trp which is present on the β-sheet site directing towards the MBD-DNA binding interface. On the other hand, Pro152Arg and two novel mutations in male Pro152His and Asn126Ile showed chromatin clustering defects. Furthermore, binding and mobility dynamics of Arg111G, Thr158Met, Pro152Arg, Pro152His and Pro152A show a gradient of impairment (severe to mild), depending on the physical and chemical properties of the substituting amino acid. All these changes are strictly site specific in the tertiary structure of MeCP2. Interestingly, a similar wide range of phenotype/clinical severity, ranging from neonatal encephalopathy to mild psychiatric abnormalities, consistent with the functional severity were observed, i. e. Thr158Met> Phe157Ile > Asn126Ile>Arg-133Cys> Pro152His > Arg167Trp> Pro152Ala >wild type, which show a direct phenotype genotype correlation in the MeCP2 hemizygous male patients.

3112T

Topoisomerase II beta interacts with Ctcf and the cohesin complex at evolutionarily conserved points of genome control. *L. Uusküla-Reimand[1,2], H. Hou[1,3], P. Samavarchi-Tehrani[4], M. Vietri Rudan[5], M. Liang[1,3], A. Medina-Rivera[1,11], H. Mohammed[8,12], E. Young[1], D. Schmidt[9,12], P. Schwalie[6,10], J. Reimand[7], S. Hadjur[5], AC. Gingras[3,4], M. Wilson[1,3].* 1) Genetics and Genome Biology Program, SickKids Research Institute, Toronto, ON, Canada; 2) Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia; 3) Department of Molecular Genetics, University of Toronto, Canada; 4) Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada; 5) Research Department of Cancer Biology, Cancer Institute, University College London, London, UK; 6) European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom; 7) The Donnelly Centre, University of Toronto, Toronto, ON, Canada; 8) Present address: The Babraham Institute, Cambridge, United Kingdom; 9) Present address: Syncona Partners LLP, London, United Kingdom; 10) Present address: Laboratory of Systems Biology and Genetics, Lausanne, Switzerland; 11) Present address: International Laboratory for Research in Human Genomics, Universidad Nacional Autónoma de México, Juriquilla, Querétaro, México; 12) Cancer Research UK, Cambridge Institute, University of Cambridge, Cambridge, United Kingdom.

Type II DNA topoisomerases are critical enzymes that regulate DNA topology by generating double stranded DNA breaks. Topoisomerase II beta (Top2b) is required for vertebrate development and has been implicated in treatment related cancers and chemotherapeutic side effects. Characterizing the protein interaction networks of TOP2B and genome-wide occupancy is required for understanding its role in vertebrate development and the mechanisms behind off-target effects in chemotherapy. Using proteomic analysis (BioID), chromatin immunoprecipitation (ChIP), and high-throughput chromosome conformation capture (Hi-C) we detect proximal TOP2B protein interactions *in vitro* and characterize the genomic landscape of Top2b binding *in vivo*. We find that TOP2B interacts with various chromatin regulators and localizes to active chromatin regions that show hallmarks of function including DNAse I hypersensitivity and conserved transcription factor binding. Specifically, TOP2B interacts with CTCF and several members of the cohesin complex. Top2b/Ctcf/cohesin proteins bind chromatin with a distinct spatial organization and are enriched at the boundaries of chromosomal domains. Our findings show that TOP2B is robustly positioned to resolve topological constraints at critical points of genome control and suggest a novel mechanism by which Top2b-mediated double strand breaks could impact human disease and treatment related phenotypes.

3113F

Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *K. D. Hansen[1,2], J. P. Fortin[1].* 1) Dept. of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA; 2) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD, USA.

Introduction: Analysis of Hi-C data has shown that the genome can be divided into two compartments called A/B compartments. These compartments are cell-type specific and are associated with open and closed chromatin. Existing Hi-C experiments have largely been performed in cell lines or for samples where unlimited input material is available. Here we show that A/B compartments can be reliable estimated using epigenetic data from several different platforms: the Illumina 450k DNA methylation microarray, DNase hypersensitivity sequencing, single-cell ATAC sequencing and single-cell whole-genome bisulfite sequencing. This work makes A/B compartments readily available in a wide variety of cell types, including many human cancers. Results: We first estimated A/B compartments from two Illumina 450k Methylation array datasets (Fibroblasts and EBV-transformed lymphoblastoid cell lines) by exploiting the fact that correlations of methylation levels between two regions located in the closed compartment are substantially higher than between other regions. At resolution 100kb, we showed good correlation with A/B compartments estimated from public Hi-C experiments performed on the GM12878 and IMR90 cell lines. Next, we showed that A/B compartments can be estimated similarly from DNase hypersensitivity sequencing by using two publicly available datasets: 70 HapMap samples and 4 replicates of the IMR90 cell line. Remarkably, our method can be further applied to single-cell epigenetic data to study within-individual genome compartmentation by using a publicly available single-cell ATAC sequencing experiment and a single-cell whole-genome bisulfite sequencing experiment. As an valuable application, we estimated A/B compartments for 11 cancer types from The Cancer Genome Atlas (TCGA) by using the publicly available 450k methylation datasets. We emphasize that TCGA does not included assays measuring chromatin accessibility such as DNase or various histone modifications, therefore making our estimated compartments a valuable data source for the community. Conclusion: This work makes A/B compartments readily available in a wide variety of cell types, including many human cancers. Our reliable estimation of the genome compartmentation from single-cell epigenetic data potentially allows our method to be used on rare types of cells.

3114T

High Quality Next-Generation Sequencing Libraries for Sub-Nanogram Amounts of ChIP DNA. *C. Couture[1], J. Irish[1], A. McUsic[1], J. Laliberte[1], L. Kurihara[1], R. Spurbeck[1], C. Schumacher[1], V. Makarov[1], D. Langlais[2], G. D. Gilfillan[3].* 1) Swift Biosciences, Ann Arbor, USA; 2) McGill University, Montreal, Canada; 3) Norwegian High Throughput Sequencing Centre, Oslo, Norway.

ChIP-Seq is widely used to produce genome-wide profiles for chromatin proteins and transcription factors. However, ChIP DNA samples are inefficiently converted into next-generation sequencing (NGS) libraries due to the limited quantity and quality of DNA samples following de-crosslinking, especially when only low cell numbers are available (patient samples, biopsies). To overcome these limitations, we developed a library preparation method that utilizes as little as 100 pg of ChIP DNA to generate highly complex NGS libraries with exceptional coverage of GC-rich regions, thus reducing the sequencing depth requirements. To validate the range of input levels required for comprehensive detection of precipitated regions using our method, we generated libraries from two ChIP DNA amounts using different master ChIP samples made with antibodies against the interferon response regulator IRF8, the hematopoietic transcription factor PU. 1, or the histone modification H3K4me3. ChIP DNA was quantified by Bioanalyzer and 1 ng or 100 pg were used to prepare a series of NGS libraries that were sequenced on the Illumina®platforms. In all three ChIP samples, when correcting for sequencing depth, we observed the same number of binding sites for samples prepared from 1 ng or 100 pg with about 90% overlapping peaks between the two different inputs suggesting that 100 pg of starting material provides the same profile as 1 ng. The peaks that were exclusive to one or the other input had low quality peak scores suggesting low confidence peaks just above the background threshold. To further analyze PU. 1, we performed deep sequencing and analyzed the number of peaks as a function of the number of reads. The 1 ng and 100 pg rarefaction curves were similar and reached the maximum at the same depth of sequencing indicating preserved complexity even at low input. Our new library preparation method pushes the limits for low input ChIP samples. Our highly efficient method, coupled with the high complexity of the resulting libraries, generates high quality sequencing data and eliminates the need for deep sequencing or additional replicates to obtain confident peak calls. Our proven library preparation method also supports inputs as low as 10 pg for whole genome sequencing and doesn't require precise input quantification (a limitation of other library preparation protocols). We are currently working to further reduce our ChIP input requirement to 10 pg.

3115F

Whole genome sequencing identifies complex and balanced *de novo*structural variation in autism. *D. Antaki[1,2,3,4], W. M Brandler[1,2,3,4], M. Gujral[1,2,3,4], A. Noor[1,2,3,4], G. Rosanio[1,2,3,4], D. J Barrera[1,2,3,4], T. C Chapman[1,2,3,4], G. N Lin[1,3], D. Malhotra[1,2,3], A. C Watts[1], L. C Wong[1,9], J. A Estabillo[9], T. E Gafomski[2,4], O. Hong[2,3,4], A. Bhandari[2,3,4], R. Owen[10], A. Gore[5], G. Reiner[6], K. K Vaux[6], C. M Strom[10], K. Zhang[1,5], K. Zhang[1,8], A. R Muotri[1,4], N. Akshoomoff[9], S. M Leal[11], K. Pierce[7], E. Courchesne[1,7], L. M Iakoucheva[3], C. Corsello[9], J. Sebat[1,2,3,4].* 1) University of California, San Diego, La Jolla, CA; 2) Beyster Center for Genomics of Psychiatric Diseases, University of California San Diego, La Jolla, California 92093; 3) Department of Psychiatry, University of California San Diego, La Jolla, California 92093, USA; 4) Department of Cellular and Molecular Medicine, University of California San Diego, La Jolla, California 92093, USA; 5) Department of Bioengineering, University of California San Diego, La Jolla, California 92093, USA; 6) Department of Medicine, University of California San Diego, La Jolla, California 92093, USA; 7) Department of Neuroscience, University of California San Diego, La Jolla, California 92093, USA; 8) Department of Ophthalmology, University of California San Diego, La Jolla, California 92093, USA; 9) Rady Children's Hospital, San Diego, California 92123, USA; 10) Quest Diagnostics Nichols Institute, San Juan Capistrano, California 92675, USA; 11) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, Texas 77030, USA.

Whole Genome Sequencing (WGS) provides a more complete ascertainment of de novo mutation compared to microarrays or target sequencing. An exhaustive investigation of structural variation in a large ASD cohort has not yet been reported. We investigated the contributions of de novo structural variants including copy number variants (CNVs), transposable elements, and complex structural rearrangements to autism spectrum disorder (ASD) by WGS and custom mutation detection in 235 subjects, including 71 with ASD, 26 sibling controls and their parents. Deletions, Duplications, Inversions and a variety of complex structural rearrangements were detected using a custom SV calling pipeline, and de novo mutations were detected using a Gaussian Mixture Model-based SV genotyper. A high rate of de novo copy number variation was observed in controls (15% of subjects), revising previous estimates. CNVs in ASD cases were enriched for genes that are mutated in independent studies of neurodevelopmental disorders. Recurrent mutations detected here and in independent studies implicate specific genes in ASD, including of CDH8, TMEM185A and CACNG2. WGS captures a wide spectrum of human mutation and promises to accelerate genetic discovery in ASD.

3116T

The copy number variation (CNV) database for autism spectrum disorder (ASD). *X. Liu[1], J. Nomura[1,2], Y. Yamaguchi[3], T. Takumi[1].* 1) Mental Biology Team, RIKEN Brain Science Institute , Wako, Saitama , Japan; 2) Graduate School of Biomedical Science, Hiroshima University, Hiroshima, Japan; 3) Neuroinformatics Japan Center (NIJC), RIKEN Brain Science Institute, Wako, Saitama, Japan.

Copy number variations (CNVs) are defined as duplication or deletion of a section of DNA with a length usually from 1 kb to several Mb. They are widespread in the human genome and account for a substantial proportion of genetic variation. Recent studies have demonstrated that CNVs are critical pathogenetic factors to autism spectrum disorder (ASD) - a severe early onset nerurodevelopmental disorder characterized by social impairments and communication difficulties. With the accumulating CNV data, it becomes urgent to build an up-to-date and comprehensive, accurate and well-annotated, flexible, yet easy to use database for both research and clinical settings. Here, we present an ASD CNV database based on manually curated information from literatures and Japanese data from our own study. In addition, we adopt a novel approach to visualize complex CNV data and also provide human to mouse synteny maps for all CNVs, which is particularly relevant to animal studies. In summary, this database is a useful resource to tackle the genetic complexity of ASD and also may help clinicians for better diagnoses and care for ASD individuals.

3117F

Building a dosage map of the genome to assist in CNV interpretation. *A. E. Hare[1], B. R. King[2], E. L. Easse[2], D. H. Ledbetter[1], C. L. Martin[1].* 1) Autism & Developmental Medicine Institute, Geisinger Health System, Lewisburg, PA; 2) Computer Science Dept. , Bucknell University, Lewisburg, PA.

In order to assess the pathogenicity of copy number variants (CNVs) in developmental brain disorders (DBD), we constructed dosage maps of CNVs (>1 kb) and compared shared and unique regions of variation in the genome between cases and controls. While previous similar maps have been constructed by merging overlapping CNVs, we employed a different approach utilizing CNV frequency to identify regions of variation. Each map was constructed by quantifying the number of dels/dups at each basepair along a chromosome and joining adjacent nucleotides that have the same number of variants (+/-5) into variant blocks. In controls, we created maps of 1,075,110 deletions (dels) and 130,487 duplications (dups) from 23,842 controls in the Database of Genomic Variants (DGV; October 16, 2014 release date). In a control population, blocks with a high number of CNVs are predicted to be dosage tolerant (DT) and not likely to confer risk for disease, while blocks with little to no CNVs are potentially dosage sensitive (DS). For the control deletion map, we defined the DS cut-off as <150 dels (corresponding to a population frequency threshold of <0. 01% in DGV) and DT as ≥150 dels by using a ROC curve analysis based on known haploinsufficient regions or genes curated as part of the ClinGen project. In controls, 0. 89% of the genome is DT and dels in these regions (avg size = 16. 7 kb) are predicted to be benign. Similarly, using triplosensitivity data from ClinGen for our control duplication map we found that 0. 97% of the genome is DT and dups in these regions (avg size = 55 kb) are also predicted to be benign. For cases, we constructed maps of 10,987 CNVs from 13 studies of DBD cases in dbVar with clinical assertions of uncertain or likely pathogenic/pathogenic. Genomic regions with a high number of CNVs in cases and not in controls are predicted to be DS. In cases, pathogenic dels encompass 11. 0% of the genome, while pathogenic dups span 24. 7% of the genome. Uncertain dels cover 3. 61% of the genome and uncertain dups span 0. 99% of the genome. For dels and dups in the pathogenic and uncertain maps, there was <1. 0% overlap with DT regions defined in the control maps; therefore we were able to determine that these regions are DS. By comparing the overlap of novel CNVs to DT regions in controls and DS regions in cases, the dosage maps presented here can assist in the assessment of variant pathogenicity to inform clinical interpretations.

3118T

Mapping the "Dark Matter" of the Genome – Complex Structural Variations and Towards True Contiguity of *de novo* Assembly with NanoChannel Technology. *H. Cao[1], A. Hastie[1], E. T. Lam[1], A. Pang[1], W. Andrews[1], T. Anantharaman[1], T. Chan[1], X. Zhou[1], J. Reifenberger[1], H. Saghbini[1], M. Austin[1], P. Sheth[1], Z. Y. Zyu[1], Z. Dzakula[1], X. Xun[2], P. Kwok[3].* 1) BioNano Genomics Inc. , San Diego, CA; 2) BGI, Shenzhen, China; 3) UCSF, San Francisco, CA.

Large genomic structural variations (SV, >1 kb) known to be associated with complex traits and diseases are found to be more prevalent than previously thought. In spite of the advancements made in high-throughput short read next-generation sequencing, a fair portion of the human genome architecture is unresolved or ambiguously characterized as gaps and unknown structural or heterozygous information, also known as the "dark matter" of the genome. Rapid whole genome mapping using NanoChannel Arrays represents a new standard of single-molecule platform independent and yet complementary to DNA sequencing for accurate genome assembly and structural variation analysis. Extremely long intact DNA molecules of hundreds to thousands of kilobases fluorescently labeled at sequence motifs and linearized in true nanofluidic channels enable direct image interrogation of the genome's architecture at a high resolution. *De novo* assembly of these single molecules yields long contiguous genome maps, spanning highly repetitive regions and complex structures in their native form. Here we present results from analysis on human diploid and cancer genomes, non-model and large complex genomes. Two methods were employed either by the *de novo* genome mapping process or direct alignment of long raw molecules against a digitally digested and "barcoded" reference genome yielding detection of chromosomal abnormalities such as translocations, arm breakpoints or other lesions. Using ultra long molecules and the Irys® System, unknown SV events, such as chromothripsis, that often incorporate fragmented chromosome arms shattered during cancer transformation can be bridged. In these genomes, we detected haplotype differences and hundreds of large SVs revealing the locations, orientations, and copy numbers of complex SVs that are biologically and clinically relevant; and precisely mapped viral component integration sites within the human genome believed to be linked to genome instability and oncogenesis. Now, a population scale comparative whole genome study to identify comprehensive genomic SV on a single platform with a large patient cohort is feasible due to the standardized high quality data and automated efficiency of data generation enabled by NanoChannel technology.

3119F

Characterization of Common Maternal Copy Number Variations observed in Non-Invasive Prenatal Testing. *Y. Sun, E. Almasri, Y. Wu, A. R. Mazloom, M. Ehrich, T. Burcham, S. Kim.* Sequenom, Inc. , San Diego, CA.

Low coverage genome wide sequencing is a common method for non-invasive prenatal testing to detect fetal aneuploidies. Our method uses normalized sequencing reads to determine the local over- or under-representation of sequence counts, to infer fetal copy number variations (CNV). However, because circulating cell free DNA (ccfDNA) in maternal plasma is a mixture of maternal and fetal DNA, maternal CNVs can be detected by the same principle. Here, we present the population-level characterization of maternal CNVs identified within a research-consented de-identified cohort of over 51,000 samples. More than 40% of the samples in the cohort have at least one putative maternal CNV, with an average size of 145 Kb. These CNVs have significant overlap (p<0. 01, Fisher's exact test) with the regions reported in the Database of Genomic Variants (DGV, http://dgv. tcag. ca/) despite the latter involving a wide range of experimental methodologies, such as qPCR, aCGH, MassSpec and FISH for CNV identification across worldwide populations and with varying resolution. Specifically, we have identified 28 samples within our cohort that share the largest most frequent 2. 3 Mb CNV event on chromosome 8p23 which is consistent with the regions reported in DGV. Additionally, we have identified a total of 8 samples, in which copy number variations were located in regions of known cancer genes (EGFR, RB1 and SMARCB1). Furthermore, it is observed that deletion events are significantly lower in Guanine-Cytosine (GC) rich regions than in GC poor regions (p<0. 01, Wilcoxon rank sum test). As gene-rich regions are correlated to regions of high Guanine-Cytosine base composition, these results are suggestive of selective pressures against detrimental deletions. Overall, we analyzed the frequency and genomic distribution of maternal CNVs in a large cohort of pregnant women, and found many CNVs that are associated with various pathologies.

3120T

Copy Number Variation Associated With Bronchopulmonary Dysplasia. *A. Ahmad[1], S. Bhattacharya[1], T. J. Mariani[1], A. M. Iqbal[2].* 1) Department of Pediatrics, University of Rochester Medical Center, 601 Elmwood Ave. , Rochester - NY, 14642; 2) Department of Pathology and Laboratory Medicine, University of Rochester Medical Center, 601 Elmwood Ave. , Rochester- NY 14642.

Bronchopulmonary Dysplasia (BPD) is common in premature babies born before 29 weeks of gestational age and weighing <2. 2 lb. Variability in the incidence and severity of BPD among premature infants with similar risk factors suggests that genetic susceptibility plays a significant role in its pathogenesis. The goal of this study is to leverage existing clinical DNA microarray data repositories to identify CNV that are either predictive of, or associated with, BPD. We conducted a retrospective analysis of clinical DNA microarray data (n=2800) from our institution. We identified 19 subjects diagnosed with BPD, and two controls groups with no lung-related disease; full-term (n=41) and pre-term (n=23). DNA microarray experiments were previously performed on 4x44K Agilent platform and data were re-analyzed using Feature Extraction v10. 10 and Genomic Workbench v7. 0 software (Agilent Technologies, USA). To test for differences in the frequency of individual CNV between BPD and control groups, we used Fisher's Exact test. To test for development-associated changes in expression for genes residing in the CNV regions, we interrogated a previously published normal human lung development expression dataset. To test for pathways and processes represented by genes residing within CNV regions, we used Ingenuity pathways analysis tools. Among 19 subjects with confirmed diagnosis of BPD, three loci, at 11q13. 2, 16p13. 3. and 22q11. 23-q12. 1, were affected in more than one subject. The frequency of all three recurrent loci were significantly higher (p<0. 05) in BPD subjects as compared with at least one control group. We found fifteen of 21 genes (71. 4%) within the recurrent CNV loci demonstrated significant changes (p<0. 05) in expression between the pseudoglandular and canalicular stages of human lung development, approximately at the time of preterm birth for subjects with the greatest risk of BPD. Genes within the recurrent CNV loci were involved in various development- and disease-related processes. We also identified canonical pathways represented by the 21 genes, which included AHR signaling, xenobiotic metabolism signaling and glutathione-mediated detoxification. We have identified regions in the genome that are recurrently affected in subjects with BPD. Our data should be useful to aid in the identification of genetic factors associated with increased risk for chronic lung disease in preterm infants.

3121F

SMN2 copy number frequency in a sample of the Brazilian Population. *M. Goncalves, P. G. COUTO, L. B. ALVIM, T. SANTOS, N. P. LOPES, A. G. ASSINI, E. L. CONCENTINO, G. C. LOPES, N. REGINA, A. M. CASTRO.* INSTITUTO HERMES PARDINI, VESPASIANO, MG. , BRASIL.

INTRODUCTION: Spinal Muscular Atrophy (SMA) is one of the most common autosomal recessive neuromuscular disease and is usually caused by deletions of the survival motor neuron 1 (*SMN1*) gene. A closely related gene, *SMN2*, modifies the disease severity. Due to a single nucleotide polymorphism (840C>T), *SMN2* produces less full-length transcript than *SMN1* and cannot entirely prevent neuronal cell death at physiologic gene dosages. Determination of SMN2 copy number frequency have not been reliably estimated in Brazilians and is important to define, along with the age of onset and the severity of symptoms, the type of SMA (I, II, III or IV) when the individual do not present any copy of SMN1. **METHODS:** Genomic DNA was isolated from peripheral blood cells from 162 Brazilian individuals using Gentra Puregene Blood Kit (Qiagen). The DNA was quantified using Nano-View and the ratio 260/280 nm was measured to obtain the purity of DNA. The Multiplex ligation-dependent probe amplification (MLPA) for genetic testing of SMA was based on the commercially available SALSA MLPA kit P021 to determine *SMN1* and *SMN2* copy numbers. The electrophoresis capillary was carried out in an ABI3730 DNA analyzer. The 36cm capillary was used with POP7 polymer and the GeneScan 500 ROX dye Size Standard. **RESULTS:** Our sample consisted of 79 men (48. 8%) and 83 women (51. 2%). Of all 162 samples tested we found that 6/162 (3,7%) had 0 copies of *SMN2*, 15/162 (9,26%) had 1 copie of SMN2, 56/162 (34,57%) had 2 copies of *SMN2*, 54/162 (33,33%) had 3 copies os *SMN2*, 24/162 (14,81%) had 4 copies of *SMN2*, 6/162 (3,7%) had 5 copies of *SMN2* and 1/162 (0,62%) had 6 copies of *SMN2*. **CONCLUSIONS:** According to the literature, the *SMN2* copy number distribution in Brazilians were similar to Europeans that are more likely to harbor 2 copies of *SMN2* but its different from sub-Saharan Africans that are more likely to harbor 1 copies of the gene. This finding is important to consider in SMA genetic diagnostic in individuals with zero copies of *SMN1*, to determine the severity of the disease.

3122T

Characterization of NAHR-mediated type-1 *NF1* deletions with breakpoints within the PRS2 hotspot. *H. Kehrer-Sawatzki[1], M. Hillmer[1], D. Wagner[1], M. Daiber[1], V. -F. Mautner[2], D. N. Cooper[3], L. Messiaen[4].* 1) Institute of Human Genetics, University of Ulm, Ulm, Germany; 2) Department of Neurology, University Medical Centre, Hamburg-Eppendorf, Germany; 3) Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, UK; 4) Medical Genomics Laboratory, Department of Genetics, University of Alabama at Birmingham, USA.

NF1 microdeletions are observed in 5% of neurofibromatosis type 1 (NF1) patients. The majority of *NF1* microdeletions span 1. 4-Mb as determined by MLPA, the most commonly used detection technique. Most of the 1. 4-Mb deletions exhibit breakpoints located within the low-copy-repeats, NF1-REPa and NF1-REPc, which exhibit 97. 5% sequence homology within 51-kb. The 1. 4-Mb deletions with breakpoints in the NF1-REPs are designated “type-1 *NF1* deletions” and have been considered to be caused by non-allelic homologous recombination (NAHR). However, owing to the wide spacing of the MLPA probes, it is impossible to distinguish between type-1 *NF1* deletions mediated by NAHR and atypical *NF1* deletions not mediated by NAHR. In this study, we analysed 67 independent type-1 *NF1* deletions initially identified by MLPA. We performed SNP genotyping as well as breakpoint-spanning PCRs and found that 95% of *NF1* deletions designated type-1 according to MLPA were mediated by NAHR and hence are indeed type-1. Our study also confirms that PRS2, spanning 4-kb, is a very strong NAHR hotspot since it harbours 78% of all type-1 deletion breakpoints. We sequenced the breakpoint-spanning PCR products of 48 *NF1* deletions with breakpoints in PRS2 and noted considerable differences in terms of the location of the regions of strand exchange (ERs) between NF1-REPa and NF1-REPc. The ERs are flanked by sequence variants that allow NF1-REPa and NF1-REPc to be distinguished. Within the ERs (ranging in size from 50-bp to 975-bp), NF1-REPa and NF1-REPc exhibit complete sequence identity. In total, 13 different ERs were noted within a 4-kb region. This heterogeneity in ER position may be explicable in terms of the predicted positions of PRDM9 binding which in turn direct the location of the recombination-initiating DSB as well as the extent of the 5’ to 3’-resection and the branch migration of the double Holliday junction. Notwithstanding the heterogeneity of ER locations, we observed 3 recurrent ERs that harboured the breakpoints of 38 independent *NF1* deletions. Further, 44 (92%) of the 48 ERs were located within 1-kb of the PRDM9-binding sites in both directions. If 1. 4-kb regions flanking the PRDM9-binding sites are considered, all 48 ERs were located within these regions. These findings imply that the current model of PRDM9-induced chromatin reorganization at recombination hotspots, which was established in the mouse, is also applicable to the NAHR events underlying type-1 *NF1* deletions.

3123F

Whole-exome sequencing identifies a novel 2. 5 kb duplication in *INSR* in a patient with Donohue syndrome. *M. Schotik[1,2], F. Beleggia[1], F. Percin[3], J. Altmueller[1,4], H. Thiele[4], P. Nuernberg[4], B. Wollnik[1].* 1) Institute for Human Genetics, University Hospital of Cologne, Cologne, NRW, Germany; 2) Cologne Graduate School of Ageing Research, University of Cologne, Cologne; 3) Department of Medical Genetics, Faculty of Medicine, Gazi University, Ankara, Turkey; 4) Cologne Center for Genomics, University of Cologne, Cologne, Germany.

Identification of structural variations in the genome by whole-exome sequencing (WES) constitutes a major challenge. Here we use three different software programs for copy-number variation (CNV) analysis to identify a two-exons duplication in the *INSR* gene in a patient with Donohue syndrome. Donohue syndrome (OMIM 246200) or Leprechaunism is a rare, autosomal, recessively inherited congenital disorder. It is characterized by severe insulin resistance, pre- and postnatal growth retardation, and typical physical features such as acanthosis nigricans, lack of adipose and muscular tissues, and an enlarged abdomen. We studied a Turkish girl affected by Donohue syndrome, born to consanguineous parents. After excluding point mutations in *INSR* by Sanger sequencing, we performed WES and CNV analysis. Using the programs CODEX, XHMM, and FishingCNV, we successfully identified a homozygous duplication of two exons, with an approximate size of 2. 5 kb spanning exons 10 and 11 of *INSR* (NM_000208). This duplication was confirmed to be in the homozygous state in the patient and in the heterozygous state in the parents by qPCR analysis. Moreover, using a nested-PCR approach, we could characterize the duplication as a tandem event, thereby resolving the genetic organization of the mutation. Sequencing of fibroblast *INSR* mRNA from the parents showed that the mutated allele is nearly absent and is thus likely to undergo nonsense-mediated decay, consistent with the predicted generation of a premature stop codon. To our knowledge, tandem duplications have not been described as a mutational mechanism in Donohue syndrome, indicating that a comprehensive molecular diagnostic approach might be necessary to identify all mutations in *INSR*. Additionally, our study provides evidence that WES can be successfully used to identify duplications in the lower kilobase range, below the resolution of standard array-based technologies. Therefore, WES is a rational first choice as a diagnostic tool to detect a wide range of different mutational mechanisms.

3124T

Copy number variation in early-onset Alzheimer's disease. *M. N. Bamne[1], E. McDade[2], X. Wang[1], F. Y. Demirci[1], E. Feingold[1], W. E. Klunk[3], M. I. Kamboh[1].* 1) Department of Human Genetics, University of Pittsburgh, Pittsburgh, PA; 2) Department of Neurology, University of Pittsburgh, Pittsburgh, PA; 3) Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA.

Early-onset Alzheimer's disease (EOAD) accounts for about 2% of all Alzheimer's disease cases. Mutations in three genes – APP, PS1 and PS2 – are associated with EOAD; however, mutations in these genes explain only about 50% of EOAD cases, indicating the existence of additional genetic factors. Recent studies have identified rare copy number variants (CNVs) in familial and sporadic EOAD. In this study, we performed a genome wide screen for CNVs using a high-resolution oligonucleotide array-based comparative genomic hybridization (a-CGH) (Agilent SurePrint G3 Human 1M platform) in 10 sporadic EOAD subjects having *APOE* 3/3 genotype (4 males and 6 females of Caucasian descent; age-of-onset range= 41-58 years). After filtering out CNVs reported in dbVAR (http://www. ncbi. nlm. nih. gov/dbvar) and DGVa (http://www. ebi. ac. uk/dgva) databases, we identified a total of 61 CNVs that have not been reported previously. Of these, 8 were shared between more than one EOAD subjects included in this study. Of the 61 novel CNVs, 20 were duplications and 41 were deletions. Further characterization of these CNVs and investigation in additional EOAD samples is in progress.

3125F

Copy number variant frequency and genomic diversity in the Danish population. *M. Bertalan[1], W. Mazin[1], S. Weinsheimer[1], T. Sparsø[1], M. Hollegaard[2], A. Ingason[3], T. Werge[1], BROAD Institute, iPSYCH.* 1) Mental Health Centre, Sct. Hans and Institute of Biological Psychiatry, Psychiatric Center Sct. Hans, Roskilde, Denmark; 2) Section of Neonatal Genetics and the Danish Neonatal Screening Biobank, Statens Serum Institute. Copenhagen, Denmark; 3) deCODE.

Introduction and findings: Copy number variant (CNV) frequencies are typically estimated in association studies from case-control samples subjected to ascertainment bias. Here we present, for the first time near unbiased population estimates of CNV frequencies in the Danish population.

The majority of the predicted CNVs are found in specific regions of the genome (hotspots) that are not randomly distributed over the genome. We observed that some chromosomes are enriched for CNV hotspots, including chromosomes 15, 16, 17, and 22, while other chromosomes seem depletedby such structural alterations, for example chromosomes 8 and 13.

Some hotspots are from genomic regions that have been previously associated with mental diseases, such as 1q21. 1, 15q13. 3, 16p13. 11, and 22q11. 2. However, many hotspots have not been previously associated with or studied in relation to disease. The large number of hotspots demonstrates that CNVs may account for a substantial amount of genetic variation in the Danish population.

We report diagnoses among CNV carriers in this large, random population sample as recorded in the near complete, nationwide Danish health registers, thus setting the scene for subsequent CNV association and survival analysis studies at the population level.

The study offers the first compilation of CNV frequencies and hotspots in a large population. These results improve our understanding of genomic variants and diseases, revealing genomic diversity and its evolution.

Methods: Dried blood spots (DBS) have been collected in Denmark since 1981, creating one of the largest and near complete population-based biobanks [Danish Neonatal Screening Biobank; DNSB] with samples from more than 2 mio. individual born in Denmark. Clinical phenotypic information on all DNSB subjects are retrieved from the comprehensive Danish health registers, allowing the application of epidemiology study designs in disease genetics.

A random sample were drawn from DNSB of more than 30,000 subjects born 1981-to-2006 corresponding to approx 2% of subjects born in the period. DNA was extracted from DBS, and subsequently whole-genome amplified and genotyped on PsychChip BeadChip array(Illumina) and CNVs were predicted using three complementary methods: *i*PsychCNV, PennCNV and iPattern. In this study we target CNVs tagged by more than 20 SNPs on the array, which are all considerable larger than the classical CNV size-definition of 1 kb.

3126T

DNA replication dynamics and its role in CNV instability. *L. Chen[1], W. Zhou[1], C. Zhang[2], Y. Chen[1], L. Jin[1], F. Zhang[1].* 1) School of Life Sciences, Fudan University, Shanghai, China; 2) Boston Children's Hospital, Boston, MA.

The fidelity of DNA replication is the guarantee of all activities in organisms. Many efforts have been made on discovering the overall dynamics of this complex and ultimate event. Copy number variation (CNV) in the human genome is of vital importance to human health and evolution. However, much of the molecular basis of CNV mutagenesis remains to be elucidated. Considering the DNA replicative mechanism of 'fork stalling and template switching' for CNV formation, we hypothesized that replication fork progression could be important for CNV mutagenesis. However, molecular assays of replication fork progression at the genome level are technically challenging. We therefore conducted an algorithm defining a statistic R for the estimation of DNA replication dynamics based on the available data of DNA replication timing. After comparing the R values of the CNV breakpoint regions to those of genome-wide average level in the genomes of human, mouse and *Drosophila*, a significant association of DNA dynamics and CNV instability is observed. Notably, the homology-mediated and VNTR-mediated CNVs contribute the most to the correlation, suggesting repeat-induced DNA replication error and consequent CNV formation. Our concordant observations suggest an important role for DNA replicative mechanisms in CNV mutagenesis and genome instability. Recently, the variance of DNA replication timing between individuals has been reported. Therefore, we further investigated the variance of replication timing between individuals in the Chinese populations. We hypothesize that human genetic variants may be responsible for variances of DNA replication timing and dynamics, which further suggests the interaction between CNV mutagenesis and DNA replication dynamics.

3127F

Identification of Copy Number Variation in the genome of Japanese population. *I. Danjoh[1], K. Tsuda[1,2], N. Konno[1], R. Saito[1,2], J. Yasuda[1].* 1) Medical Megabank Organization, Tohoku University, Sendai, Miyagi, Japan; 2) Life Science Analytics Center, Toshiba Corporation Healthcare Company, Sendai, Miyagi, Japan.

Eastern part of Japan was suffered from a big earthquake 4 years ago. The victims have a larger risk of acquiring diseases in the devastating environment. Advanced, genome-based precision medicine based on the precise genomic information of Japanese will be an ideal solution to restore the medical services in the damaged area. Copy number variations (CNVs) are critical factors for our mission: they are associated with various diseases, in addition to causing SNP miscalling. Therefore, we have analyzed CNVs in the genome of Japanese population to figure out the aspect of the Japanese population. The genomic DNAs from 1450 Japanese healthy volunteersin ToMMo cohort, 683 males and 767 females, who passed the standard quality control (QC) procedures for SNPs and samples including the estimation of "cryptic relatedness", were genotyped with Omni 2. 5 microarray (Illumina Inc.) on which approx. 2. 5 million SNPs are arrayed. Then the genotype data were used for further CNV analysis. CNVpartition, which is one of the module installed in GenomeStudio software and it called CNVs by likelihood-based algorithm, and PennCNV were used for CNV calling. As a result of CNV calling, we identified 3075 CNV region (CNVRs). We will report the details on CNVR region and the comparison with other Asian populations in this meeting.

3128T

Copy Number Variation in Cynomolgus monkeys linked to tissue specific gene expression. *A. R. Gschwind[1,2], T. Heckel[3], U. Certa[3], A. Reymond[1].* 1) Center for Integrative Genomics, University of Lausanne, Lausanne, Switzerland; 2) SIB Swiss Institute of Bioinformatics, Lausanne, Switzerland; 3) Roche Pharmaceutical Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, Switzerland.

Macaque monkeys are key model species for various fields of biomedical research such as simian immunodeficiency virus pathogenesis, transplantation biology, drug development and safety testing. Cynomolgus monkeys (*Macaca fascicularis*) are the most widely used non-human primate species for drug safety testing in pharmaceutical companies and naturally experimental results might be influenced by variation in biological processes among the individuals sampled. Knowledge of genetic factors contributing to variability with respect to biological drug responses could help to design better experimental approaches, which in turn would help to reduce, refine or even replace animal experiments. We attempt to investigate the importance and implications of genetic variation on cellular processes by combining genome-wide information on copy number variation (CNV) and gene expression data from 24 Cynomolgus monkeys originating from four different populations used in pharmaceutical research (Mauritius, Vietnam, China and the Philippines). Using array comparative genome hybridization (aCGH) data and a CNV calling pipeline combining three different methods, we assess copy number variation among our cohorts. These results are then combined with gene expression data from five different tissues (heart, kidney, liver, lung and spleen) to map expression quantitative trait loci (eQTLs) in proximity of the detected CNVs. We discover eQTLs in all tissues, acting mostly in a tissue specific manner. Of interest many of these loci are found in the kidney, a key organ for drug excretion. Using further downstream analyses, we attempt to get information on cellular processes possibly affected by these gene regulatory changes and to make statements on potential implications for drug safety testing.

3129F

High-Throughput Screening for Rare Copy Number Variants across a ~45,000 Subject Pediatric Biobank by qPCR. *C. Kao, R. Pellegrino, J. Garifallou, F. -X. Wang, J. Snyder, H. Hakonarson.* Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA.

Many rare copy number variations (CNVs) are known to be causative for a number of human syndromes and disorders. However, these CNVs often were discovered in the context of the most extreme phenotypes, with the assumption that the variation is pathognomonic because subsequent encounters in the clinic arise from genetic screening following clinical presentation. This ascertainment bias can thus mask the true phenotypic variability in presentation for these syndromic CNVs. We performed a non-biased screen across a large pediatric cohort of ~45,000 subjects in the biobank at the Children's Hospital of Philadelphia at five known pathogenic loci: the 22q11. 21, 16p11. 2, 15q13. 3 regions and the NRXN1 and MYT1L genes. A method was developed using quantitative, real-time PCR that enabled high-throughput screening across ~45,000 subjects in an efficient and accurate manner at a fraction of the cost compared to an array-based format. We present here the technical parameters for establishing and optimizing this workflow, including the key variables needed to assess for proper quality control. Sample quality of the DNA, the tolerance of the "master-mix" reagent, and the PCR cycling conditions are critical for consistent and reliable performance across different samples and with different primer/probe sets. In particular, different commercial "master mix" reagents have different tolerances for variations in DNA quality, primer/probe sequences, and storage at ambient temperature. For problematic samples of poor quality, we find that clean up improves performance for a large proportion of samples, despite neglible improvement of A260/280 and A260/230 ratios in many instances. For QC, the comparison of performance across different loci is critical for assessing the accuracy of an apparent copy number change at any single loci. Importantly, we found that a commonly used RNAse P reference exhibits copy number variability in ~1% of cases, which confounds the screening for rare CNVs occuring at similar or lower frequencies. The CNVs found by our qPCR-based workflow were validated by array-based approaches with high concordance (>95%). Thus, the CNV dataset generated here is robust and highly accurate, which can be used to estimate population frequencies/prevalence. Additionally, the detailed phenotyping available through electronic health records linked to each subject/sample will enable us to broadly assess phenotypic variability presented by each CNV.

3130T

Phenome-wide Copy Number Burden Association on a Range of Phenotypes in 10,000 Participants from the eMERGE Network. *D. Kim[1], J. Glessner[2], A. Lucas[1], Y. Bradford[1], S. S. Verma[1], J. H. Haines[3], P. Peissig[4], J. Denny[5], M. Basford[5], D. Crosslin[6], J. Harley[7], L. Rasmussen-Torvik[8], I. Kullo[9], E. Larson[10], P. Avillach[13,14], M. Brilliant[4], J. Connolly[2], G. Tromp[11], P. Crane[6], J. A. Pacheco[8], H. Kakonarson[2], C. McCarty[12], M. D. Ritchie[1,11].* 1) The Pennsylvania State University, University Park, PA; 2) Children's Hospital of Philadelphia, Philadelphia, PA; 3) Case Western Reserve University, Cleveland, OH; 4) Marshfield Clinic, Marshfield, WI; 5) Vanderbilt University, Nashville, TN; 6) University of Washington, Seattle, WA; 7) Cincinnati Children's Hospital, Cincinnati, OH; 8) Northwestern University, Evanston, IL; 9) Mayo Clinic, Rochester, MN; 10) Group Health, Seattle, WA; 11) Geisinger Health System, Danville, PA; 12) Essential Institute of Rural Health, Duluth, MN; 13) Boston Children Hospital, Boston, MA; 14) Center of Biomedical informatics, Harvard Medical School, Boston, MA.

Recent studies on copy number variations (CNV) have suggested that an increasing burden of CNVs is associated with susceptibility or resistance to disease. A large number of genes or genomic loci contribute to complex diseases such as autism. CNV burden as - measured by the total number of base pairs deleted or duplicated in an individual - can be used as a measure of genomic instability to identify the association between global genetic effects and phenotypes of interest. However, no systematic CNV burden association analysis has been done on a wide range of phenotypes. We performed a phenome-wide CNV burden association analysis to explore global CNV burden effects associated with multiple phenotypes in subjects from the electronic Medical Records and Genomics (eMERGE) Network, a multi-ethnic cohort with electronic medical records linked to DNA biobanks. We tested for significant associations between CNV burden, measured as base pairs of duplication, deletion, or total copy number change, in 10,538 samples and 21 phenotypes based on ICD9 diagnosis and Lab values from five eMERGE sites. Logistic regression models were fit to the data to evaluate the associations between CNV burden and case/control phenotypes. Linear regression models were used to examine associations for continuous phenotypes, adjusting for age, sex, the first three principal components of ancestry, genotyping centers, and eMERGE sites. We found several statistically significant associations between blood-related phenotypes and CNV burden including duplication burden – erythrocyte sedimentation rate (P= 4. 4x10-5), total CNV burden – erythrocyte sedimentation rate (P= 3. 6x10-5), duplication burden – LDL (P= 1. 7x10-2), total CNV burden – LDL (P= 4. 0x10-2), deletion burden – red blood cell count (P= 1. 1x10-2), and deletion burden – glycosylated hemoglobin measurement (P= 4. 8x10-2). These findings indicate that blood-related phenotypes might be more associated than other traits with global CNV burden. This is interesting given the nature of blood cells constant process of developing new cells, erythropoiesis and warrants further phenotype exploration.

3131F

Structural variation at the glycophorin locus from genome sequencing of Gambian trios. *E. M. Leffler[1], Q. S. Le[1], G. Band[1], K. Kivinen[2], G. B. J. Busby[1], M. Jallow[3], F. Sisay-Joof[3], J. Stalker[2], K. A. Rockett[1,2], D. P. Kwiatkowski[1,2], C. C. A. Spencer[1], The Gambia Genome Diversity Project.* 1) Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, Oxfordshire, United Kingdom; 2) Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, United Kingdom; 3) MRC Unit, Atlantic Boulevard, Fajara, The Gambia.

Characterising variation in complex regions of the genome remains difficult despite medical and evolutionary importance. One such challenging region is the glycophorin gene cluster on chromosome 4 encoding the putatively non-translated *GYPE* as well as *GYPA* and *GYPB*, which underlie the MNS blood group system and serve as red blood cell surface receptors for*Plasmodium falciparum*. This locus has been suggested to be under balancing selection reflecting a continued evolutionary battle between host and parasites. However, further interpretation and study is complicated by the high similarity among the three homologous copies, each about 100 kb, which makes assembly, mapping and variant calling problematic. Here, we develop a method to infer large copy number variants that are robust to the homology of the region using a hidden Markov model and sequence coverage data. We apply the method to trios from four ethnic groups in the Gambia and identify three common events, including two deletions that involve complete removal of*GYPB*, with 20% of Gambian individuals carrying non-diploid copy number in the region. We then call structural variants in worldwide 1000 Genomes populations, observing individuals with*GYPB*copy number ranging from 0 to 3 and additional duplications and deletions encompassing*GYPA*or*GYPE*. Finally, we integrate this variation into analyses of association with malaria and signatures of selection.

3132T

Effects of copy number variable regions on the transcriptome and phenotypic variation in Finnish population-based cohorts. *H. Mattsson[1,2], A. Joensuu[1,2], V. Salomaa[3], T. Lehtimäki[4], O. Raitakari[5,6], J. Kettunen[1,7,8], M. Perola[1,2,9].* 1) Dept of Health, Unit of Genomics and Biomarkers, National Institute for Health and Welfare, Helsinki, Finland; 2) University of Helsinki, Institute for Molecular Medicine Finland (FIMM), Helsinki, Finland; 3) Dept of Health, Unit of Health Monitoring, National Institute for Health and Welfare, Helsinki, Finland; 4) Dept of Clinical Chemistry, Fimlab Laboratories and School of Medicine, University of Tampere, Tampere, Finland; 5) Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland; 6) Dept of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland; 7) Computational Medicine, Institute of Health Sciences, University of Oulu, Oulu, Finland; 8) NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland; 9) University of Tartu, Estonian Genome Center, Tartu, Estonia.

Copy number variants (CNVs) represent a significant source of genetic diversity and an important evolutionary force. There is also growing evidence linking copy number variation to complex disease susceptibility. The pathogenic effects of CNVs on human health are not fully understood but may be explained by various ways including modifications in gene dosage and gene expression. We conducted an association analysis of 4258 copy number variable regions (CNVRs) with peripheral blood leucocyte transcriptome in order to determine the effect of common variants on gene expression in *cis* in a Finnish population cohort the Young Finns Study (YFS, N=3596). A linear regression model adjusted for potential confounders was used to examine the associations between CNVRs and expression levels of 37,022 transcripts. A comparable model was used to evaluate phenotypic consequences of CNVRs that showed significant association with gene expression. For replication we used a *cis* window of +/-1Mb around each CNVR considering the differences in CNV breakpoints produced by various genotyping platforms. After taking multiple testing into account, associations of 216 CNVRs with transcript levels were observed (p<4. 9e-07) in the YFS. The CNVR on chromosome 10q11. 22 associated with expression of *ANXA8* in the YFS and showed further association with diastolic blood pressure in women (p=0. 038). We replicated these associations in an independent cohort DILGOM (N=278) with p-values 2. 3e-06 and 4. 2e-02 respectively. In conclusion, the present study highlights the importance of integrative analyses of genome-wide copy number data and gene expression profiles in identification of common disease variants which may offer more in-depth awareness on genotype-phenotype associations. Our results also suggest possible differences between sexes in CNVR contribution to complex traits through intermediate phenotypes.

3133F

Detecting copy number variations by analyzing whole exome sequencing data using the eXome Hidden Markov Model. *S. Miyatake[1], E. Koshimizu[1], A. Fujita[1], R. Fukai[1], E. Imagawa[1], C. Ohba[1], I. Kuki[2], A. Araki[3], Y. Makita[4], T. Ogata[5], M. Nakashima[1], Y. Tsurusaki[1], N. Miyake[1], H. Saitsu[1], N. Matsumoto[1].* 1) Department of Human Genetics, Yokohama City University, Yokohama, Japan; 2) Department of Pediatric Neurology, Pediatric Medical Care Center, Osaka City General Hospital, Osaka, Japan; 3) Department of Pediatrics, Kansai Medical University, Hirakata, Japan; 4) Education Center, Asahikawa Medical University, Asahikawa, Japan; 5) Department of Pediatrics, Hamamatsu University School of Medicine, Hamamatsu, Japan.

Whole exome sequencing (WES) is becoming a standard tool for detecting nucleotide changes, and is now known as a possible method to detect copy number variations (CNVs). To date, several algorithms have been developed for such analyses, although verification is needed to establish if they fit well for the appropriate purpose, depending on the characteristics of each algorithm. Here, we performed WES CNV analysis using the eXome Hidden Markov Model (XHMM). We validated its performance using 27 rare CNVs previously identified by microarray as positive controls, finding that the overall detection rate for known CNVs was as high as 59%, and even higher (89%) with three or more targets. Larger CNV size increases the detection rate, and in general, CNVs > 200 kb are likely to be detected. Overall, XHMM tended to call smaller CNVs than microarrays. In this study, 22. 2% of total calls were < 10 kb, and 71. 7% < 100 kb, indicating it may be possible to detect exonic deletions. However, we found that about half of these calls were false-positives, regardless of the quality score. It may be important to minimize conditional experimental differences for higher true positive rates. Although adjacent read information is usually limited, XHMM data can be useful for identifying CNV breakpoints. We were able to successfully determine breakpoints of 7- and 22-kb deletions, which are usually difficult to detect by microarrays. Identifying breakpoints can provide precise information, including in-frame gene fusions. Next we applied XHMM to genetically unsolved patients, demonstrating successful identification of pathogenic CNVs: 1. 5–1. 9-Mb deletions involving *NSD1* in patients with unknown overgrowth syndrome leading to the diagnosis of Sotos syndrome, and 6. 4-Mb duplication involving *MECP2* in affected brothers with late-onset spasm and progressive cerebral/cerebellar atrophy confirming the clinical suspect of *MECP2* duplication syndrome. As it is not known if a nucleotide variant or gross CNV is the culprit before analysis, an “exome-first” approach may save time and cost of unnecessary microarray analysis. This approach is even more effective with target CNVs > 200 kb.

3134T

Understanding the genetic structure of FCGR genes region for Kawasaki disease using sequence capture of target fragments for long reads sequencing. *E. Png[1], P. F. De Sessions[1], L. Low[1], H. Liany[2], V. Kumar[2], C. E. Tacke[3], S. Nagelkerke[4], J. Geissler[4], S. Davila[2], T. W. Kuijpers[3,4], M. Hibberd[1].* 1) Infectious Diseases, Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672; 2) Human Genetics, Genome Institute of Singapore, 60 Biopolis Street, Singapore 138672; 3) Pediatric Hematology, Immunology and Infectious Diseases, Emma Children’s Hospital, Academic Medical Center (AMC), Amsterdam, The Netherlands; 4) Department of Blood Cell Research, Sanquin Research and Landsteiner Laboratory, Academic Medical Center, Amsterdam, The Netherlands.

Kawasaki disease (KD) is an acute inflammatory disorder associated with vasculitis, usually affecting the coronary arteries. In recent genome wide association studies (GWAS), we identified a non-synonymous single nucleotide variant (SNV) (rs1801274, P= 7. 35x10-11, OR 1. 32, 95% C. I. 1. 22-1. 44) in the *FCGR2A* gene to be significantly associated with KD susceptibility. The *FCGR2A* gene is situated among the *FCGR2/3* locus, which contains highly homologous genes encoding the five major Fc-gamma receptors (*FCGR*) for immunoglobulin G (IgG). Incidentally, the infusion of intravenous IgG is the standard effective treatment for KD. Yet, genotype variations in the *FCGR* genes have been demonstrated to influence IgG response in certain KD patients. Because of the immunologically importance of this locus in susceptibility and managing KD, we need a better understanding of the genetic mechanism underlying the *FCGR* genes associations. However, the coverage of probes in the GWAS array is sparse over this *FCGR2/3* locus, due to the presence of a large segmental duplication and copy number variations (CNV). Thus, GWAS SNPs alone are unlikely to help decipher the full spectrum of genetic variations that are significant to KD. In order to identify potentially functional variants, we performed targeted re-sequencing of the entire *FCGR2/3* locus in 6 sets of trio families with individuals affected or not affected by KD. We synthesized NimbleGen SeqCap EZ library baits, which is estimated to cover 92% of the target region of interest (ROI) (chr1:161,500,415-161,683,654), to pull down fragments in the ROI for sequencing using the PacBio System. The sequenced data is aligned to the human reference using BWA and GraphMap software. SNV is identified using GATK Unified Genotyper. Larger structural variations are called using PBHoney and cnvSeqCap for confirmation. On average we have 12500 circular consensus sequences per library, with a mean read quality of insert of >99%. More than 4Mb of sequences from each library mapped to the ROI allowing each target base to be covered >20x. The genotype of variants identified through sequencing will be analyzed for concordance with those called independently using a multiplex ligation-dependent probe amplification (MLPA) assay. Haplotypes analysis and imputation will be used to infer the genotype of variants identified through sequencing, to the entire KD cohorts that have subset of variants genotyped using the MLPA assay.

3135F

Genic intolerance to copy number variation in 60,000 individuals and applications to identifying risk genes in schizophrenia. *D. Ruderfer[1], M. Fromer[1], T. Hamamsy[1], M. Lek[2,3], K. Samocha[2,3], K. Karczewski[2,3], J. Moran[3], S. McCarroll[3], C. Hultman[4], P. Sullivan[5], P. Sklar[1], M. Daly[2,3], D. MacArthur[2,3], S. Purcell[1,2,3], Exome Aggregation Consortium.* 1) Icahn School of Medicine at Mount Sinai, New York, NY; 2) Massachusetts General Hospital, Boston, MA; 3) Broad Institute, Cambridge, MA; 4) Karolinska Institutet, Solna, Sweden; 5) University of North Carolina at Chapel Hill, Chapel Hill, NC.

Copy number variation (CNV), in particular a gain or loss of coding sequence, is known to contribute to phenotypic diversity and risk of diseases including schizophrenia (SCZ). Here we characterize genic intolerance to CNV – defined as observing fewer CNVs in a gene than expected – using data from nearly 60,000 individuals from the Exome Aggregation Consortium (ExAC). We hypothesized that CNVs in intolerant genes would be more likely to have deleterious effects. We further show how this resource and the derived genic intolerance values can be used to relate CNV burden to risk of SCZ. We calculated genic likelihoods of CN per individual, allowing us to distinguish between a diploid state for the extent of a gene, versus a high likelihood of partial or complete gene CNV, versus the absence of a confident call. Individuals averaged 0. 81 deleted and 1. 75 duplicated genes although 63% and 46% of individuals did not have any deleted or duplicated genes, respectively. A single gene was involved in 34% of CNVs and 62% of single gene CNVs are likely to disrupt (partial deletion or duplication) the gene. Genes with fewer CNVs than expected were defined as more intolerant. In the absence of accurate models of CNV mutation rates across the genome, we calculated genic intolerance from the residuals of a logistic regression of CNV frequency on gene length, read depth, GC content, sequence complexity and presence between pairs of segmental duplications. CNV intolerance was significantly correlated with a haplo-insufficiency constraint score based on loss-of-function variants in the same sample (deletions r=0. 23, duplications r=0. 11). Genes highly expressed in the brain were the most intolerant to CNVs (p=7. 66x10-16); similarly, the mostly highly intolerant genes (n=157) were involved in neuronal and axon development. In a subset of individuals with SCZ and matched controls, cases were more likely to have a CNV in an intolerant gene (22% versus 18%). Over and above the increased burden, genes hit by CNV in cases were also more intolerant compared to controls (p=0. 007). Although directly using ExAC CNV data as a convenience control sample runs a high risk of bias, here we also demonstrate improved power to detect SCZ loci when considered along with an appropriate matched control sample.

3136T

Complex structural variation in the MHC locusinfluences schizophrenia risk by shaping expression of *complement component 4*. *A. Sekar[1,2], A. R. Bialas[3,4], H. de Rivera[1,2], T. R. Hammond[3], A. Davis[1,2], J. Presumey[4], N. Kamitaki[1,2], K. Tooley[1,2], G. Genovese[1,2], M. Baum[1,2,3], V. Van Doren[1], S. Rose[2], R. E. Handsaker[1,2], M. J. Daly[2,5], M. C. Carroll[4], B. Stevens[2,3], S. A. McCcarroll[1,2], Schizophrenia Working Group of the Psychiatric Genomics Consortium.* 1) Department of Genetics, Harvard Medical School, Boston, MA, USA; 2) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA; 3) Department of Neurology, F. M. Kirby Neurobiology Center, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA; 4) Program in Cellular and Molecular Medicine, Children's Hospital Boston, MA, USA; 5) Analytical and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA.

Complex forms of structural variation affect hundreds of human genes, but their biological significance is largely unknown. We found that schizophrenia's strong but previously unexplained association with common genetic variation in the Major Histocompatibility Complex (MHC) locus arises in part from diverse structural alleles of the complement component 4 (*C4*) gene. These *C4* structural alleles promoted widely varying levels of *C4A* and *C4B* expression in different individuals' brains. Among 28,799 cases with schizophrenia and 35,986 controls, alleles of *C4* associated with schizophrenia in proportion to their tendency to promote higher brain expression of *C4A*. We found human C4 protein localized at dendrites, axons, neuronal cell bodies, and synapses. In mice, C4 mediated synapse elimination during postnatal development. These results implicate excessive complement activity in schizophrenia and may help explain the reduced numbers of synapses in the brains of affected individuals. The implication of the complement pathway suggests novel strategies for treating schizophrenia.

3137F

Copy number variant conferred risk of mental disorders in the Danish population. *T. Sparsø[1], SM. Weinsheimer1[1], M. Bertalan[1], W. Mazin[1], M. Hollegaard[2], T. Werge[1], SSI, BROAD & iPSYCH investigators.* 1) INSTITUTE OF BIOLOGICAL PSYCHIATRY, Mental Health Services of Copenhagen, Copenhagen, Denmark; 2) Section of Neonatal Genetics and the Danish Neonatal Screening Biobank, Statens Serum Institut, Copenhagen, Denmark.

Introduction and findings: It is well established that copy number variants (CNV) confer risk of both somatic and mental disorders. Risk of disease is typically estimated from case-control studies subjected to ascertainment bias among others. Here we present, for the first time near unbiased population estimates of CNV conferred risk of mental disorders in the Danish population. We here report the study of a population-based sample of more than 80. 000 subjects born 1981-to-2006, corresponding to approx. 5% of the individuals born in this period, and including all subjects with autism, ADHD, schizophrenia, bipolar disorder and major depression, and healthy controls. We identified more than 100 hotspots in the genome where CNVs are shared in the Danish population with frequencies ranging from 0. 01% to 5%, and for each we estimate the conferred risk for the five target disorders and combinations thereof. The initial association analyses are complemented by survival analysis to provide estimates of risk corrected for sex, age, calendar time, place of birth, maternal and paternal age, parental origin and family history of psychiatric disorders. The study offers the first compilation of population-based risk estimates of CNVs for mental disorders. These results improve our understanding of structural variants in the genome that cause mental disorders and prompt studies of the underlying biology. *Methods:* Dried blood spots (DBS) have been collected in Denmark since 1981, creating one of the largest and near complete population-based biobanks [Danish Neonatal Screening Biobank; DNSB] with samples from more than 2 mio. individuals born in Denmark. Clinical phenotypic information on all DNSB subjects are retrieved from the comprehensive Danish health registers, allowing the application of epidemiology study designs in disease genetics. A random sample were drawn from DNSB of more than 25,000 subjects born 1981-to-2005 corresponding to approx 2% of subjects born in the period. Furthermore, all subjects born in this period and assigned a clinical diagnosis of autism (n=12,735), ADHD (n=13,004), schizophrenia (n=2,311), bipolar disorder (1,121), affective disorders (n=14,985) or anorexia (2,012) recorded in the Danish health registers were selected for analysis. CNVs were predicted using data obtained from the Infinium PsychArray BeadChip (Illumina) using three complementary methods: *i*PsychCNV, PennCNV and iPattern.

3138T

Comprehensive comparative performance analysis of high-resolution array platforms for genome-wide CNV detection in humans. *A. Urban, R. R. Haraksingh.* Psychiatry and Genetics, Stanford University, Palo Alto, CA.

High-resolution microarray technology can comprehensively and efficiently detect copy number variants (CNVs) in the human genome. It is routinely used for fundamental applications in basic and translational research as well as in clinical practice – namely in clinical cytogenetics, large association studies, the tracking of genome stability in stem cell culture, and the orthogonal confirmation of CNVs detected by sequencing methods. A new generation of array designs from several manufacturers (Affymetrix, Agilent, and Illumina) has recently become available. These designs combine high probe densities (interrogating hundreds of thousands and up to almost 5 million loci on a single array) with optimized probe placement strategies and experimental workflows, with the expectation of providing an essential workhorse for CNV-level genome analysis for the coming years. We carried out a comprehensive comparative analysis of the relative detection and resolving power of the 16 currently available array platforms for genome-wide CNV analysis. We achieved unbiased comparison by hybridizing DNA from the well-characterized genome of HapMap/1000 Genomes Project subject CEU-NA12878 to all arrays and performing data analysis using both manufacturer-recommended and platform-independent software for each array. These are the same study design features that we used in [Haraksingh et al. , 2011] now applied to the all the currently and newly available array designs and platforms. Our results confirm some general expectations e. g. regarding the benefits of increasing probe numbers, while providing quantitative measures of power of detection. Furthermore we also show that array design specifics can have striking and unexpected effects on array performance, which could only be revealed through direct benchmarking using a previously well-characterized genome and direct comparison of array designs. For example adding probes specific to the exonic regions of the genome to an already high-powered genome-wide design only somewhat increases the power of detection of a given array design and may lead to a considerable increase in false positive CNV calls. These results will help to inform appropriate platform selection for a given individual project or application (e. g. clinical diagnostics vs. association study vs. cell line monitoring), and also enable the genomics community to gauge the CNV-analytical power of already existing and ongoing array-based genomics studies.

3139F

A single rearrangement event resulting in two concomitant genomic disorders: the *PMP22-RAI1* contiguous gene duplication syndrome. *B. Yuan[1], T. Harel[1], S. Gu[1], P. Liu[1], A. Coe[2], M. Wright[3], L. Burglen[4], S. Chantot-Bastaraud[4], V. Gelowani[1,5], C. R. Beck[1], C. M. B. Carvalho[1], S. W. Cheung[1], A. Munnich[6], P. L. Magoulas[1], L. Potocki[1], J. R. Lupski[1,5,7,8].* 1) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA; 2) University Hospitals, Coventry and Warwickshire NHS Trust, University Hospital, Coventry, UK; 3) Institute of Human Genetics, Newcastle University, Newcastle, UK; 4) APHP, Service de Génétique et embryologie Médicales, CHU Paris Est - Hôpital d'Enfants Armand-Trousseau, Paris, France; 5) Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas, USA; 6) Service de Génétique, INSERM U781, Hôpital Necker-Enfants Malades, Institut Imagine, University Sorbonne-Paris-Cité, Paris, France; 7) Department of Pediatrics, Baylor College of Medicine, Houston, Texas, USA; 8) Texas Children's Hospital, Houston, Texas, USA.

The genomic duplication associated with Potocki-Lupski syndrome (PTLS) maps in close proximity to the duplication associated with Charcot-Marie-Tooth disease type 1A (CMT1A) on chromosome 17p11. 2-17p12. PTLS is characterized by hypotonia, failure to thrive, reduced body weight, intellectual disability, and autistic features. CMT1A is a common autosomal dominant distal symmetric peripheral neuropathy. The key dosage sensitive genes *RAI1* and *PMP22* are associated with PTLS and CMT1A, respectively. Recurrent duplications accounting for the majority of subjects with these conditions are mediated by nonallelic homologous recombination between low copy repeats (LCRs). The LCR pairs flanking *RAI1* and *PMP22* do not share extensive homology with each other; thus, duplications encompassing both loci are rare and potentially generated by a different mutational mechanism. We characterized genomic rearrangements duplicating both *PMP22* and *RAI1*, including nine potential complex genomic rearrangements, in twenty-two subjects by high-resolution array comparative genomic hybridization and breakpoint junction sequencing. Insertions and microhomologies were found at the breakpoint junctions, suggesting potential replicative mechanisms for rearrangement formation. At the breakpoint junctions of these nonrecurrent rearrangements, enrichment of repetitive DNA sequences was observed, indicating that they might predispose to genomic instability and lead to rearrangement. Clinical evaluation revealed blended CMT1A and PTLS phenotypes with a potential earlier onset of neuropathy. Moreover, additional clinical findings may be found due to the extra duplicated material included in the rearrangements. Our genomic analysis suggests replicative mechanisms as a predominant mechanism underlying *PMP22-RAI1* contiguous gene duplications and provides further evidence supporting the role of complex genomic architecture in genomic instability.

3140T

Comprehensive breakpoint analyses for genomic disorders reveal the correlation of CNV size and complexity with clinical severity at the *PLP1* locus. *L. Zhang[1], J. Wang[2], C. Zhang[1], D. Li[2], J. Xiao[1], W. Zhou[1], C. Carvalho[3], C. Liu[4], Y. Luo[5], L. Jin[1], J. Lupski[3], F. Zhang[1], Y. Jiang[2].* 1) Fudan university, Shanghai, China; 2) Department of Pediatrics, Peking University First Hospital, Beijing, China; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA; 4) Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Beijing, China; 5) The Research Center for Medical Genomics, MOE Key Laboratory of Medical Cell Biology, College of Basic Medical Science, China Medical University, Shenyang, China.

Human genomic copy-number variation (CNV) can be responsible for many human diseases, especially genomic disorders. The molecular mechanisms of genomic disorders include dosage effect, position effect, and etc. Pelizaeus-Merzbacher disease (PMD, MIM#312080) is a classic genomic disorders that can be caused by the nonrecurrent CNVs at the *PLP1* locus. However, it is technically challenging to get the exact breakpoints of the PMD-associated CNVs. Here we reported 49 cases with *PLP1* gene CNVs, 19 cases of which have complex CNVs. In addition to using previously reported methods of traditional long-range PCR and target-DNA capture sequencing, we also newly developed Asymmetry Primers PCR (APP), which is a high-efficiency method of CNV breakpoint analysis. Therefore, we successfully obtain the CNV breakpoints in more than 90% of the PMD cases. Intriguingly, the CNV sizes are associated with the severity of the clinical features in PMD patients (P<0. 05). Additionally, complex CNV structures, especially the inverted duplications, tends to induce more severe clinical features (P<0. 05). In conclusion, we found the size and the genomic orientation of the pathogenic CNVs at *PLP1* affected the phenotype of genomic disorders. We also showed that the APP method can be used for analyzing CNV breakpoint sequences, especially for the challenging cases with complex CNVs in the genomic regions with complex genome architecture.

3141F

Correlations between *AMY1* copy number, diet and BMI in ethnically diverse African populations. *K. E. Johnson[1], C. C. Elbers[1,2], S. R. Thompson[1,2], A. Ranciaro[1,2], D. W. Meskel[3], A. L. Mandel[4], G. Belay[3], S. A. Tishkoff[1,2].* 1) Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA; 2) Department of Biology, School of Arts and Sciences, University of Pennsylvania, Philadelphia, PA; 3) Department of Biology, Addis Ababa University, Addis Ababa, Ethiopia; 4) Monell Chemical Senses Center, Philadelphia, PA.

Amylase 1 (*AMY1*) is expressed in human saliva and is the first step in digestion of starches. Previous research has shown that *AMY1* copy number varies significantly between human populations, and that higher copy number may have been selected in populations with a starchy diet (Perry *et al,Nature Genetics*, 2007). To further investigate the relationship between subsistence methods and *AMY1* copy number, we estimated copy number using qPCR in a diverse data set of 1390 individuals from 20 populations in Tanzania and Ethiopia. Pastoralist populations, which consume a low-starch diet, are associated with significantly lower copy number compared to other subsistence methods (P = 3. 99x10-7). Agriculturalism is also associated with a decrease in copy number (P = 0. 0003), while hunter gathering and agropastoralism had no significant effect on copy number. *AMY1* copy number has also been associated with obesity and BMI (Falchi *et al, Nature Genetics*, 2014), with increased copy number leading to decreased risk of obesity. We tested for association between *AMY1* copy number and BMI in our diverse sample, and found no association in any of the populations alone nor in the combined sample (P = 0. 347). However, these populations contain no individuals that would be classified as obese by BMI. In conclusion, we found significant differences in copy number between African populations practicing different subsistence methods, but were unable to replicate the association between *AMY1* copy number and BMI.

3142T

Customizable exon-centric target enrichment for copy number and targeted mutation detection. *K. Jeong, A. Vadapalli, B. Peter, A. Ashutosh, D. Joshi, J. Ghosh, D. Roberts.* Agilent Technologies, Santa Clara, CA.

Structural variations in the genome can be determined from NGS data with either whole genome sequencing (WGS) or targeted enrichment using exome or gene panels. WGS has the potential to provide a single platform solution for detecting copy number variations (CNVs). However, it is prohibitively expensive in identifying mutations such as single nucleotide polymorphisms (SNPs), and insertions and deletions (INDELs) that require high sequencing read depths. Although whole (WES) or targeted exome sequencing is a cost-effective method for identifying mutations in exons, high read depths from several neighboring exonsare often required to provide confident copy number determination. WES also does not provide accurate aberration boundaries in potentially important regulatory regions outside of exons such as in introns or promoters. To enable higher resolution copy number calls in and around genes/exons of interest, we have augmented exonic tiling with custom target enrichment probes designed specifically to improve copy number (CN) resolution in these regions while maintaining exonic coverage of SNPs/INDELs. OneSeq is an NGS target enrichment solution providing simultaneous detection of genome-wide CNVs and copy neutral loss of heterozygosity (cnLOH) in addition to SNPs and INDELs at desired loci. Herein we describe custom OneSeq, enabling higher resolution, exon-centric CN determination in any genes of interest. To this end, we have created a library of high resolution target enrichment probes targeting exon-proximal regions. When deployed in conjunction with existing exonic probes, they improve detection of biologically relevant CNVs in and near exons, while simultaneously detecting exonic SNPs and INDELs. The targeted CN regions can be customized to address only a subset of genes, thus maximizing sequencing efficiency. To demonstrate the efficacy of custom OneSeq, we compare customized OneSeq panels to targeted exonic sequencing alone. Custom OneSeq improves detection of copy number changes with resolution often spanning just a few exons. Additionally, we identify aberrations in samples with either SNP/INDEL or CNV at a particular locus, and in some cases, compound heterozygotes with different aberrations in both alleles of the same locus. The addition of customizable, exon centric OneSeq probes provides a means to detect exonic, intronic and intergenic CNVs, together with SNPs/INDELs in a single assay.

3143F

Modifying DNA extraction protocol alters human brain copy number estimation. *C. Proukakis[1], K. Gancheva[2], A. Soenmez[1], C. Grace[2], R. Valli[3], D. Pease[1], A. Ejaz[1], H. Houlden[4], A. Pittman[4], S. Vattathil[5], E. Maserati[3], J. W. Taanman[1], A. H. Schapira[1], E. Nacheva[2].* 1) Clinical Neuroscience, UCL Institute of Neurology, London, England, United Kingdom; 2) Academic Haematology, Royal Free Campus, UCL, London, UK; 3) Genetica Umana e Medica, Dipartimento di Medicina Clinica e Sperimentale, Université dell'Insubria, Varese, Italy; 4) Molecular Neuroscience, UCL Institute of Neurology, London, England, United Kingdom; 5) Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA.

Background Mosaicism may be relatively common, including in the brain, and a role in common sporadic diseases such as Parkinson disease (PD) has been proposed. Sub-integer copy number (CN) changes could indicate mosaicism, but potential confounding influences of DNA extraction methods on CN estimates remain largely unexplored. **Methods and results**We analysed brain DNA extracted using spin columns from the cerebellum and frontal cortex (FC) of 4 PD cases and 4 controls, initially using a custom aCGH design focused on PD genes. We did not detect aberrations in PD genes. Further review suggested possible widespread subtle CN changes, mostly losses in cerebellum, predominantly in GC rich regions, even when using matched FC DNA as reference. CN determination of 2 genes (TSC2, EIF2C1) by droplet digital PCR (ddPCR) was consistent with mosaic losses, despite normal results in peripheral monocyte DNA. SNP array did not, however, detect any B-allele frequency (BAF) deviations. We therefore considered the possibility that we were incompletely extracting certain genomic regions, leading to falsely low CN estimates. Varying the protocol by using overnight protease digestion, lower starting tissue amount, and a non-spin column method (Puregene), radically altered results. ddPCR showed almost complete normalization. When DNA from cerebellum and FC of the same brain extracted with Puregene was hybridized together on aCGH, the losses previously seen were no longer evident. Losses were, however, still seen in direct hybridizations of certain combinations of DNA extracted differently from the same cerebellum. Increasing the proteinase K amount in Puregene extraction affected aCGH results in a cerebellar sample, but had no clear effect on the FC. **Discussion and conclusions** As apparent losses were present in both aCGH and ddPCR, mosaicism was a plausible explanation. The clear influence of DNA extraction method on the result, however, is consistent with a previous report of systematic bias introduced during DNA extraction from rat tissues. GC rich regions in particular may be inadequately recovered from spin columns, particularly if large amounts of starting material are used, or with any protocol, if protein digestion is incomplete. The reason for the clearer appearance of this phenomenon in the cerebellum may be the higher cell density. CN estimation using any technique may be misleading after suboptimal DNA extraction.

3144T

***De Novo* Copy Number Variation in Taiwanese Trios Affected by Schizophrenia.** *T. Poterba[1], J. Maller[1], D. Howrigan[1], K. Chambert[1], J. Moran[1], H. Hwu[6], W. J. Chen[6], S. V. Faraone[5], S. J. Glatt[5], M. Tsuang[4], S. McCarroll[1, 3], B. M. Neale[1, 2].* 1) Stanley Center, Broad Institute of Harvard and MIT, Cambridge, MA; 2) Massachusetts General Hospital, Boston, MA; 3) Harvard University School of Medicine, Boston, MA; 4) University of California San Diego, La Jolla, CA; 5) SUNY Upstate Medical University, Syracuse, NY; 6) National Taiwan University, Taipei, Taiwan.

Schizophrenia is a highly heritable psychiatric disorder, but the collective understanding of the genetic mechanisms behind its pathology is incomplete. Both common SNP variants and large rare copy number variants (CNVs) have been shown to harbor substantial risk for schizophrenia. However, the contribution of *de novo* CNVs to schizophrenia risk is not well characterized, with previous reports of ~600 trios with schizophrenia. This work demonstrated a role for recurrent *de novo* events in schizophrenia (Kirov et al, 2012). Therefore, we performed a large-scale genome-wide experiment to identify CNVs, both*de novo*and inherited, associated with risk for schizophrenia. We collected 1121 parent-child trios from Taiwan, generally with only the child affected with Schizophrenia. This cohort is double the size of the largest schizophrenia trio cohort reported in the literature (Kirov et al, 2012). All trios were genotyped on PsychChip, a custom Illumina array with 588k total SNPs, 255k of which are common (minor allele frequency > 0. 05) in East Asian populations. We called CNVs on samples passing initial QC using PennCNV (Wang et al, 2007), and these calls were filtered based on size and quality. We compiled and compared calls within trios to generate one list of inherited CNVs and one of de novo events found in the probands. Initial results are largely consistent with the literature, with about 4% of our probands possessing a*de novo*large CNV, many of which localized to known disease regions like 22q11 deletions, 16p11 duplications, and 3q29 deletions. We also observed an enrichment of extremely large (greater than 2 MB) CNVs, occurring in roughly 1% of probands.

3145F

Sequence analysis and characterization of active human Alu subfamilies. *M. K. Konkel[1], J. A. Walker[1], A. B. Howes[1], M. C. Ranck[1], C. C. Fontenot[1], J. Storer[1,2], D. Kural[3,4], M. P. Strömberg[3], C. Stewart[3,5], G. T. Marth[3,6], M. A. Batzer[1], the 1000 Genomes Consortium.* 1) Department of Biological Sciences, Louisiana State University, Baton Rouge, LA; 2) Department of Molecular, Cellular and Developmental Biology, The Ohio State University, Columbus, OH; 3) Department of Biology, Boston College, Chestnut Hill, MA; 4) Seven Bridges Genomics, Inc. , Cambridge, MA; 5) The Broad Institute of MIT and Harvard, Cancer Genome Computational Analysis, Cambridge, MA; 6) Eccles Institute of Human Genetics, University of Utah, Salt Lake City, UT.

Structural variations (SV), such as deletions, duplications, and insertions, are common within humans. Insertions are commonly caused by transposons also referred to as, mobile element insertions (MEIs). Alu elements are one type of MEI, and are major contributors to genome variation among humans. In recent years, MEI detection from short-read sequencing data has become a way to identify MEIs on a large scale. Subfamily assignment for Alu insertions from short-read data has also been performed. However, comprehensive wet-bench validations are often limited. To better understand the properties of MEIs validated by PCR and toward a gold standard for MEIs, we performed Sanger sequencing for 322 previously PCR validated polymorphic Alu insertions. For each element, we determined features associated with target-primed reverse transcription (TPRT), the mechanism by which MEIs are typically inserted into the genome. These hallmark features include target site duplications, an Adenosine-rich (A)-tail, and an endonuclease cleavage site. Moreover, we determined the precise genomic breakpoint coordinates, subfamily assignment, and divergence from the respective subfamily consensus sequence. More than 98% of the sequenced MEIs contained the expected features as well as a breakpoint indicative of insertion through the TPRT mechanism. Five of six events previously identified as insertions, were in fact deletions roughly the size of an Alu element (about 300 bp). These events were distinguished from traditional MEIs because the breakpoint resided within an Alu element and/or within the flanking sequence, further supported by an alignment to the orthologous region of the chimpanzee genome (and presumably the ancestral genomic state). The remaining locus, represented an Alu insertion, fixed in the human population with an A-tail of varying length due to the expansion of a microsatellite sequence. All of the sequenced Alu loci were members of the AluY lineage. AluYa5 is the most active Alu subfamily in human populations, followed by AluYb8 and AluY. In addition, we identified several other subfamilies, including three previously uncharacterized groups that we have termed AluYb7a3, AluYb8b1, and AluYa4a1. In summary, many Alu subfamilies are currently active within human populations, including a surprising level of AluY retrotransposition. Human Alu subfamilies exhibit continuous evolution with potential drivers generating a succession of new Alu lineages.

3146T

Biochemical analysis of chimeric human and mouse AMCase proteins expressed in *Escherichia coli*. *K. Okawa[1], M. Ohno[1,2], A. kashimura[1], Y. Kobayashi[1], M. Sakaguchi[1], Y. Sugahara[1], F. Fumitaka[1].* 1) Department of Chemistry and Life Science, Faculty of Engineering, Kogakuin University, Hachioji, tokyo, Japan; 2) Research Fellow of Japan Society for the Promotion of Science (DC2), Chiyoda-ku, Tokyo, Japan.

Chitinases are enzymes that digest chitin. Acidic mammalian chitinase (AMCase) is implicated in asthma, allergic inflammation and food processing. Human and mouse AMCase share 82% sequence identity and 86% sequence similarity. Despite these structural similarities, these two homologues significantly differ with respect to their enzymatic characters at enzymatic activities. The chitinolytic activity of the recombinant human AMCase was less than that of the mouse counterpart. To determine regions responsible for the reduced chitinolytic activity in human AMCase, we constructed chimeric human-mouse AMCase proteins and expressed in *E. coli.* We successfully expressed chimeric AMCase proteins in *E. coli* and measured the chitinolytic activity of the proteins. We found that the N-terminal region of human AMCase containing conserved active site residues significantly reduced the chitinolytic activity among the chimeric enzymes. Thus, the N-terminal region of human AMCase impairs the chitinolytic activity of the enzyme.

3147F

Whole Genome Reference Panel of One Thousand Japanese Individuals and Bioinformatics in ToMMo. *M. Nagasaki.* Tohoku Medical Megabank Organization, Tohoku University, Tokyo, Japan.

Tohoku University Tohoku Medical Megabank Organization (ToMMo) located at the north-eastern Japan is now developing a biobank (the final goal is 150,000 volunteers with millions samples) that combines medical and genome information during the process of rebuilding the community medical system and supporting health to foster the reconstruction from the Great East Japan Earthquake on Mar/11/2011. One of the missions of ToMMo is to reveal a fine genetic architecture of Japanese population to tackle the further genome wide associate study analysis by combining the knowledge, which is daily accumulated in the ToMMo prospective genome cohort project, e. g. questionnaire data, physiological data, medical treatment records and other omics data from serum, plasma and immortalized lymphocytes. The first goal is to sequence one thousand samples to cover MAF > 0. 5% variants including short insertion, deletion, and large structural variants in Japanese for constructing the Japanese whole-genome reference panel (1KJPN). In our experimental design, to minimize the biases caused by the different equipment, protocol and bioinformatics analysis, we performed whole genome sequencing of one thousand samples with 30x high coverage using the HiSeq 2500 rapid run mode and analyzed by the same bioinformatics pipeline. This talk presents our bioinformatics analyses and preliminary findings while constructing 1KJPN.

3148T

EyeSeq - Toward a Comprehensive Genomic Research and Diagnosis Resource for Inherited Retinal Degenerations. *L. Shen, D. Navarro-Gomez, M. Maher, E. M. Place, M. B. Consugar, J. Comander, J. L. Wiggs, E. A. Pierce, X. Gai.* Ocular Genomics Institute, Dept. Ophthalmology, Massachusetts Eye and Ear Infirmary and Harvard Medical School, Boston, Massachusetts, USA.

Inherited retinal degenerations (IRDs) are a large family of highly heterogeneous blinding diseases characterized by progressive death and dysfunction of rod and cone photoreceptor cells. Existing resources of IRD-associated variants include the commercially available HGMD, RetinoGenetics (http://www.retinogenetics.org/) with data on 194 genes derived from 943 publications, Retina international (http://www.retina-international.org/) with data on 78 IRD genes, and single disease databases like USHbase with 20 genes only but richer annotations. These resources are limited in data and analytical capabilities. This has presented a significant hurdle for genetic diagnosis of IRDs, especially with the emergence of NGS technologies and associated techniques such as whole-exome sequencing. The Ocular Genomic Institute (OGI) at Massachusetts Eye and Ear Infirmary conducts both research and CLIA certified diagnostic service for genetic eye diseases including IRDs, early-onset glaucoma and optic atrophy. We established a comprehensive DNA variant resource EyeSeq that focuses on these diseases and enables phenotype and genotype integration. The disease portal organizes diseases in MESH hierarchical tree and links to curated phenotype and genomic data in a single place. EyeSeq interlinks eye diseases and phenotype dictionaries from internal symptom dictionary, OMIM, RetNet, ClinGen, CTD Base, and HPO. The pathogenic variants are compiled from ClinVar, HGMD, Ensmebl phenotype, and the community sources, and supplemented with in-house pathogenic variant data, as well as control data sets for easier interpretation. They include over 17,000 variants, including 7,000 open to the general public, in 888 genes that are currently associated with 446 diseases as defined by all related clinical terms. A web-based diagnosis tool is developed to aid such interpretation by extracting data from the EyeSeq resource, while capturing newly identified pathogenic variants from diagnostic lab in return. Additionally, web-based annotation tools are created to collect community expert input, including ClinVar compatible pathogenicity data submission tools. Together, EyeSeq aims to be the community genomic resource to aid in both genetic eye disease research and diagnosis.

3149F

Significance ranking of sequence variants in Parkinson Disease variant database. *J. Vance[1], K. Nuytemans[1], L. Wang[1], G. W. Beecham[1], C. Van Broeckhoven[2].* 1) John P. Hussman Institute for Human Genomics, Miami Morris K. Udall Parkinson Disease Research Center of Excellence (UPDRCE), University of Miami, Miller School of Medicine, Miami, FL; 2) Department of Molecular Genetics, VIB and Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium.

Next generation sequencing (NGS) has significantly increased the rate of rare sequence variants (RVs) reported in PD patients. Interpretation of the functional significance of RVs in PD development can be difficult however, due to the vast amounts of data and disparate data sources. Current available databases do not address the contribution of variants to PD pathogenesis and/or do not accommodate NGS data. We designed the PD variant database to address these needs. Variants are extracted from the literature (known and candidate genes) and existing NGS datasets (public databases or through collaboration). The variants are evaluated in each of three evidential categories by using predefined criteria that will be manually curated: 1) “Population Genetic evidence” (e. g. population frequency, family segregation) and 2) “Functional Genetic evidence” (e. g. binding specificity, RNA expression) extracted from literature and data repositories and 3) “*in-silico* evidence” (e. g. PolyPhen2 for missense, ESEfinder for splice, RegulomeDB and GWAVA for non-coding variants, conservation measures). Each variant will then be placed in one of six ranks, based on the strength of evidence from the three categories [Table 1]. For each variant, summary data across the available datasets will be included rather than individual level data. Currently, data on all variants reported in all known PD genes in literature or identified in the WES at Miami UPDRCE have been imported in the database (~674k variants). All ranks and supporting data will be available to the user through the website (launch mid2015, online by ASHG meeting 2015). Additional summary data for each dataset and variant (e. g. QC, frequency) will be linked to the contributing lab’s info or publication to facilitate collaborative efforts. This database will allow users to quickly evaluate variants when interpreting sequence data, leading to a more focused follow-up in a research or clinical setting. The summary data format avoids the presence of individual identifiers, facilitating rapid availability of NGS data to the PD research community and thus potential collaborative efforts.

Table 1. Rankings for variants

	Rank 1	Rank 2	Rank 3
Genetic support	Highest	High	High
In-silico support	High	High or Intermediate	High or Intermediate
Functional support	Any	High	Intermediate or Low
	Rank 4	Rank 5	Rank 6
Genetic support	Intermediate	Low	Low
In-silico support	Any	High	Intermediate or Low
Functional support	Any	Intermediate or Low	Low

3150T

Distribution and clinical impact of functional variants in 31,000 whole exome sequences from the DiscovEHR study. *F. Dewey[1], D. J. Carey[2], L. Habegger[1], S. Balasubramanian[1], J. B. Leader[2], S. N. Fetterolf[2], C. Van Hout[1], C. O'Dushlaine[1], S. A. Pendergrass[2], H. L. Kirchner[2], O. Gottesman[1], J. Staples[1], J. Packer[1], C. Gonzaga-Jauregui[1], A. Lopez[1], J. D. Overton[1], S. Mellis[1], A. Murphy[1], A. Economides[1], R. Phillips[1], N. Stahl[1], G. D. Yancopolous[1], F. D. Davis[2], W. A. Faucett[2], J. G. Reid[1], M. D. Ritchie[2], M. F. Murray[2], A. R. Shuldiner[1], A. Baras[1], D. H. Ledbetter[2].* 1) Regeneron Pharmaceuticals, Inc. , Tarrytown, NY; 2) Geisinger Health System, Danville, PA.

The DiscovEHR collaborative project between the Regeneron Genetics Center and Geisinger Health System MyCode community health initiative aims to catalyze genomic discovery and genomic medicine by coupling high throughput sequencing to a large, integrated healthcare population utilizing longitudinal electronic health records (EHR). All participants are consented for re-contact phenotyping and return of clinically-actionable results. Here we describe initial insights from deep whole-exome sequencing of over 31,000 participants. We sequenced the exons of 18,951 protein-coding genes in 31,058 participants of predominantly European ancestry with clinical phenotypes described in the EHR. The median duration of EHR data in these participants was 14 years, during which a median of 87 clinical encounters, 687 laboratory tests and 7 procedures were captured. In these participants we found ~5. 8 million single nucleotide variants, ~700,000 insertion deletion events, and ~180,000 putative loss of function (LoF) variants in the protein coding regions of the genome. The overwhelming majority of these genetic variants were rare (MAF < 1%) and over half were singletons. At this sample size, over 90% of genes, including genes encoding existing drug targets and a list of 76 genes (56 designated by ACMG plus 20 additional) conferring risk for actionable, highly penetrant genetic diseases, harbor rare heterozygous putative LoFs. Over 15% of genes contain rare homozygous putative LoFs. Each participant harbored a median of 31 rare putative LoFs. Linking these data to EHR-derived clinical phenotypes, we highlight examples of protective and harmful clinical associations that support or invalidate therapeutic targets. In individuals harboring deleterious variants in the clinically actionable gene list, we highlight examples of clinical phenotypes derived from the electronic health record that reflect this genetic disease predisposition, and outline plans to return and act clinically on these findings. In a large collection of exome-sequenced participants with longitudinal EHRs, we identified hundreds of thousands of protein-disrupting alleles and provide concrete examples of the use of this data for genetic association discovery, identification and genetic validation of therapeutic targets, and use of sequencing data in clinical care. This collaboration will serve as a blueprint for large-scale genomic medicine initiatives such as the Precision Medicine Initiative.

3151F

Analysis of structural variants in 2,200 whole-genome sequenced myocardial infarction cases and controls. *E. M. Schmidt[1], J. Chen[2], O. L. Holmen[3], K. Hveem[3], R. E. Mills[1], H. M. Kang[1,4], C. J. Willer[1,2].* 1) Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; 2) Department of Internal Medicine, Division of Cardiovascular Medicine, University of Michigan, Ann Arbor, MI; 3) HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger, Norway; 4) Department of Biostatistics and Center for Statistical Genetics, School of Public Health, University of Michigan, Ann Arbor, MI.

Myocardial infarction (MI) is a major cause of death throughout the world, with both common and rare single genetic mutations contributing to early-onset risk. Discovery of large structural variants (SVs) will give a more complete understanding of the genetic etiology of this complex disease. We perform whole-genome sequencing at 5x coverage of 2,200 individuals from Nord-Trøndelag, Norway as part of the population-based HUNT study. Cohort participants include cases with early-onset MI, and 1:1 age- and sex- matched healthy controls. We hypothesize that there are different frequencies of large structural variants in MI cases compared to controls and apply established and complementary SV detection algorithms to identify deletions, duplications, inversions, and translocations. Using the non-redundant union set of calls, we test for differences in both the prevalence and distribution of large structural variation among individuals with MI compared to their matched controls. In addition to analyzing association of genotypes with MI status, we examine the relationship between SVs and single nucleotide variants associated with coronary artery disease (CAD) and related traits identified by genome wide association studies. A top association shows a 505bp deletion near a CAD-associated collagen locus, which is a known structural component of the inner surface of blood vessels. We also investigate novel or Norwegian-specific SVs not present in the 1000 Genomes Project. Structural variants identified here will be imputed into a larger HUNT cohort of 30,000 samples for further study. This research provides broader insight into the genetic architecture underlying myocardial infarction and other heart disease-related clinical phenotypes.

3152T

Structural variants and obesity: Analysis of *FTO* intron 1 enhancer. *S. S. Sundseth[1], M. W. Lutz[2], R. Saul[3], R. McPherson[4], R. Dent[4], A. D. Roses[1,2].* 1) Cabernet Pharmaceuticals, Chapel Hill, NC, USA; 2) Deptartment of Neurology, Duke University Medical Center, Durham, NC, USA; 3) Polymorphic DNA Technologies, Alameda, CA, USA; 4) University of Ottawa Heart Institute, Ottawa, ON, Canada.

Over the last decade, genome wide association studies (GWAS) identified many loci associated with obesity, however the precise causal variants and the molecular mechanisms underlying those genetic associations remain largely unknown. In the post-GWAS era, we have concentrated on the identification of causal genetic factors. There is now evidence that many important causal variants for human traits are caused by structural variants (SVs) rather than SNPs, and that SVs are located in non-coding intronic and intergenic regions that affect gene regulation. Our strategy is to (1) use bioinformatic tools to identify non-coding regulatory genomic regions likely to be involved with obesity, (2) create and annotate our database of SVs from which to compile lists of SV targets within such regions, (3) apply proprietary scoring methods to identify those SVs most likely to have a biological effect, (4) analyze the top-scoring SVs in DNA samples from obese case and lean control subjects, (5) use phase-defined DNA sequencing and phylogenetic analysis to characterize the inheritance patterns of the variants. The approach has been successful for neurodegenerative diseases (Alzheimer's Disease, Lewy Body Variant-AD and Parkinson's Disease), but this is our first application to a "non-neurological" disease. We focused on the 47kb enhancer region within intron 1 of *FTO* that contains several SNPs previously associated with BMI and fat mass in multiple GWAS. We examined the 47kb *FTO* enhancer for SVs, scored them for putative function, and then sequenced the 4kb CUX1 element within the 47kb enhancer in Caucasian obese case/lean control subjects. To test for interactions between variants that may have an association with phenotype, recursive partitioning and statistical modeling was done. The analysis was done twice: (1) using case/control status as the dependent variable and (2) using the BMI measurements as an interval dependent variable. In this study one SNP, but no SVs were associated with BMI in the CUX1 region. We are currently analyzing additional SVs in the 47kb *FTO* enhancer for association with BMI and related phenotypes (waist circumference, fat mass, hyperlipidemia, T2D.) Identifying functional genetic variants involved in obesity and related phenotypes can lead to clinically relevant applications including earlier diagnosis, accurate prognosis, enhanced understanding of mechanism(s) of action and new therapeutic strategies.

3153F

***COL4A3* Mutations Cause Focal Segmental Glomerulosclerosis.** *L. Hu[2], J. Xie[1], X. Wu[2], H. Ren[1], W. Wang[1], Z. Wang[1], X. Pan[1], X. Hao[1], J. Tong[1], J. Ma[1], Z. Ye[3], G. Meng[4], Y. Zhu[2], K. Kiryluk[5], X. Kong[2], N. Chen[1].* 1) State Key Lab of MedicalGenomics, Ruijin Hospital, Shanghai, China; 2) State Key Laboratory for Medical Genomics, Institute of Health Sciences, Shanghai Jiao Tong University School of Medicine (SJTUSM) and Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS), Shanghai, China; 3) Nephrology Department, Huadong Hospital Affiliated to Fudan University, Shanghai, China; 4) State Key Laboratory for Medical Genomics, Shanghai Institute of Hematology, Rui Jin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China; 5) Division of Nephrology, Department of Medicine, College of Physicians and Surgeons, Columbia University, New York, USA.

Focal segmental glomerulosclerosis (FSGS) is a histologically identifiable glomerular injury often leading to proteinuria and renal failure. To identify its causal genes, whole-exome sequencing and Sanger sequencing were performed on a large Chinese cohort that comprised 40 FSGS families, 50 sporadic FSGS patients, 9 independent autosomal recessive Alport's syndrome (ARAS) patients, and 190 ethnically matched healthy controls. Patients with extrarenal manifestations, indicating systemic diseases or other known hereditary renal diseases, were excluded. Heterozygous *COL4A3* mutations were identified in five (12. 5 %) FSGS families and one (2 %) sporadic FSGS patient. All identified mutations disrupted highly conserved protein sequences and none of them was found in either public databases or the 190 healthy controls. Of the FSGS patients with heterozygous *COL4A3* mutations, segmental thinning of the glomerular base membrane (GBM) was only detected in the patient with electronic microscopy examination results available. Five ARAS patients (55. 6 %) had homozygous or compound-heterozygous mutations in *COL4A3* or *COL4A4*. Serious changes in the GBM, hearing loss, and ocular abnormalities were found in 100 %, 80 %, and 40 % of the ARAS patients, respectively. Overall, a new subgroup of FSGS patients resulting from heterozygous *COL4A3* mutations was identified. The mutations are relatively frequent in families diagnosed with inherited forms of FSGS. Thus, we suggest screening for *COL4A3* mutations in familial FSGS patients.

3154T

Long single-molecule RNA sequencing detects novel variants in humans. *H. Yang[1], S. Mangul[1], F. Hormozdiari[1], E. Tseng[2], A. Zelikovsky[3], E. Eskin[1].* 1) Department of Computer Science, UCLA, Los Angeles, CA; 2) Pacific Biosciences, Menlo Park, CA; 3) Department of Computer Science, Georgia State University.

RNA sequencing is essential for understanding biological processes but current method such as short-read sequencing is not able to fully describe an entire RNA molecules. Previously long-read sequencing was infeasible due to high error rate which disables identification of genetic variants. However, single-molecule long-read RNA sequencing allows to sequence complete haplotype isoforms, which allows parental haplotype identifications. As short-read sequencing relies on statistical analysis on the covered regions, it introduces a significant bias toward higher mapping rates for of the reference allele at the heterozygous loci and disregards the variants in minor alleles that are expressed in low level. HapIso (Haplotype-specific Isoform reconstruction) tolerates high error-rate of long-read sequencing and identifies partition the isoform reads into parental alleles. Phasing the reads according to the allele of origin allows our method to efficiently distinguish between read errors and the true biological mutation. HapIso uses a k-means clustering algorithm to group the reads into two meaningful clusters, maximizing the similarity of the reads within cluster, and minimizing the similarity of the reads from different clusters. Each cluster corresponds to a parental haplotype. We use family pedigree information to evaluate the approach. Furthermore, single-molecule sequencing was able to find novel variants that were not able to be identified in short-read sequencing. 2755 variants were identified through single-molecule sequencing that were not identified by short-read sequencing, while 1323 variants were identified in both single-molecule sequencing and short-read sequencing; the newly identified variants were confirmed by family based validation using trio family data.

3155F

Survey of Human Cytomegalovirus Gene Polymorphism in Infants Infected Congenitally or Postnatally from Gunagzhou, China. *H. Li[1], B. WANG[1], Z. DENG[1,2], T. ZHOU[1].* 1) Dept. Biotechnology, Jinan University, Guangzhou, Guangdong, China; 2) Psychiatry and Behavioral Sciences Keck School of MedicineUniversity of Southern California, 1501 San Pablo Street, Los Angeles,CA.

Human cytomegalovirus (HCMV) is the leading cause of intrauterine viral infection, affecting 0. 3–0. 7% of live births in industrialized countries and 1–5% in developing countries. CMV may cause pregnancy complications such as intrauterine growth restriction and birth defects. Congenital CMV (cCMV) infection may result in spontaneous abortion, stillbirth, microcephaly, hepatosplenomegaly, jaundice, petechiae, or chorioretinitis Approximately 10% of newborns infected congenitally are symptomatic at birth; if symptoms of CMV infection are present, they are frequently non-specific for CMV infection. When disabilities, such as hearing loss, mental retardation, or visual impairment, become apparent in next months or years, it is generally too late to identify cCMV infection as the culprit. Both host and viral factors may affect the outcome of infection. CMV virulence may depend on genetic variability in several regions of the genome. The CMV genome, the largest of all human herpesviruses, is a linear 235–240 kbp double- stranded DNA and contains more than 200 open reading frames (ORFs). In this study, we collected the urine samples of 62 newborn infants with congenital or postnatal CMV infection, who represented disease outcome and were born in Guangzhou, China. Four low passage HCMV clinical strains, D3, D2, D52, C30, were successfully isolated and identified from these urine samples. CMV polymorphisms were studied in UL133-UL148 genes from these clinical strains. Genotyping was performed by a sequencing analysis of PCR-amplified fragments. The results demonstrated that (1) there are abundant polymophism in UL133-UL148 genes from D3, D2, D52, C30, comparing with other HCMV clinical isolates; (2) UL136-UL148 genes are not essential for viral reproductivity and infection in vitro. Further more, we will establish which genotypes in several genes associated with immunodeficiency might play a role in congenital and postnatal CMV infections in infants, and attempted to explore the relationship among genetic variations, disease outcome and viral gene polymorphisms.

3156T

STR-Seq: a next generation technology for massively parallel microsatellite sequencing and genotyping. *G. Shin[1], S. Grimes[2], B. Lau[2], H. Lee[1], E. Hopmans[2], H. P. Ji[1,2].* 1) Division of Oncology, Stanford University School of Medicine, Stanford, CA; 2) Stanford Genome Technology Center, Stanford University, Palo Alto, CA.

Short tandem repeats (STR), otherwise known as microsatellites, are multiallelic in terms of germline variation. As a result, STR genotyping is highly informative for genetic applications such as forensics DNA fingerprinting. In addition, STRs are prone to accumulating somatic mutations in cancers that lack DNA mismatch repair mechanism and are informative for clonal analysis. Despite their wide application in a variety of fields, the analysis of STRs with next generation sequencing-based (NGS) methods is limited by several major issues including: i) Only the reads which encompass an entire STR locus are informative; ii) PCR amplification during library preparation can introduce artifactual "stutter" mutations that confound accurate genotyping. Consequently, analysis on STRs requires more sequencing depth with finely controlled target selection, but current target enrichment methods such as bait-hybridization are of limited utility. In this study, we developed a novel targeted sequencing technology, short tandem repeat sequencing (STR-Seq), by which a novel sequencing technology can generate STR-spanning reads for thousands of microsatellites. This technology is highly scalable and eliminates traditional sequence alignment. This technology involves programming an Illumina sequencing flowcell to select STR / microsatellite loci targets in a massively parallel fashion. Unlike conventional traditional bait hybridization or PCR amplification methods, targeting primer-probes are engineered into a flowcell, which maximize the efficiency and specificity of target selection. We demonstrate simultaneous analysis of approximately 2,500 STR loci in a single lane with individual reads cover the entire STR. We also developed a targeted fragmentation process to improve target selection and coverage. In addition, to eliminate stutter noise, sequencing libraries were prepared by an amplification-free method, thus each read corresponds to a single DNA molecule. Finally, we developed a novel bioinformatics pipeline for quantitation of STR repeat motifs and associated variants in phase with the STR. The process has two internal validation steps for high-confidence genotype calling, which use the target capture sequence and flanking sequences of the repeat regions as the validation targets. We demonstrate that this technology has concordance with conventional methods with dramatically higher throughput.

3157F

The importance of heterozygous protein truncating variants in the human genome. *A. Telenti[1], I. Bartha[2], A. Rausell[3], P. McLaren[2], P. Mohammadi[4], M. Tardaguila[3], N. Chaturvedi[2], J. Fellay[2].* 1) J Craig Venter Institute, La Jolla, CA; 2) École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland; 3) Swiss Institute of Bioinformatics, Lausanne, Switzerland; 4) Computational Biology Group, ETH Zurich, Switzerland.

Sequencing projects have identified large numbers of rare stop-gain and frameshift variants in the human genome. As most of these are observed in the heterozygous state, they test a gene's tolerance to haploinsufficiency and dominant loss of function. We analyzed the distribution of truncating variants across 16,260 protein coding autosomal genes in 11,546 individuals. We observed 39,893 stop-gain and frame-shift variants affecting 12,062 genes, which significantly differed from an expectation of 12,916 genes under a model of neutral *de novo* mutation ($p<10-4$). This suggests that there is a set of genes in the human genome that cannot tolerate heterozygous protein truncating variants (PTVs) because of early life lethality. Genes lacking any PTVs have fewer paralogs, are more likely to be part of protein complexes and are more connected in protein-protein interaction networks than the rest of the genes. Furthermore, they have characteristics of essentiality and a higher probability of CRISPR-Cas9 editing compromising cell viability. The set of genes not carrying stop-gain or frameshift variants is enriched in OMIM genes annotated with 'haploinsufficient' or 'dominant negative' keywords and are overrepresented in functional categories such as transcription regulation, developmental processes, cell cycle, and nucleic acid metabolism. Assuming that a fraction *f* of genes do not carry *de novo* PTVs while the remaining genes do so according to their neutral mutation rates, *f* can be estimated by fitting a model to the observed distribution of PTVs. By doing so we estimated that at least 10. 8% of human genes do not tolerate heterozygous PTVs. The study of PTVs delineates the essential genome and, more generally, identifies rare heterozygous variants as an unexplored source of diversity of phenotypic traits and diseases.

3158T

Whole genome sequencing of six baboon species reveals high levels of polymorphism including extensive functional variation. *M. Raveendran[1], R. A. Harris[1], M. Dahdouli[1], D. Rio Deiros[1], C. J. Jolly[2], J. Phillips-Conroy[3], L. Cox[4], D. Muzny[1], K. Worley[1], R. A. Gibbs[1], J. Rogers[1].* 1) Baylor College of Medicine , Houston, TX; 2) New York University, New York, NY; 3) Washington University School of Medicine, St. Louis, MO; 4) Texas Biomedical Research Institute, San Antonio, TX.

Baboons, genus *Papio*, are a commonly used research model for studies of various human diseases and health-related phenotypes. One baboon species, the olive baboon (*Papio anubis*) is widely used in research related to obesity, diabetes, cardiovascular disease, atherosclerosis, neuroimaging and epilepsy. To facilitate genetic analyses of risk factors for disease, we generated a de novo reference assembly for the olive baboon genome, and sequenced four additional *P . anubis* individuals. The Illumina HiSeq 100bp reads were mapped to the baboon reference, Panu_2. 0 (http://www. ncbi. nlm. nih. gov) and the single nucleotide variants (SNVs) called using both Atlas2 and GATK. Among these five individuals, we identified 5,840,388 SNVs and using the Ensembl variant effect predictor tools found 27,743 mis-sense variants, 299 de novo stop codons and 5,738 splice site variants. There are five other recognized species of *Papio* baboons, some of which are also used in biomedical research studies: *P. hamadryas* (osteoporosis), *P. cynocephalus* (atherosclerosis, cardiovascular disease, infectious disease), *P. ursinus* (bone formation) and *P. papio* (epilepsy). Although closely related and known to form hybrids in the wild, the species within this genus differ in social behavior, ecology, body size and other fundamental characteristics. To investigate genetic differences among species, as part of the baboon genome analysis consortium, we sequenced (~25x Illumina whole genome coverage) two individuals representing *P. hamadryas* (Ethiopian wild caught), two *P. papio* (Guinea), two wild-caught *P. cynocephalus* (Yellow), two *P. ursinus* (Chacma) and three wild-caught *P. kindae* (Kinda) baboons. Again the intersection of SNVs calls (16,524,065) from Atlas2 and GATK were used for subsequent analyses. We identified 4. 5M SNVsin *P. hamadryas*, 6. 5M in *P. cynocephalus*, 3. 2M in *P. papio*, 4. 8M in *P. ursinus* and 8. 2M in *P. kindae*. The average SNVs per individual ranges from 2. 8M (Guinea baboons) to 5. 2M (kinda baboons). We identified 51,188 missense, 551 stop gained, and 9,212 splice region variants. Overall, 7,453 variants were scored by Polyphen2 as probably damaging and SIFT as deleterious. The extensive genetic variation, much of which is expected to have functional consequences, can be used to extend baboon models of human health and disease by more extensively incorporating genetic variation into research study design and analysis.

3159F

Calling mitochondrial DNA (mtDNA) variants from whole exome sequencing data. *P. Zhang[1], H. Ling[1], K. Hetrick[1], E. Pugh[1], D. Witmer[1], J. Ding[2], N. Sobreira[3], D. Valle[3], K. Doheny[1].* 1) Center for Inherited Disease Research, (CIDR), Johns Hopkins University, Baltimore, MD; 2) Laboratory of Genetics, National Institute on Aging, NIH, Baltimore, MD; 3) Institute of Genetic Medicine, Johns Hopkins University School of Medicine.

As part of the Baylor Hopkins Center for Mendelian Genomics (CMG) (http://www. mendelian. org/), CIDR has performed whole exome sequencing (WES) for over 900 samples to discover the genetic basis for Mendelian conditions. When potential causal variants are not identified from coding regions in the WES data we need to look beyond the exome. In this project, we examine whether the off-target reads of mitochondrial DNA (mtDNA) generated by WES can be used to study mtDNA variation. As many large-scale genetic studies are collecting WES data, the method and protocol developed by us could have wide application. We extracted off-target mtDNA reads from our WES data and tried to identify homoplasmies (positions with all the same non-reference alleles) and heteroplasmies (positions with a mixture of two or more alleles) using MitoCaller [Ding et al. PLoS Genetics (2015), in press], a software designed to call both homoplasmies and heteroplasmies from next-generation sequencing data. We observed a mtDNA median depth of coverage (DOC) ranging from 3x to 500x in the 767 samples, out of which 21. 5 % had a median DOC >100x. The DOC varied significantly by DNA source. We used Agilent SureSelect HumanAllExonV4 capture and sequenced on Illumina HiSeq2500 platform. The mean on-target (exonic) DNA coverage in these samples ranged from 50x to 322x with a median DOC of 93x (percentage DOC on target at least 10x ranges from 85. 9% to 99. 6%). After applying MitoCaller to our WES data, we called 0 to 8 (median 1) heteroplasmies per sample for individuals with a median mtDNA DOC > 100x, using a 10% minor allele fraction threshold. As expected, analyses using trios showed high maternal and low paternal inheritance of homoplasmy. We also annotated the called mtDNA variants with ANNOVAR to assess their functional implications with the phenotype of interest. We will further evaluate how the DOC and DNA sources affect the quality of homoplasmy and heteroplasmy calls, and continue to search for variants that are associated with the Mendelian traits combining the other information available.

3160T

Sequencing and typing of HLA class I genes by using a single-molecule, real-time sequencing technology. *Y. Kuroki[1,2], N. Nariai[1], A. Ono[1], K. Tateno[1], N. Inagaki[1], S. Saito[1], M. Nagasaki[1], J. Yasuda[1], M. Yamamoto[1].* 1) Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Miyagi, Japan; 2) Genome Medicine, National Center for Child Health and Development, Tokyo, Japan.

Human leukocyte antigen (HLA) is the human version of the major histocompatibility complex (MHC), which has important functions in cellular interactions and immune reactions. The human HLA region consists of more than 250 genes, which are arranged as duplicated and clustered structures spanning of 3. 78 Mb genomic region on the chromosome 6p21. The genes located in this region are known as the susceptibility genes for various diseases, such as the common diseases and the autoimmune disorders. HLA genes play a critical role in the occurrence and the development of these diseases; however the molecular basis and the detailed relationship between the genes and HLA-related disorders have not been fully explained. One of the reasons for this is the complexity of the genome structures–in other words, the diverse sequences and different structures known as the polymorphisms within the population. To reveal the polymorphisms and the types of each allele, here we have sequenced the HLA class I genes (HLA-A, HLA-B, and HLA-C) by using single-molecule, real-time (SMRT) sequencing technology. The sequence data obtained by SMRT sequencing were subjected to computational typing for each allele. We analyzed more than 200 Japanese samples belonging to the 'ToMMo reference sequence panel', which form part of the sequence data from the prospective genome cohort study by Tohoku Medical Megabank Organization (ToMMo). Currently, we have identified several candidates for novel sequences that do not match the sequences already deposited in the IMGT/HLA database. We will present an overview of this project and report its current status at the meeting.

3161F

Linking Protein Annotation and Genome Variation in ClinGen Knowledge Base. *P. McGarvey[1], M. Harris[1], S. Rao[1], X. Feng[2], R. Patel[2], A. Milosavljevic[2], S. Madhavan[1].* 1) Innovation Center for Biomedical Informatics, Georgetown University Medical Center, Washington DC; 2) Bioinformatics Research Laboratory, Department of Molecular & Human Genetics, Baylor College of Medicine, Houston, Texas.

Gene variants may cause deleterious effects via many mechanisms including: gene transcription & splicing; protein translation; and disruption of protein structure & function. Understanding the effects of polymorphisms on protein function is a key component of gene & variant curation but much of this information is not readily available. UniProtKB contains decades of literature-based and semi-automated curation describing protein function including variation. Standard NGS annotation tools identify simple effects like translation stops & missense mutations, but more subtle effects like disruption of an enzyme active site, protein-binding site or post-translational modifications are often not included. There is a need to make this valuable functional information available to the genomic community to connect DNA variation to the protein and the biological consequence. To integrate UniProtKB, information needs to be mapped from the multi-dimensional space of protein structural features into the coordinate space of transcripts and genomic sequences. We map protein features to the genome using software developed to integrate information from UniProtKB and RefSeq. Each UniProt sequence is mapped to a Refseq protein entry. The corresponding RefSeq RNA location on the genome including the intron/exon boundaries and transcription start site is used to map the amino acid sequence to the genome. Splicing introduces complexity as functional domains on an amino acid sequence and even individual codons may have multiple locations on the genome due to intron splicing. The software outputs protein to genome coordinates in tabular, and JSON formats. The JSON output is used to load the data into ClinGen Knowledge Base via APIs using a document model created for protein sequence features. Annotated protein variations are loaded as simple allele documents and linked to related documents on gene, reference sequences, protein features and evidence to be used in the ClinGen Pathogenicity Calculator. Results from a test set of genes will be presented and illustrated by examples where genomic variants are aligned with UniProt features such as Active Sites, Domains and other features that could explain a variant's pathogenicity. Also shown are novel disease associated protein variations with published evidence not currently in ClinVar. The results suggest that this integration can help the classification of pathogenic variants that ultimately will support decisions on actionability.

3162T

Somatic mutation analysis in a healthy adult's blood and sperm. *C. Zeng[1], Y. Hong[1], W. Chen[1], D. Zhang[1], Y. Fu[2].* 1) Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China; 2) Human Genetics Center and Division of Biostatistics, School of Public Health, University of Texas, Houston, TX.

Somatic mutations commonly occurred in cells during the DNA replication and DNA repair process. Although studies on somatic mutations were focused on their association with diseases, this type of mutations was continually generated in our bodies and most of them were neutral as inferred by current research without exact test. Studying somatic mutations in normal tissues will give us insight into the mutation hot spots, the accumulation of mutations and their impacts in cell development. However, a big challenge to study somatic mutations is their very low frequency that are hardly to be distinguished from the background noise of next generation sequencing. Here we developed a method by estimating the error probability to detect low frequency somatic mutations. With this approach we also analyzed the dynamics of mutation distributions in blood and sperm sampled from an adult. To filter the errors produced in the process of whole exome sequencing, a probability based error estimating model was constructed using a priori error distribution generated by control sequences with no mutation. The probability of a call as to be a mutation or an error was then calculated accordingly. Applying this model, hundreds of probably true mutations were found in each tissue. Similar mutation numbers were seen in two tissues, whereas the distributions of the mutation frequency spectrums between blood and sperm appear to be different. Greater number of high frequency mutations were detected in blood, perhaps because of large effective cell population size in early stage of development and higher number of cell divisions in blood than that in sperm. Interestingly, the observed spectrums in both tissues showed significant bias towards low frequency mutations in comparison with the simulation of mutation generation starting from one cell. One possibility for this phenomenon is the change in mutation rate during the process of cell development. Moreover, in five time-point samples, three from blood and two from sperm with the 12 month interval for each tissue, only the very high frequency somatic mutations were shared among samples. However, for the hundreds of low frequency mutations, only one or two were overlapped in either blood or sperm. This is most likely due to the different lineages of active hemopoietic stem cells sampled in time points, or a relatively large population size in sperm.

3163F

The Korean Reference Genome project: construction of the reference panel for imputation analysis. *Y. Kim, B. Kim, B. Han.* Center for Genome Science, National Institute of Health, Heungdeok-gu, Chungcheongbuk-do, South Korea.

Next Generation Sequencing (NGS) technology is becoming powerful tool for discovery of genetic variants across entire chromosomes in genomic study. However, NGS is not yet feasible to apply in a large scale population based genome study due to its relatively high cost and required high computing power. In this context, imputation analysis gathered much attention as an alternative tool for comprehensive analysis of all genetic variants that are estimated using sequencing data as the reference panel. The 1,000 Genomes project data is widely used as a public reference panel. There are about 2,500 samples from 20 populations in the phase 3 dataset. Among them, 504 samples with East Asian ancestry were included. However, Korean samples were not included in the 1,000 Genomes project. Geographically, the Republic of Korea is located between China and Japan in North-East Asia. In this study, we describe the Korean Reference Genome (KRG) project which was initiated in 2012. We sequenced 400 Korean samples (10-30x) and constructed the study specific reference panel containing about 20 million single-nucleotide variants. We demonstrated that the KRG reference panel increased imputation performance over the 1,000 Genomes reference panel. The KRG reference panel will be valuable resource and would be great aid for imputation based association analysis in the Korean population.

3164T

Association study of *TAF2* with aspirin exacerbated respiratory disease and FEV1 decline. *J. Kim[1], C. Park[2], H. Shin[1,3].* 1) Research Institute for Basic Science, Sogang University, Seoul, South Korea; 2) Division of Allergy and Respiratory Medicine, Soonchunhyang University Bucheon Hospital, Bucheon, South Korea; 3) Department of Life Science, Sogang University, Seoul, South Korea.

Aspirin exacerbated respiratory disease (AERD) is a clinical syndrome characterized by a severe decline in forced expiratory volume in one second (FEV1) when taken with non-steroidal anti-inflammatory drugs (NSAIDs) such as aspirin. Our previous genome-wide association study (GWAS) has identified several variants, including a nonsynonymous variant, of *RNA polymerase II, TATA box binding protein (TBP)-associated factor* (*TAF2*, a cofactor for TFIID-dependent RNA polymerase II-mediated transcription) as potential susceptible loci for AERD. To validate the potential association between *TAF2* and AERD, we performed an extended study using additional single nucleotide polymorphisms (SNPs), different genetic models, and haplotype analysis in a larger cohort. In the logistic analysis, *TAF2* polymorphisms showed potential association signals for the development of AERD ($P < 0.05$). In addition, regression analysis showed significant associations between *TAF2* polymorphisms and FEV1 decline after aspirin provocation in asthmatics ($P < 0.05$). Despite the needs for replications in other populations and further functional evaluations, our findings suggest that *TAF2* polymorphisms might be associated with AERD and FEV1 decline, or its related symptoms.

3165F

An integrative computational workflow for eQTL analysis. *N. Prodduturi, C. Wang, H. Sicotte, JP. Kocher.* Department of Health Sciences Research,Mayo Clinic, 200 first street SW, Rochester, MN 55905 Rochester, MN.

Background: Genome-wide association studies (GWAS) have led to the discovery of many variants associated with risk, prognosis, outcome, and pharmacogenomics drug response. Even after the functional variant is identified by fine mapping, the functional mechanism of action of this variant remains to be elucidated. Currently, only 20% of variant-phenotype associations can be explained by amino acid changes. The remaining association involve variants located in non-coding regions that could be functional by affecting transcripts regulation. If this assumption is correct, an association between genotype and gene expression could be established using expression quantitative trait loci analysis(eQTL). This analysis can either detect cis-eQTLs, that have direct influence on regulatory elements affecting genes near the DNA variant, or trans-eQTLs that affect distal genes or genes on a different chromosome. Integrative eQTL analysis can therefore help explain the functional role of clinically relevant SNPs. Despite the availability of tools to perform eQTL computation, a comprehensive workflow is still lacking. **Results:**We developed an easy to use workflow consisting of powerful tools in public domain, such as MatrixEQTL, plink, and ANNOVAR. In addition, we designed additional components to reduce false positives and facilitate results interpretation. This workflow is portable and can be used on single machines or computer clusters, using the Sun Grid Engine (SGE) cluster management software. Genotyping and expression data with the genomic coordinates and strand information are the required input files for the workflow. The package provides multiple options to customize eQTL analysis: choosing different statistical methods, running permutation based hypothesis-testing, and retrieving flexible forms of annotation tables. Moreover, the workflow provides users with additional reports, which include input data QC reports, eQTL results with FDR corrected p-values, annotations, fold changes, and information about correlated SNPs in blocks of linkage-disequilibrium (LD). Collectively, this versatile workflow is helpful to facilitate interpretation of genomic variants and formulation of mechanistic hypothesis.

3166T

Fine-mapping red blood cell traits using eQTL in erythroblasts differentiated *ex vivo* from CD34+ hematopoietic stem cells. *S. Lessard[1,2], G. Lettre[1,2].* 1) Montreal Heart Institute, Montreal, Quebec, Canada; 2) Université de Montréal, Montreal, Quebec, Canada.

INTRODUCTION: A challenge arising from genome-wide association studies (GWAS) is the identification of causal relationships between variants and specific genes, since most associated variants fall outside coding regions. A powerful approach to circumvent this problem is expression quantitative trait loci (eQTL) mapping, but this requires access to relevant human tissues or cells. Red blood cell (RBC) traits are ideal for testing the eQTL method: (1) they are highly heritable (h2~0. 4-0. 8), (2) they are clinically relevant, being implicated in cancers, hemoglobin disorders and malaria resistance, and (3) RBC progenitors are accessible and amenable to *in vitro* culture. **METHODS & RESULTS:** Erythroblasts differentiated *ex vivo* from CD34+ hematopoietic stem cells obtained from 24 anonymous human donors were genotyped on the HumanOmniExpress array (Illumina). We imputed genotypes using the 1000 genomes phase 3 reference haplotypes. We measured gene expression levels by RNA-sequencing (>40M reads per sample), mapping reads and evaluating expression values using Tophat2 and Cufflinks2. We first identified genes displaying allelic imbalance (AI). For each sample at heterozygous sites, we tested if the reads alternative-to-reference allele ratio deviated from an expected ratio corrected for mapping bias using a binomial test. For each gene with AI P<5x10-8 (N=355), we mapped candidate eQTLs. First, we tested if surrounding SNPs (±100kb) were associated with the gene's expression by linear regression. Second, we tested if the candidate eQTL SNP was concordant with AI using a *t*-test, based on the hypothesis that heterozygotes should display AI, whereas homozygotes should display the expected ~50:50 ratio. We determined that both statistics are poorly correlated (R2 = 0. 01) and can thus be combined using Fisher's method to increase confidence in the results. With this approach and despite our small sample size, we identified at least one potential eQTL with false discovery rate (FDR) <0. 05 for 159 genes. We found several promising loci, including the SNPs rs1541252 and rs10900585, which are respectively associated with mean corspuscular hemoglobin concentration and malaria resistance, and are eQTLs for the *ATP2B4* gene. **CONCLUSION:** We identified several candidate eQTLs in erythroblasts, which provides a framework to fine-map RBC traits-associated variants. We aim to validate these eQTLs by engineering deletions of potential regulatory regions using the CRISPR/Cas9 system.

3167F

Identification of rare and common variants associated with gene expression in a complex Amish pedigree. *C. Brown[1], Y. Park[1], R. Kember[1], B. Emert[1], S. Elwyn[1], D. Craig[2], B. Engelhardt[3], D. Rader[1], M. Bucan[1].* 1) Genetics, University of Pennsylvania, Philadelphia, PA; 2) The Translational Genomics Research Institute, Phoenix, AZ; 3) Department of Computer Science and Center for Statistics and Machine Learning, Princeton University, Princeton, NJ.

The majority of heritability of complex traits in humans is associated with non-coding variants of modest effect size. In an effort to characterize the mechanisms by which non-coding variants contribute to phenotypic variation, numerous studies have identified genetic variation associated with gene expression (eQTLs) in a variety of tissues. While eQTLs have provided tremendous insight into genome wide association studies (GWAS) of disease, they suffer from limited statistical power because their sample sizes are often several orders of magnitude smaller than the corresponding GWAS. As a result, existing eQTL studies tend to identify genetic variants of large effect, which may miss many alleles relevant for disease. In population-based samples, power limitations become increasingly problematic with less frequent alleles. Recently, several groups have developed haplotype-based approaches that jointly model allele specific expression (ASE) and eQTL effects that increase power to detect genetic effects on expression. However, such approaches require accurate phasing between causal regulatory SNPs and coding SNPs and therefore have less benefit when these SNPs are near linkage equilibrium. To address these limitations, we have performed an eQTL and ASE study in a large (n=497), complex (9 generation) Old Order Amish pedigree, combining whole genome and RNA sequencing. We demonstrate that this study design has a number of advantages for the ascertainment of genetic variants associated with gene expression. High quality, long-range haplotypes improve the ability to jointly model QTL and ASE effects at distal regulatory elements and improve detection of allele specific transcript isoform usage. Moreover, by studying a large family from a founder population, we are able to detect associations with large effect genetic variants that are rare in European population cohorts. Finally, we use the unique pedigree structure of this family to estimate heritability components for gene expression and splicing.

3168T

Characterizing regulatory variation using a panel of induced pluripotent stem cells. *NE. Banovich[1], YI. Li[2], SM. Thomas[1], MC. Ward[1], PY. Tung[1], JE. Burnett[1], MD. Myrthil[1], CL. Kagan[1], I. Gallego Romero[1], BJ. Pavlovic[1], JK. Pritchard[2,3], Y. Gilad[1].* 1) Human Genetics, University of Chicago, Chicago, IL; 2) Departments of Genetics and Biology, Stanford University, Stanford, CA; 3) Howard Hughes Medical Institute, Stanford University, Stanford, CA.

Human induced pluripotent stem cells (iPSCs) provide unprecedented potential to study gene regulation in multiple differentiated cell types from a single individual. This tool also allows researchers to perform experimental perturbations in disease relevant tissues. A first step for the study of gene regulation in iPSCs is to generate a panel large enough to study the effects of genetic variation on gene regulation. To this end, we reprogrammed lymphoblastoid cell lines (LCLs) from 70 individuals into iPSCs. These individuals have been used extensively in previous work, making them attractive for further studies on gene regulation. In this study, we characterized regulatory variation, by collecting gene expression data and identifying expression quantitative trait loci (eQTLs). In our hands, the number of eQTLs identified in iPSCs are comparable to those identified in somatic cell types with similar sample sizes. This result was surprising given our observation that gene expression in iPSCs is highly homogeneous among individuals. Interestingly, we found that eQTL effect sizes are smaller on average in iPSCs compared to their counterparts in somatic cells. This however may be explained by standard errors of our estimated effect sizes in that are an order of magnitude lower on average in iPSCs. To further improve our understanding of regulatory variation in these lines, we measured DNA methylation and chromatin accessibility (ATAC-seq). Finally, to demonstrate the broad utility of iPSCs, we differentiated iPSCs from ten individuals into cardiomyocytes and collected RNA-seq data. Altogether we have generated a panel of iPSCs of sufficient size to identify QTLs associated with multiple regulatory phenotypes. Our initial analyses have identified eQTLs and highlighted differences between the effect of genetic variation on gene expression in iPSCs and somatic tissues. This panel will be a valuable resource in the study of gene regulation for our group and others.

3169F

"Buffet-Style" eQTL Discovery – Assessing the Impact of PEER Adjustment on Replicability of Cis and Trans eQTL Associations in Whole Blood. *P. J. Castaldi[1,2], M. Obeidat[3], M. H. Cho[1,4], E. K. Silverman[1,4], P. D. Pare[3], D. D. Sin[3], . C. P Hersh[1,4].* 1) Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA; 2) Division of General Medicine and Primary Care, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA; 3) University of British Columbia Centre for Heart and Lung Innovation, St Paul's Hospital, BC Canada; 4) Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA.

Background: It is common practice to adjust for probabilistic estimation of expression residuals (PEER) factors in eQTL analyses, based on the rationale that this eliminates systematic noise and increases power. However, PEER factors also capture biologic signal. Thus, PEER adjustment could be beneficial or detrimental to eQTL discovery, and this could vary for *cis* and *trans* eQTLs. Using two blood eQTL datasets of 130 and 490 subjects with chronic obstructive pulmonary disease (COPD), we systematically evaluated the impact of adjustment for a range of PEER factors on eQTL yield, replicability, and retention of baseline (i. e. non PEER-adjusted) eQTL. **Methods:** Gene expression data were generated for 130 subjects with COPD from the ECLIPSE Study using the HG-U133 Plus 2. 0 GeneChip (the ECLIPSE 1 sample), and additional expression data were generated for 490 non-overlapping subjects with COPD from ECLIPSE using the HumanGene ST 1. 0 chip (ECLIPSE 2 sample). In each sample, up to 80 PEER factors were generated from RMA-normalized and rank transformed expression data. ECLIPSE 2 data were analyzed both at the gene and exon levels. A series of 1,440 *cis* and *trans* eQTL analyses were performed for chromosomes 15 and 22, using MatrixeQTL with a 1MB cis window, adjusting for 0 through up to 80 PEER factors, and the impact of PEER adjustment on *cis* and *trans* eQTL discovery and replication was assessed. **Results:** PEER adjustment increased the yield of cis eQTLs, though in ECLIPSE 1 there were diminishing returns after adjustment for >15 PEER factors. ECLIPSE 1 cis eQTL associations replicated in ECLIPSE 2 at a rate ranging from 55-78% and 33-54% at the gene and SNP-gene level, respectively. Notably, 20-30% of significant non-PEER adjusted eQTLs were no longer significant after PEER adjustment, though these eQTLs replicated at a similar rate to PEER-identified eQTLs. For *cis* eQTL analysis, collecting all significant results across PEER-adjusted analyses increased eQTL yield without sacrificing replicability, after adjusting for the increased number of tests. For *trans* eQTL discovery, PEER adjustment often decreases yield. **Conclusions:** PEER factor adjustment aids *cis* eQTL discovery, but often impairs *trans* eQTL discovery. Concatenating significant *cis* eQTL results across non-PEER and PEER-adjusted analyses, i. e. "buffet-style" eQTL discovery, is a more powerful approach than the traditional method of selecting a single number of PEER factors to use for adjustment.

3170T

Local epistasis between regulatory and coding variants contributes to disease in diverse human tissues. *S. E. Castel[1,2], T. Lappalainen[1,2].* 1) New York Genome Center, New York, NY; 2) Department of Systems Biology, Columbia University, New York, NY.

Genetic variants that affect gene expression (expression quantitative trait loci, eQTL) are common in human populations and it is well known that they canassociate to disease. Interactionbetween these common regulatory variants and rare putatively deleterious protein coding variants may explain some of this association, but has not been extensively studied. Such interaction, termedepistasis, could result in the increased penetrance of acoding variant if it finds itself on the more highly expressed haplotype as a resultof acis-eQTL. In this work, we put forward a novel population genetics basedmodel that describes theepistatic risk resulting from interaction between regulatory and deleterious codingvariants within a gene. We use the model to predict eQTL frequency spectra thatwould lead to epistatic interactions within heterozygous individuals. Using datafrom the Genotype Tissue Expression (GTEx) and 1000 Genomes projects wefind a both a reduction in strength and number of eQTLs in the allele frequencyspectra where we predict negative interactions would occur, suggestive ofpurifying selection acting against negative epistasis. By integrating epistaticsignatures with gene and variant level annotations we identify a set of genes thatare at high risk for negative interactions that could result in disease. Finally, weuse the transcriptome as a molecular phenotype to illustrate the effect of thisphenomenon within individuals. Our results suggest thatlocal epistatic interactionbetween cis-eQTLs and coding variants is common, and plays arole in disease.

3171F

Induced pluripotent stem cells retain patient-specific gene expression patterns largely driven by cis-regulatory variation. *G. E. Hoffman[1], I. Carcamo-Orive[2], P. Cundiff[3], J. W. Knowles[2], N. D. Beckmann[1], S. D'Souza[3], S. Whalen[4], D. Papatsenko[1], A. Patel[3], F. Abbasi[2], G. M. Reaven[2], M. Shahbazi[2], R. Chang[1], G. Pandey[1], I. Lemischka[3], T. Quertermous[2], E. E. Schadt[1].* 1) Department of Genetics and Genomic Sciences, Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY; 2) Stanford School of Medicine, Cardiovascular Institute, Palo Alto, CA; 3) Department of Developmental and Regenerative Biology, Black Family Stem Cell Institute, Icahn School of Medicine at Mount Sinai, New York, NY; 4) Gladstone Institutes, University of California, San Francisco, CA.

Patient-derived induced pluripotent stem cells (iPSC) are a powerful *in vitro* model of human disease owing to their ability to self-renew and differentiate into other cell lineages. iPSC's can be derived from adult somatic cells and it is well established that the reprogramming process overrides the gene expression pattern of the originating cell type. Yet the degree to which iPSC's retain a patient-specific expression pattern or converge to a more homogeneous iPSC pattern remains poorly understood. Moreover, the degree to which any retained patient-specific expression pattern is driven by genetic compared to epigenetic factors remains unknown. Here we characterize the drivers of variation in gene expression in the largest study of iPSC heterogeneity to date comprising 317 RNA-Seq experiments from 101 individuals (50 insulin resistant, 51 insulin sensitive) and ~3 iPSC clones per individuals derived through a non-integrative Sendai virus methodology. Applying a novel linear mixed model approach to quantify the contribution of multiple drivers of variation in expression, we observe that cross-individual variation is the major driver of gene expression variability, while other factors such as originating cell type and reprogramming batch have only a minor effect. Thus iPSC's demonstrate strong retention of patient-specific gene expression patterns. Across all genes, a median of ~50% of expression variability is driven by cross-individual variation, yet many genes have substantially higher or lower percentages of cross-individual variation. We observe that highly variable genes are enriched for cis-eQTL's detected in this dataset, while stable genes are depleted for cis-eQTL's. This relationship is highly significant (p-value < 1x10-75 by logistic regression test) and demonstrates that much of the cross-individual variation in gene expression is driven by detectable cis-regulatory variation rather than retained epigenetic factors. This analysis represents a step towards validating iPSC's as a robust system to study disease biology and regulation of gene expression in a patient-specific genetic background. We further interpret our analysis of eQTL's and the divers of gene expression variation in iPSC's in the context of insulin biology, genome-wide association studies, GTEX and the Roadmap Epigenomics Project.

3172T

Dissecting the Genetics of the Human Transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci. *H. Kirsten[1,2,3], H. Al-Hasani[4,5,6], L. Holdt[7], A. Gross[1,2], F. Beutner[2,8], K. Krohn[9], K. Horn[1,2], P. Ahnert[1,2], R. Burkhardt[2,10], K. Reiche[5,6,11], J. HackerMøller[5,6,11], M. Löffler[1,2], D. Teupser[7], J. Thiery[2,10], M. Scholz[1,2].* 1) Medical Faculty / IMISE, University Leipzig, Leipzig, Germany; 2) LIFE - Leipzig Research Center for Civilization Diseases, University of Leipzig, Leipzig, Germany; 3) Cognitive Genetics, Fraunhofer Institute for Cell Therapy and Immunology- IZI, Leipzig, Germany; 4) Analysis Strategies Group, Department of Diagnostics, Fraunhofer Institute for Cell Therapy and Immunology - IZI, Leipzig, Germany; 5) Department for Computer Science, University of Leipzig, Leipzig,Germany; 6) Young Investigators Group Bioinformatics and Transcriptomics, Department Proteomics, Helmholtz Centre for Environmental Research - UFZ, Leipzig, Germany; 7) Institute of Laboratory Medicine, Ludwig-Maximilians-University, Munich, Germany; 8) Department of Internal Medicine/ Cardiology, Heart Center University of Leipzig, Leipzig,Germany; 9) Interdisciplinary Center for Clinical Research, Faculty of Medicine, University of Leipzig, Leipzig,Germany; 10) Institute of Laboratory Medicine, University of Leipzig, Leipzig, Germany; 11) RNomics Group, Department of Diagnostics, Fraunhofer Institute for Cell Therapy and Immunology - IZI, Leipzig, Germany.

Background and statement of purpose: Genetics of gene expression (eQTLs or expression QTLs) has proved an indispensable tool for understanding biological pathways and pathomechanisms of trait associated SNPs. However, power of most genome-wide eQTL studies is still limited. Therefore, our aim was to performed a large eQTL study in peripheral blood mononuclear cells of 2,112 individuals increasing the power to detect trans-effects genome-wide. Furthermore, going beyond univariate SNP-transcript associations, we aimed to investigate relations of eQTLs to biological pathways, polygenetic effects of expression regulation, trans-clusters, and enrichment of co-localised functional elements. **Methods:**We used Illumina HT-12 v4 Expression BeadChips and Affymetrix Axiom CEU SNP-chips Illumina-HT12-v4 expression chips imputed to HapMap2 CEU. After thorough pre-processing, 2,625,374 autosomal SNPs and 18,738 probes corresponding to 13,338 genes were included in the analysis, which was done primarily in the R language. **Summary of results:**We found eQTLs for about 85% of analysed genes, 18% of genes were trans-regulated. Local eSNPs were enriched up to a distance of 5 Mb to the transcript challenging typically implemented ranges of cis-regulations. Pathway enrichment within regulated genes of GWAS-related eSNPs supported functional relevance of identified eQTLs. We demonstrate that nearest genes of GWAS-SNPs might frequently be misleading functional candidates. We identified novel trans-clusters of potential functional relevance for GWAS-SNPs of several phenotypes including obesity-related traits, HDL-cholesterol levels, and haematological phenotypes. We used chromatin immunoprecipitation data for demonstrating biological effects. Yet, we show for strongly heritable transcripts that still little trans-chromosomal heritability is explained by all identified trans-eSNPs, however, our data suggests that most cis-heritability of these transcripts seems explained. Dissection of co-localised functional elements indicated a prominent role of SNPs in loci of pseudogenes and non-coding RNAs for the regulation of coding genes. **Conclusion:**In summary, our study substantially increases the catalogue of human eQTLs and improves our understanding of the complex genetic regulation of gene-expression, pathways and disease-related processes.

3173F

The effects of human genetic variation on the gene regulatory cascade. *Y. Li[1], B. van de Geijn[2], A. Petti[3], A. Raj[1], D. Knowles[4], J. Blischak[2], Y. Gilad[2], J. Pritchard[1,5,6].* 1) Department of Genetics, Stanford University, Stanford, CA; 2) Department of Human Genetics, University of Chicago, Chicago, IL; 3) Genome Institute, Washington University in St. Louis, MO; 4) Department of Computer Science, Stanford University, Stanford, CA; 5) Department of Biology, Stanford University, Stanford, CA; 6) Howard Hughes Medical Institute, Stanford University, CA.

Understanding the mechanisms through whichgenetic variation impacts gene regulation and complex traits such as disease risk is central to the growth of genomics in medicine. Noncoding, in particularregulatory, variation is thought to play acentral role in disease, however it is as yet unclear what are the most important mechanisms affected. Here, we answer this question by quantifying the impact of genetic variation on gene regulation and diseases by jointly mapping genetic variants that affect the levels of histone modifications, DNA methylation, transcription rate, stable mRNA levels, ribosome occupancy and protein in lymphoblastoid cell lines.

We report two major bottlenecks that reduce the effects of genetic variation on the gene regulatory cascade, which limit their functional importance: (1) up to half of all genetic variants that affect histone modification levels do not appear to affect mRNA transcription rates and (2) although the vast majority of variants that affect mRNA transcription also affect protein expression levels, their effect sizes are often partially buffered. We also found that as many as 30% of all QTLs that affect protein expression levels do not appear to affect chromatin-level traits. Instead, they tend to modulate gene expression levels directly by affecting splicing and/or RNA decay. Lastly, using a polygenic model to partition heritability between different classes of QTLs across several common immune-related diseases, we estimate that variants that percolate through the regulatory cascade tend to have larger effects on disease than SNPs that appear to be buffered.

3174T

Quantifying tolerance of genes to *cis*-regulatory variation from allelic expression data. *P. Mohammadi[1,2], S. E. Castel[1,2], T. Lappalainen[1,2].* 1) New York Genome Center, New York, NY; 2) Department of Systems Biology, Columbia University, New York, NY.

Several approaches have been presented to assess the tolerance of coding genes against functional impact of genetic variation. However, noncoding variants are also likely to play a significant role in genetic predisposition to disease, as exemplified by the overrepresentation of disease-associated genetic variants among expression quantitative trait loci (eQTL). Yet, the extent to which individual genes tolerate common and rare regulatory variation remains largely elusive, which complicates the interpretation of putative disease-causing variants discovered in growing whole genome sequencing data sets. Allelic expression analysis is an elegant integrative approach for capturing functional impact of *cis*-regulatory variation over the expressed genome and is minimally obscured by other confounding regulatory mechanisms. Here we present a generative probabilistic model to describe allelic expression data as net outcome of one or more unobserved regulatory variants. We apply the method to derive a model-based clustering for identifying intrinsic patterns of allelic expression within a given dataset. Using simulation studies, we demonstrate that the model is able to robustly capture the expected spread of gene expression within population, a quantity we use to assess total regulatory tolerance of each gene. Application of this method to allelic expression data from GEUVADIS and GTEx consortia allowed us to estimate regulatory tolerance of most coding genes. We show that our results are in high concordance with equivalent scores calculated from eQTL data ($R=.66$, $p<10\text{-}110$). While eQTL data is inherently limited to loci with common genetic variant, our approach using allelic expression data is able to capture low-frequency regulatory variants as well. We observed that genes with higher tolerance to *cis*-regulatory variation tend to have higher tolerance to coding mutation as assessed by Residual Variation Intolerance Score ($p<10\text{-}74$), and are also less likely to be haploinsufficient ($p<10\text{-}40$). We further investigate the distribution of the scores between different classes of genes. We conclude that the presented regulatory intolerance score can be useful in interpretation and follow up analysis of disease-associated variants, including prioritization of rare variants observed in whole genome sequencing data.

3175F

Characterizing and mitigating winner's curse in expression quantitative trait loci (eQTL) studies. *G. Darnell[1], J. Tung[2], C. Brown[3], S. Mukherjee[4,5], B. E. Engelhardt[6].* 1) Lewis-Sigler Institute, Princeton University, Princeton, NJ; 2) Department of Evolutionary Anthropology, Duke University, Durham, NC; 3) Department of Genetics, University of Pennsylvania, Philadelphia, PA; 4) Departments of Statistical Science, Mathematics, and Computer Science, Duke University, Durham, NC; 5) Institute for Genome Sciences and Policy, Duke University, Durham, NC; 6) Computer Science Department, Princeton University, Princeton, NJ.

In quantitative genomic studies, effect sizes are often overestimated in low-power situations, leading to ascertainment bias referred to as Winner's curse [Zöllner et al. , 2007]. Indeed, this phenomenon may explain the correlation that is often observed between minor allele frequency and effect size in both genome-wide association studies (GWAS) and quantitative trait loci (QTL) studies. In particular, statistically significant single nucleotide polymorphisms (SNPs) are enriched for low minor allele frequency SNPs and a high effect size. Yet, our ability to verify this claim via detection of *truly* causal variants remains questionable. In this work, we illustrate the prevalence of Winner's curse in eQTL studies, and we describe a simple variation on the common statistical test for association to mitigate the effects of Winner's curse in these studies. First, through extensive simulations using data from the Genotype-Tissue Expression (GTEx) v6 project data, we assess whether or not current methods commonly identify truly causal variants, and the extent to which these studies are victims of Winner's curse. With this study design, our simulations capture the LD structure present in the human genome and imputed variants as in real genomic study data. Then, we developed a Bayesian statistical framework that incorporates the expected relationship between minor allele frequency and effect size by explicitly accounting for the error inherent in estimating variant effect sizes in limited-power situations (i. e. , for low minor allele frequency SNPs). Simulations show that our method mitigates the effect of Winner's curse by correcting for the ascertainment bias that results from assessing significance using a standard t-statistic. Applied to real eQTL study data, we find that we recover different allele frequency spectrum and power spectrum results across our eQTLs, and that our eQTLs show greater enrichment with respect to cis-regulatory elements, validating our approach.

3176T

Predicting tissue-specific gene expression by leveraging the genotype-tissue expression (GTEx) datasets. *Z. Qi[1,2], Y. Guan[1,2,3].* 1) Department of Pediatrics, Baylor College of Medicine, Houston, TX; 2) USDA/ARS Children's Nutrition Research Center, Houston, TX; 3) Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX.

Tissue-specific gene expressions have direct relevance to disease phenotypes, and knowing gene expression in disease-relevant tissues is advantageous in both detecting novel associations and dissecting known genetic associations. Unfortunately, many genome-wide association study (GWAS) datasets have no companion gene expression essay, let alone tissue-specific ones. This project aims to develop novel methods that can impute tissue-specific gene expression into GWAS data by leveraging expression quantitative trait loci (eQTL) from genotype-tissue expression (GTEx) datasets. Using a full Bayesian approach, we first reproduce SNP-eQTLs by performing the single-SNP association analysis between gene-expression and genotypes in cis. Our SNP-eQTL results will be compared against those in the GTEx database. Then we perform haplotype analysis to detect hap-eQTLs, using a method that can integrate out phase uncertainty. The hap-eQTLs is complementary to SNP-eQTLs in that hap-eQTLs are often detected in regions that harbor allelic heterogeneity. For those loci that have significant hap-eQTLs but insignificant snp-eQTLs, we investigate, using RNA-seq data, whether the gene have multiple isoforms expressed. This knowledge is useful in defining gene expression unit. For each unit of gene expression, be it a gene or an isoform, we use a Bayesian variable selection procedure to combine SNP-eQTLs and hap-eQTLs. This variable selection procedure produces a posterior weighted predictive model that is more powerful than penalized regression methods such as LASSO and Elastic-Net. We produce one predictive model for each tissue-specific expression unit, and apply the model to impute tissue-specific gene expression of a GWAS dataset. We have a whole blood gene expression array data on a cohort that is similar in size with GTEx, and this cohort will be used to test our predictive models. Both our SNP-eQTL and hap-eQTL can jointly analyze multiple correlated phenotypes. Correlation of multiple measurements of the same gene will increase the power in predicting tissue-specific gene expression. The predicted gene expression, in disease-relevant tissues, can be used as a surrogate genotype to test for disease associations.

3177F

Fine mapping of QTLs with heterogeneous subgroups, as illustrated in cross-tissue and cross-population eQTL analysis. *X. Wen[1], Y. Lee[1], R. Pique-Regi[2], F. Luca[2].* 1) Dept Biostatistics, Univ. of Michigan, Ann Arbor, MI; 2) Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI, USA.

Recently, QTL studies have been extended to incorporate data from multiple potentially heterogeneous subgroups. The subgroups may represent different tissue environments, cellular conditions, population groups in different application contexts. For example, the GEUVADIS project analyzes expression-genotype data from 5 population groups;the GTEx project collects eQTL data from > 30 human tissues. In this presentation, we present an elegant integrative analysis approach that leverages QTL data collected from multiple potentially heterogeneous subgroups, and demonstrate its power using two applications of eQTL fine mapping analysis. In the first application, we analyze the GEUVADIS data aiming toidentify multiple independent cis-eQTL signals that are consistent across populations, accounting for population heterogeneity in allele frequencies and linkage disequilibrium patterns. We find that joint analysis across population groups greatly improves the power of eQTL discovery and the resolution of fine mapping of causal eQTL. We also find convincing evidence that many genes (7% of all interrogated protein coding genes and linc-RNAs) harbor multiple independent eQTLs in their cis regions. In addition, we find that many seemingly population specific eQTL signals from single SNP analysis can be explained by multiple consistent eQTLs and varying LD patterns across populations. In the second application, we showcase some interesting examples in multi-SNP fine mapping analysis of cross-tissue eQTLs using the pilot release of the GTEx data. In particular, we will highlight some interesting discoveries of multiple tissue-specific eQTLs which otherwise could be mis-interpreted by single SNP analysis. Finally, we illustrate that our analysis framework enables high-resolution functional analysis of eQTLs by integrating genomic annotations.

3178T

Partitioning of genetic variation influencing gene expressoin by chromosome. *J. Hall, W. Bush.* Institute for Computational Biology, Case Western Reserve University, Cleveland, OH.

Neither the structure of the genome nor how it's expressed are fully understood -- many factors affect gene expression, including temporal factors, epigenetic marks, genetic mutations/variation, and chromatin conformation. Gene expression is often influenced by cis regulatory elements, such as promoters and enhancers, which are located nearby or within a respective gene. Trans regulation is less understood and thought to be less frequent but is likely mediated through chromatin looping or proximity due to chromatin conformation. We systematically explored differences of variation in gene expression explained by cis and trans regions on the resident chromosome of the gene, as well as all other chromosomes, separately, for 10,343 genes. Using 210 HapMap samples (and corresponding lymphoblastoid cell lines) we partitioned genetic variation separately for each gene then calculated genetic relationship matrices (GRMs) for each region, representing genomic sharing of additive genetic effects within the respective region. We then fit those GMRs in mixed linear models, estimating the proportion of variation in gene expression explained by each region, for all respective genes. From these data we calculated, for each gene, the amount of gene expression explained overall by local and distant genomic regions. Initial results show that, unsurprisingly, the largest proportion of variance explained was due to variation on the resident chromosome. However, in our initial results we observed that some chromosomes have more overall trans influence than others and that expression of some genes can be explained more by trans variation than cis variation. We hypothesize that this trend may be due to certain chromosomes being more accessible (i. e. physically close to more chromosomes), due to conformation, which we plan to investigate more in the future.

3179F

Functional partitioning of local and distal gene expression regulation in multiple human tissues. *X. Liu[1], H. K. Funucane[1,2], A. Gusev[1], G. Bhatia[1], B. Bulik-Sullivan[3,4], F. A. Wright[5], P. F. Sullivan[6], B. M. Neale[3,4], A. L. Price[1,7].* 1) Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, Massachusetts, USA; 2) Department of Mathematics, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA; 3) Analytic and Translational Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA; 4) Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA; 5) Bioinformatics Research Center, North Carolina State University, Raleigh, North Carolina, USA; 6) Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA; 7) Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA.

Studies of the genetics of gene expression have served as a key tool for linking genetic variants to phenotypes. Large-scale eQTL mapping studies have identified a large number of local eQTLs, but the molecular mechanism of how genetic variants regulate expression is still unclear, particularly for distal eQTLs, which these studies are not well-powered to detect. In this study, we instead use a heritability partitioning approach to dissect the functional architecture of gene regulation. We utilize a method, stratified LD score regression (Finucane et al. biorxiv) that leverages all variants (not just those that pass stringent significance thresholds) to partition heritability by functional category, and adapt this method to analyze gene expression data. For both local (within 1Mb of TSS) and distal gene regulation, we estimate the proportion of SNP-heritability explained by SNPs in each functional category and define enrichment as this proportion divided by the proportion of SNPs in each category. We compute standard errors by jackknifing the set of regression SNPs. We first analyzed a large data set of gene expression in peripheral blood (Wright et al. 2014 Nat Genet). Our results included local functional enrichment of FANTOM5 enhancers (4. 17x), 5' UTR (4. 17x) and TSS (4. 00x), and distal functional enrichment of conserved elements (4. 38x), weak enhancer (4. 37x), and broadly defined DHS (1. 57x). All of these results were highly statistically significant (P < 10-12); the distal enrichments would not be detected from the limited number of significant distal eQTLs in this data. We next analyzed a large data set of gene expression in adipose, skin and LCLs (Grundberg et al. 2012 Nat Genet). Our results show that local functional enrichment is highly consistent across different tissues with average pairwise correlation of 0. 86, whereas enrichments of distal functional elements tend to vary across different tissues (average pairwise correlation = -0. 10). For example, distal enrichment of DHS is significantly different (P = 0. 0007) between adipose (2. 76x) and skin (1. 71x), and distal enrichment of coding regions is significant in blood (4. 31x enriched, P=1. 1x10-6) but not in adipose or skin. Our study identified enriched functional categories for both local and distal gene expression regulation, and revealed variation in functional enrichment among different tissues.

3180T

Characterizing the role of STRs in gene regulation. *D. Zielinski[1], M. Gymrek[2], S. Zaaijer[1], Y. Erlich[1].* 1) New York Genome Center, New York, NY; 2) Broad Institute, Cambridge, MA.

Most variants associated with complex traits reside in noncoding DNA, suggesting an important role for*cis*-regulatory elements. Efforts to discover*cis*eQTLs (expression Quantitative Trait Loci) have mainly focused on the contribution of SNPs to gene expression. However, several candidate gene studies in human and model organisms suggest that Short Tandem Repeat (STR) variations can modulate expression levels and splicing of nearby transcripts. After identifying significant associations between STR variations and expression profiles across hundreds of samples from the 1000 Genomes Project, we are experimentally validating the effects of these candidate STRs on expression levels in human cell lines. Specifically, we are using the CRISPR/Cas genome engineering system to edit STR allele lengths at several loci and measuring the associated gene expression levels. In parallel, we are also conducting dual fluorescent reporter assays to compare intensity levels for varying STR alleles. In addition to validating our*in silico*findings, these approaches will allow us to fine map individual eQTL signals and determine whether the STR is the true causal signal at these candidate loci. .

3181F

Rare non-coding variation in a population isolate from Sardinia. *M. Pala[1,2], Z. Zappala[2], M. Marongiu[1], X. Li[2], J. Davis[2], A. Mulas[1], R. Cusano[3], F. Crobu[1], K. Kukurba[2], K. S. Smith[2], F. Reinier[3], R. Berutti[3], M. G. Piras[1], M. Zoledziewska[1], F. Busonero[1], A. Maschio[1], M. Steri[1], G. Pistis[1], E. Porcu[1], F. Danjou[1], S. Sanna[1], C. Sidore[1], A. Battle[4], A. Angius[1], D. Schlessinger[5], G. Abecasis[6], F. Cucca[1,7], S. B. Montgomery[2].* 1) IRGB, CNR, Monserrato, Cagliari, Italy; 2) Dept. of Pathology, Stanford University School of Medicine; 3) CRS4, Advanced Genomic Computing Technology, Pula, Italy; 4) Center for Computational Biology, John Hopkins University; 5) Laboratory of Genetics, NIA, Baltimore, Maryland; 6) Center for Statistical Genetics, University of Michigan, Ann Arbor, MI; 7) Dipartimento di Scienze Biomediche, Université di Sassari, Sassari, Italy.

Focused on a cohort from the founder population of Sardinia, the SardiNIA project has benefited from an excellent representation of European genetic variation with a shift toward higher frequencies for biologically relevant variants that are rare elsewhere and minimal admixture that enable high-powered studies of both rare and common genetic variants. We use whole genomes and peripheral blood transcriptomes of 624 individuals to identify genetic variants that contribute to both population and individual phenotypic differences. We identify 20,353 independent expression quantitative trait loci (eQTLs) and 11,152 independent splicing quantitative trait loci (sQTLs). We find that over half of all protein-coding and lncRNA genes are influenced by two or more independent *cis*-eQTLs (46. 7% and 46. 8% respectively) whereas pri-miRNA genes are most often influenced by only one *cis*-eQTL. In extreme cases, we find more than 10 independent *cis*-eQTLs for six protein-coding genes and four lncRNA genes. We compare eQTLs to those previously reported in two mainland European-descendent cohorts; after assessing for high reproducibility (>72%), we identify properties of highly-differentiated eQTLs in Sardinia reflective of its population history. We identify an enrichment of high frequency, large effect eQTLs in Sardinia. Considering trait loci identified in published GWAS studies, we find that eQTLs with large frequency difference between Sardinia and Europe are most enriched in SNPs associated with multiple sclerosis (p=1. 17x10-8), a disease with high prevalence on the island. In addition to the classical detection of common eQTLs, we developed a likelihood method that combines total gene expression and family relationship to detect gene expression outliers. It has been demonstrated that outlier gene expression can be driven by rare regulatory variants. Leveraging the familial relatedness of our experimental design (61 families), we identify 802 genes as expression outliers. The outliers show a significant increase in allelic imbalance with respect to non-outlier genes and contain an enrichment of shared rare variants within 1 MB of the TSS. We rank these variants by conservation and deleteriousness, and test their impact on traits measured in Sardinia. Our work provides insight into how studies of gene expression in isolated populations can inform population history, epidemiology and genetic risk factors across the genome.

3182T

Combining whole genome and RNA sequence data from multiple tissues to map causal eQTL using resampling methods. *A. Brown[1,2,3,4], O. Delaneau[1,3,4], E. Dermitzakis[1,3,4].* 1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland; 2) NORMENT, K. G Jebsen Centre for Psychosis Research, University of Oslo, Norway; 3) Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland; 4) Swiss Institute of Bioinformatics, Geneva, Switzerland.

GWAS studies have revealed many genetic regions associated with disease, but for the vast majority the precise causal variant acting remains unknown. Whole genome sequence (WGS) data gives full ascertainment of all variants in the population, crucial to identifying causal variants. As a first step, gene expression studies are ideal for investigating the potential of sequence data to identify causal variants: large expression effect sizes mean sufficient power with only hundreds of samples, and available ChIP-seq experiments can validate putative causal variants. Once causal eQTL are identified, we hope to observe downstream consequences on whole organism traits. Here, we combine 522 whole genome sequences produced by the UK10K project with RNA-seq quantified expression from 4 tissues from the Eurobats consortium (blood, fat, skin and LCLs) to map causal variants affecting expression in disease relevant tissues. A cis-eQTL analysis using sequence based genotype calls did not map many more eQTL (3613-8238 depending on tissue) than array genotype data imputed into 1000 Genomes (3676-8204 , FDR=0. 05). However, using GM12878 ChiP-seq ENCODE experiments, we observed more peak sequence genotype eQTL in transcription factor binding sites and open chromatin regions compared to array genotype eQTL (odds ratio>1 for enrichment in 80/83 experiments, FDR<0. 05 for 66/83). The implication is that eQTL identified using sequence data are more likely functional. When calling eQTLs, the specific peak significance variant depends on factors including variability in phenotype measurements and regional linkage structure. To produce a set of "candidate causal eQTL" (cc-eQTL), we applied bootstrap resampling methods. We call cc-eQTL the SNPs which were peak eQTL most frequently across bootstrap samples. These cc-eQTL are more enriched in GM12878 experiments than peak significance eQTL (OR>1 for 82/83 experiments, FDR<0. 05 for 57/83). We also saw more indels (11. 1% vs 8. 8%) and multiallelic SNPs (9. 3% vs 4. 4%) among cc-eQTL. We identify 663 genes where the same variant was chosen in >95% of the bootstrap samples, strong candidates for experimental validation. Current work is looking at tissue specificity of cc-eQTL, using relevant ChiP-seq experiments, and integration of these causal variants with GWAS signals. We will apply this methodology to the GTEx dataset, which has RNA-seq and WGS for 47 tissues, including ones central to the pathogenesis of a very wide spectrum of diseases.

3183F

Genomic modulators of gene expression in human neutrophils. *V. Naranbhai, B. P. Fairfax, S. Makino, P. Humburg, . D Wong, E. Ng, A. V. S. Hill, J. C. Knight.* Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.

Neutrophils form the most abundant leukocyte subset and are central to many disease processes. Technical challenges in transcriptomic profiling have prohibited genetical genomic approaches to date. Here we report efforts to map expression quantitative trait loci (eQTL) in peripheral blood CD16+neutrophils from 101 healthy European adults. We identify cis-eQTL for 3281 neutrophil-expressed genes including many implicated in neutrophil function and Mendelian primary immunodeficiency syndromes, with 450 of these not previously observed in myeloid or lymphoid cells. We show extensive analyses to delineate those that are not present in other cell types. Paired comparison with monocyte eQTL demonstrates nuanced conditioning of genetic regulation of gene expression by cellular context. We describe underlying histone features, DNA methylation, microRNA binding site and transcription factor binding site variation as it relates to cell-type constraint of eQTL. Neutrophil eQTL are markedly enriched for trait-associated variants particularly autoimmune, allergy and infectious disease and find several novel examples of these. We further demonstrate how eQTL in PADI4 and NOD2 delineate risk variant function in rheumatoid arthritis, leprosy and Crohn's disease. Finally, we define eQTL present in neutrophils and under apparent selection suggesting possible operative forces for variant selection. Taken together, these data help advance understanding of the genetics of gene expression, neutrophil biology and immune-related diseases. .

3184T

Genetic variants affect expression of nearly all genes, but only in a specific context. *P. Deelen, The BIOS consortium.* University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands.

The expression of nearly all genes is subject to genetic regulation. This is a first conclusion from the largest population-based RNA-sequencing project executed so far, including 4,000 total blood mRNA samples from five Dutch biobanks. The large sample size, unprecedented resolution of this study and the comparison with other large population-based studies revealed that SNPs affecting expression (eQTLs) act in a context-dependent manner. Whereas microarray-identified blood eQTLs (Westra *et al.* , Nature Genetics, 2013) replicated well in our study (87%, including 6. 8% with opposite allelic direction), eQTLs identified by RNA-seq in lymphocyte cell lines (LCL) (Lappalainen *et al.* , Nature, 2013) replicated less well (78%, including 12% with opposite allelic direction). This was not due to differences in data processing as we analyzed all RNA-sequencing data using exactly the same pipeline, but was partly due to differences in cell-types. We inferred cell-type-specific eQTLs (Westra *et al.* , PLOS Genetics, 2015), and observed that lymphocyte-specific eQTLs replicated well in LCLs. Neutrophil-specific eQTLs on the other hand showed lower replications rates and more often show opposite effects when comparing to LCLs. We also observed that higher order effects that induce, abrogate, or inverse eQTLs. Using gene-expression levels as a proxy for stimulations, we were able to identify eQTLs affected by different conditions and stimulations. This allowed us to detect, among others, erythrocyte and regulatory T cell-specific eQTLs. We also replicated known eQTL modifiers, for instance the previously identified rhinovirus dependent eQTLs (Çaliskan, *et al.* , PLOS Genetics, 2013). Additionally, we find eQTLs affected by core cellular processes such as metabolism and cell cycle. In conclusion, we observed several examples where genetic risk factors affect gene expression only in a specific context. Since it is not feasible to collect eQTL data for all tissues and cell-types in combinations with all potential stimulations and cellular states, approaches that can untangle the effects of different stimulations are essential to properly interpret the effect of regulatory genetic variation. The BIOS project is funded by the BBMRI-NL, a research infrastructure financed by the Netherlands Organization for Scientific Research (NWO project 184. 021. 007).

3185F

Identifying the effects of postmortem sample collection parameters on RNA-sequencing results across 1,640 samples. *D. S. DeLuca[1], S. Ayellet[1], T. Sullivan[1], F. Aguet[1], E. Gelfand[1], A. Roa[2], H. M. Moore[2], K. Ardlie[1].* 1) The Broad Institute, Cambridge, MA, MA; 2) Biorepositories and Biospecimen Research Branch, National Cancer Institute, Rockville, MD.
In order to study the interaction of genomics and transcriptomics across individuals and tissue types, collection of a wide panel of tissue biopsies is required. For humans, this is only possible in a postmortem setting. Projects of this format, such as the National Institutes of Health Genotype-Tissue Expression (GTEx) Project , must optimize their collection protocols to ensure the highest RNA integrity possible to best facilitate downstream analytical applications such as RNA-sequencing. When effects of postmortem interval or preservation method are unavoidable, it is essential to elucidate the impact of these factors on downstream analyses.

To this end, the Biospecimen Methodological Study (BMS) was initiated by the National Cancer Institute's Biorepositories and Biospecimen Research Branch. The experimental design is based on the collection of six tissues (adrenal gland, artery, muscle, skin, pancreas, thyroid) from 32 individuals across 4 time points ranging from 1 to 15 hours postmortem. Each tissue biopsy was subdivided and preserved in each of three preservation methods: liquid nitrogen, dry ice, and PAXgene Tissue (Qiagen). Following RNA extraction, RNA Integrity Number (RIN) values were determined and Illumina RNA-seq with polyA-selected libraries was performed on a total of 1,640 samples.

Analysis of RNA quality showed that RIN was only modestly negatively correlated with postmortem time course, with a strong influence of tissue type. There was a significant effect of preservation method on RNA integrity. For tissue types prone to rapid degradation, samples preserved with PAXgene Tissue produced RNA with higher RINs. Initial quality control of the RNA-seq results indicated almost no observable effect of preservation method or postmortem interval on the technical success of sequencing when input material had a RIN of 4. 0 or higher. In contrast, gene expression was significantly impacted by the postmortem interval. Between 800 and 3000 genes were differentially impacted by postmortem interval. Correcting for these effects is essential for avoiding spurious associations and for increasing statistical power. The results of this study have important practical methodological implications, not only for GTEx, but for all projects using post-mortem, variable ischemic time, or otherwise degraded samples for analyses of the transcriptome.

3186T

Rare regulatory variants cause cross-tissue extreme gene expression. *X. Li[1], Y. Kim[2], Z. Zappala[3], J. Davis[3], E. Tsang[3], F. Damani[2], A. Battle[2], S. Montgomery[1,3], GTEx Consortium.* 1) Department of Pathology, Stanford University, Stanford, CA; 2) Department of Computer Science, Johns Hopkins University, Baltimore, MD; 3) Department of Genetics, Stanford University, Stanford, CA.

The discovery that most GWAS variants fall in non-coding regions suggests that an individual's genetic burden of disease risk is deeply related to regulatory variants (measured as expression quantitative trait loci; eQTLs). However existing eQTL discoveries have been limited to common variants while the roles of rare variants, abundant in human populations, remain challenging to detect. We use diverse GTEx human tissue samples (8555 RNA-Seq samples from 54 tissues of 522 donors, 157 also with sequenced genome, from the v6 release) to investigate the impact of rare variants on expression and, ultimately, disease. We identify that rare variants can lead to outlier expression levels far above or below the population mean and that outlier expression effects can exhibit across multiple tissues. We first discover the degree to which outlier expression identified in a single tissue and individual can be shared across other tissues. The degree of sharing is reflected by the closeness of tissues and such a gradient is concordant between over-expression and under-expression outliers. Next, we observe a continuous enrichment of rare variants at those outliers given increased number of shared outlier tissues. This enrichment of rare non-coding variants suggests increased selection pressure for variation which influences multiple tissues. Last, we observe that rare regulatory variants associated to outlier expression influence essential genes and genes involved in complex disease, and demonstrate that such effects increase an individual's risk for diverse diseases including diabetes and hypertension.

3187F

Genetic regulation of whole blood gene expression quantified in large families pedigrees. *A. Viñuela[1,2,3], A. Martinez-Perez[4], A. Brown[1,2,3,5], A. Ziyatdinov[4], H. Brunel[4], J. C. Souto[5], M. Sabater-Lleal[6], A. Hamsten[6], A. Buil[1,2,3], J. M. Soria[4], E. T. Dermitzakis[1,2,3].* 1) Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Geneva, Switzerland; 2) Swiss Institute of Bioinformatics, Geneva, Switzerland; 3) Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland;; 4) Instituto de Investigaciones Biomedicas Sant Pau (IIB-Sant Pau). Barcelona; 5) NORMENT, K. G Jebsen Centre for Psychosis Research, University of Oslo, Norway; 6) Atherosclerosis Research Unit, Center for Molecular Medicine, Department of Medicine, Karolinska Institutet, Stockholm, Sweden.

Discovering rare variants affecting gene expression is challenging without very large sample sizes; however, a variant which is rare in the population can be seen more commonly within a particular pedigree. Therefore, methods based on linkage are well placed to discover these variants. Moreover, familial information allows for parent-of-origin studies, where the effect of a variant depends on whether it was inherited from the maternal or paternal line. Here, we describe de analysis of 935 individuals grouped in 35 pedigrees, with an average of 27 individuals per pedigree and a total of 8654 related pairs. All families have at least 10 living individuals in three or more generations, and were recruited and selected as part of a study on idiopathic thrombophilia with an age range of 3 to 101 years (mean 40). Hundreds of quantitative phenotypes were measured at time of recruitment, including anthropometric measurements, hemogram, hemostasis traits, as well as phenotypes related with platelets, platelet activity, homocysteine metabolism, inflammation, and flow cytometry of leukocytes and microparticles. We sequenced the mRNA transcriptome from whole blood for all the individuals, at the time of collection not medicated for thrombophilia, using 49 bp paired end reads. Reads were mapped onto the human genome reference using GEM, obtaining 16. 2M well mapped exonic reads per sample on average. Genotypes obtained from a mix of Illumina 1M and Illumina 500K chips were imputed into a 1000 genomes reference panel. We are performing eQTL mapping on this data, both standard association mapping using variance components to consider the familial structure in the data, and linkage analysis to discover rare variants, which co-segregate in familial individuals. In addition, by treating allelic specific expression (ASE) in family members as a quantitative phenotype, we are aiming to discover particular haplotypes which co-segregate within individuals showing ASE, which will demonstrate the presence of causal eQTL, possibly rare in the population. Furthermore, using the available cell count measures and other phenotypes, we will investigate cell type specific eQTL in this heterogeneous tissue as well as eQTL with specific effects in individuals with idiopathic thrombophilia. Finally, exploiting the many trios within the pedigrees, we will investigate particular parent-of-origin effects on gene expression.

3188T

RNA:DNA hybrids and genome instability in post-mitotic neurons. *V. Bhatia, M. Sanchez, L. Quintanilla, F. Ledesma, A. Aguilera, S. S. Bhattacharya.* Centro Andaluz de Biología Molecular y Medicina Regenerativ, CABIMER, Sevilla, Sevilla, Spain.

R-loops can impair transcription and replication and lead to genomic instability. However, not much is known about the molecular events associated with the formation and processing of R-loops in non-replicating cells and their physiological relevance. Our goals in this study have been to explore the existence and possible role of RNA-DNA hybrids in non-replicating postmitotic neurons. Nascent RNA can reinvade the negatively supercoiled DNA behind moving RNA polymerase and form a structure termed R-loops, where a ssDNA is accompanied by an RNA:DNA hybrid. mRNP biogenesis and export factors work in tandem to prevent R-loop accumulation and to warrant the maintenance of genome integrity. It has previously been shown that R-loops accumulate in absence of factors involved in mRNP biogenesis and export in human cells. Cells depleted of BRCA1 and BRCA2 double-strand break repair factors accumulate RNA-DNA hybrids and that the genomic instability in R-loop dependent manner. R-loops are considered a major threat to incoming replication fork (RF), but they are also detected in non-replicating cells. Here, we wondered whether R-loops are also a threat to genomic instability in non-replicating post-mitotic cells. Using post-mitotic neurons from mice brain and retina we have found that R-loops do accumulate in specific light sensor retinal cells called Photoreceptors. The results suggest a relationship between oxidative stress-associated DNA damage and R-loop accumulation in post-mitotic neurons. R-loops are frequently formed in cells and have been proposed to be a major source of replication stress and cancer-associated genomic instability. We show here that R-loops also accumulate in post-mitotic photoreceptor cells and affect cell survival and apoptosis signaling.

3189F

Complex multiple-nucleotide substitution mutations causing human inherited disease reveal novel insights into the action of translesion synthesis DNA polymerases. *J. M Chen[1,2,3], C. Férec[1,2,3,4], D. N. Cooper[5].* 1) Institut National de la Santé et de la Recherche Médicale (INSERM), U1078, Brest, France; 2) Etablissement Français du Sang (EFS) - Bretagne, Brest, France; 3) Faculté de Médecine et des Sciences de la Santé, Université de Bretagne Occidentale (UBO), Brest, France; 4) Laboratoire de Génétique Moléculaire et d'Histocompatibilité, Centre Hospitalier Universitaire (CHU) Brest, Hôpital Morvan, Brest, France; 5) Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, United Kingdom.

Translesion synthesis (TLS) DNA polymerases potentiate the bypass of unrepaired lesions during DNA replication. Based upon mutational signatures of a subtype (N = 477) of multiple-nucleotide substitution (MNS) mutations causing human inherited disease, we postulated two properties of TLS DNA polymerases in DNA repair: namely, the generation of *neo*-microhomologies for strand-misalignment and additional micro-lesions within the templated inserts when recruited to stalled replication forks. To provide further support for this postulate, we analyzed the mutational signatures of a new and complex subtype of pathogenic MNS mutation. Several mutations containing long templated inserts (8-19 bp), that are highly informative with regard to their underlying mutational mechanisms, harbor imprints of TLS DNA polymerase action. Dissecting the mechanism underlying the generation of the 19 bp-insert implicates repeated participation of TLS DNA polymerases in the conversion of a damaged base into a complex MNS lesion through a process of successive template switching and bypass repair.

3190T

Breakpoint analysis of 43 patients with complex genomic rearrangements and their mechanisms for formation. *G. M. Novo-Filho[1], M. M Montenegro[1], E. A. Zanardo[1], A. M. Nascimento[1], F. A. R Madia[1], T. V. M. M. Costa[1], F. B Piazzon[1], A. T. Dias[1], C. P. Barbosa[3], B. Bianco[3], C. P. Christofolini[3], C. A. Kim[2], L. D. Kulikowski[1,3].* 1) Department of Pathology, Cytogenomics Laboratory , Universidade de Sao Paulo-HC-FMUSP, Sao Paulo, Sao Paulo , Brazil; 2) Genetics Unity, Department of Pediatrics, Children Institute, Universidade de São Paulo -HC-FMUSP Paulo, Brazil; 3) Human Reproduction and Genetics Center - Department of Collective Health, FMABC, Santo André, Brazil.

Genomic rearrangements have been shown to play an important role in pathogenesis of congenital or developmental defects in human diseases. Breakpointanalysis can help define genomic location, sequence structure, genetic content and also could improve understanding of the mechanisms for formation. In this sense, we studied a cohort of 63 patients with congenital malformation and ADNPM previously identified by MLPA with distinct kits (MCR- Holland) with at least one genomic rearrangement. Subsequently, we performed high-resolution array analysis using the Infinium CytoSNP-850K BeadChip (*Illumina Inc*. , San Diego, CA), Genome-Wide Human SNP Array 6. 0®(*Affymetrix, Inc. , Santa Clara, CA*), CytoScan HD and Human Genome CGH Microarray 2x105K® (*Agilent Technologies, Santa Clara, CA*), in order to breakpoint delineation. The results were entered into the *UCSC Genome Browser* using annotation GRCh37/hg19 and confirmed by the *NCBI Map Viewer*. The analysis revealed 43 patients with complex genomic rearrangements and several different breakpoints. Between them ten patients presented breakpoints surrounded by low-copy repeats (LCRs) suggesting non-alellic homologue recombination (NAHR) as a possible accountable mechanism. Whereas microhomology-mediated break induced replication (MMBIR) and fork stalling and template switching (FoSTeS) could be associated with the cause of complex rearrangements in other 29 patients. In four of patients, who presented more than five breakpoints, a chromoanasynthesis event, mediated by MMBRI or FoSTeS, could be the cause of complex genomic rearrangement. The breakpoint analysis is essential to understanding the etiology of complex genomic rearrangements and their association with pathogenic phenotypes. Financial support: CNPq and FAPESP.

3191F

Development of a quantitative targeted RNA-Seq methodology for use in differential gene expression analysis. *M. A. Hussong, M. D. Fosbrink, E. Lader.* Biological Research Content, New Product Development, QIAGEN Sciences, LLC, Frederick, MD.

RNA Sequencing (RNA-Seq) uses the capabilities of Next Generation high-throughput sequencing (NGS) methods to provide insight into the transcriptome of a cell as it generates millions of reads. Whole transcriptome sequencing can be used to quantify gene expression on a transcriptome-wide scale, identify splice variants, quantify allele specific expression, and characterize fusion transcripts. Development of a highly reproducible and sensitive targeted quantitative sequencing method would aid in facilitating a deeper understanding and characterization of the roles of a specific set of genes, while enabling much higher sample throughput and significant cost savings relative to whole-transcriptome sequencing. In this study, we report a targeted RNA-Seq technology, GeneRead RNA Counter (GRRC), which makes use of several methodologies to deliver an extremely flexible, highly precise tool for characterizing gene expression. GRRC incorporates 12-base random molecular barcodes into each unique target strand which benefits quantifying gene expression in a given multiplexed sample. Counting the number of molecular tags allows one to determine the sequence coverage per target and adjust experimental conditions to use the read budget of any sequencing platform most efficiently. Using either the Illumina or Ion Torrent platforms, users can choose to multiplex up to 96 RNA samples from 12 to 1000-plex expression panels. No mRNA selection or rRNA removal or blocking is required. The entire protocol, from cDNA synthesis to finished library, which is ready for sequencing, can be accomplished in under one day. Custom assays for a specific target site can add the ability to distinguish between isoforms or identify allele specific expression. We explore the capabilities of this system by profiling large numbers of genes in a cell model system's response to small molecule treatment. Changes in gene expression by these treatments were measured by GRRC NGS, and fold-changes in gene expression due to chemical perturbation were characterized. Complex gene relationships in perturbed pathways were mapped using QIAGEN's Ingenuity Pathway Analysis (IPA) tool. The IPA tool also facilitated the elucidation of the impact of gene expression changes in the context of biological processes, molecular interactions, cellular phenotypes and disease. This article will provide application data for GRRC including a discussion of technical challenges faced when profiling large numbers of genes in a large cohort.

3192T

A Microfluidic System for Generating Massively Partitioned and Barcoded DNA Sequencing Libraries. *M. Schnall-Levin.* 10X Genomics, Pleasanton, CA.

10X Genomics has developed a reagent-delivery system that transforms short-read sequencers into high throughput and accurate long-read systems with single molecule sensitivity. The GemCode system partitions long molecules of DNA, then prepares sequencing libraries in parallel such that all molecules produced within a partition share the same barcode. Custom software maps short-read data to original long molecules, creating Linked-Reads that span many 10's of kilobases. This enables many applications inaccessible with short-read data including phasing SNPs and indels, and calling and phasing of large-scale structural variants. The fundamental building block of this system is a picoliter droplet that encapsulates DNA, reagents, and a single gelbead to which barcode-containing oligos are covalently attached. A plastic microfluidic consumable creates hundreds of thousands of partitions within minutes. After partitioning, the gelbeads dissolve, releasing barcode-containing oligos and activate a custom library construction reaction. The system has numerous powerful characteristics: (1) generation of long-range (10s to 100s of kilobases) information, (2) compatibility with hybrid capture, and (3) creation of high-quality sequencing libraries from 1ng of DNA. Due to current limitations of short-read data, phasing generally requires statistical methods relying on trio data and/or linkage disequilibrium. These methods are problematic when trio information is unavailable or an individual is not part of a well-studied population. We demonstrate the use of Linked-Read data to phase SNPs and indels from a diverse set of ancestries including multiple European trios, as well as Gujarati Indian, African, and Mexican samples. Results were similar regardless of origin. Whole genome sequencing generated megabase-scale N50 phase blocks, with > 95% of SNPs phased and long switch error rates < 0. 05%. For exome sequencing, we achieved > 95% of genes phased <100kb with long switch error rates < 0. 05%. We also called and phased long-range (>50KB) deletions and inversions in multiple samples and confirmed the accuracy of these calls with multiple orthogonal methods.

3193F

The dynamics of meiotic recombination initiation in laboratory mouse strains. *K. Brick[1], F. Smagulova[2,3], Y. Pu[2], R. D. Camerini-Otero[1], G. Petukhova[2].* 1) Genetics & Biochemistry Branch, NIDDK, National Institutes of Health, Bethesda, MD; 2) Department of Biochemistry and Molecular Biology, USUHS, Bethesda, MD, USA; 3) Inserm UMR1085-Irset, 263 ave du Général Leclerc, 35042, Rennes, France.

Genetic recombination is a key mechanism by which genetic diversity is generated in sexually reproducing species. In mice, humans and most other mammals, the DNA double strand breaks (DSBs) that initiate meiotic recombination are directed to a subset of genomic loci by sequence-specific DNA binding of the PRDM9 protein. This sequence specificity is conferred by a long array of C-terminal zinc fingers (ZFs) that each recognize a preferred DNA sequence. Genome-wide maps of DSB hotspots have been generated for a limited repertoire of *Prdm9* alleles in congenic mice, however to understand the interplay between *Prdm9* and genetic background, we identified and analyzed DSB hotspots in mice with different *Prdm9* alleles representing all four *Mus musculus* subspecies and in their F1 hybrids. As expected, in the six mice homozygous for *Prdm9*, almost all DSB hotspot loci differ and are defined by the *Prdm9* allele. Alleles with common ZF combinations define some shared hotspot loci yet despite this, even a single extra ZF with little predicted effect on DNA binding is sufficient to globally alter the hotspot landscape. In addition to clear dominance of some PRDM9 alleles over others, a striking finding in F1 mice is that of the appearance of "novel" DSB hotspots, not present in either parental genome. To this end, although most hotspots in hybrids coincide with parental hotspots - with no evidence of imprinting - up to 35% of hybrid hotspots are absent from the parental genomes. Gene conversion will destroy highly active PRDM9 binding sites, however these PRDM9 binding sites remain untouched in mice with different *Prdm9* alleles. Thus, in hybrids, with one copy of each parental genome and of each parental *Prdm9* allele, DSB formation can once again occur at these "eroded" binding sites. We will also discuss curiously frequent instances of DSB hotspots that coincide with the sites where DSBs form in the absence of PRDM9 as the use of such hotspots may impede meiotic recombination.